Impact of Undergraduate Medical Training on Housestaff Problem-Solving Performance: Implications for Problem-Based Curricula

Vimla L. Patel, Ph.D., D.Sc.; José F. Arocha, Ph.D.; Michael S. Leccisi, M.A.

Abstract: This article reports a study comparing the problem-solving performance of housestaff with undergraduate medical training in either conventional or problem-based schools. After reading two clinical cases, residents were required to write differential diagnoses and pathophysiological explanations. Biomedical and clinical knowledge used and reasoning strategies were identified. The results suggest that housestaff performance is influenced by the nature of instruction. Housestaff trained in a conventional curriculum (CC) focused on patient information, separated biomedical from clinical knowledge, and used data-driven strategies. Housestaff from problem-based learning curricula (PBLC) organized their knowledge around generated inferences, integrated biomedical and clinical knowledge, and used hypothesis-driven strategies. Data-driven reasoning appears to be impeded in PBLC, suggesting that PBLC students have difficulties in acquiring problem schemata. Previous investigations also found this pattern to be true for medical students trained in two different curriculum formats. Although all housestaff generated equal numbers of diagnostic hypotheses during the reasoning process, housestaff from the conventional curriculum generated a greater number of accurate hypotheses than residents in PBLC. These results are discussed in relation to assumptions in health professional curricula about the adequacy of hypothetico-deductive methods of reasoning as teaching mechanisms and the need for clinical and biomedical knowledge integration.

Dr. Patel is Professor of Psychology and Education, Departments of Medical Informatics and Psychiatry, Columbia University; Dr. Arocha is Assistant Professor of Cognitive Studies in Medicine, Centre for Medical Education, McGill University, Montreal, Canada; and Mr. Leccisi is a doctoral student in the Centre for Medical Education, McGill University. Direct correspondence to Dr. Vimla L. Patel, Department of Medical Informatics, Vanderbilt Clinic Blg., 5th Floor, Columbia University, 622 West 168th Street, New York, NY 10032; patel@dni.columbia.edu.

Key words: cognition, clinical problem solving, problem-based learning, conventional curriculum

Submitted for publication 4/11/01; accepted 8/9/01

Editor’s note: For those of us engaged in dental graduate education, there are lessons to be learned by substituting the word “housestaff” in the title with, for example, “periodontics residents” or “oral and maxillofacial surgery residents,” etc.
knowledge and diagnostic tasks, researchers have found evidence of superior diagnostic performance by students in either conventional curricula or PBLC in accuracy of knowledge and accuracy of diagnosis.8,10

Studies that have attempted to investigate the effects of different curricula beyond the student years have also been conducted. Santos-Gomez and colleagues13 carried out a study to evaluate performance differences between graduates of PBLC and CC schools. Through a Likert-type scale given to physician supervisors, nurses, and housestaff themselves, these authors investigated several factors such as knowledge, communication skills, independent learning ability, critical thinking, teamwork, patient education, attention to health care costs, and self-assessment. The authors found that whereas PBLC graduates were rated as having superior communication skills and being more attentive to health care costs, CC graduates were rated as being more knowledgeable. No other factors were found to be statistically significant. A survey study conducted by Shin et al.14 reported differences between graduates from a PBLC school and a CC school in terms of the former being more up-to-date regarding current clinical practice guidelines for hypertension. They attributed this to the emphasis in PBLC on promoting a more independent learning style. Another survey,15 however, failed to find differences between graduates from both types of curricula in pursuing continuing medical education courses.

Problem-based curricula have also been implemented in various dental schools around the world,16-22 where evaluations of the benefits and drawbacks of PBLC have also reached inconclusive results. Ferguson,22 for instance, found that although PBLC dental students gave evidence of possessing a much broader knowledge, their knowledge of the domain was less detailed. Also, they were able to better integrate knowledge.21 Similarly, other evaluations21 have reported that students in PBLC in dentistry learn as well as their CC counterparts, but have greater group and cognitive analytic and communicative skills. Consistent with results from medicine, PBLC dental students seem to be more motivated and perceive themselves as being better prepared for learning.24-26

The reasons for these results are not clear. In a recent evaluation of the effectiveness of PBLC as compared with CC, Colliver9 argued that despite the theoretical and empirical basis of PBLC, the research thus far suggests that it does not meet the expectations that supporters of the curriculum had hoped for. Colliver makes the point that in reality the theory behind PBLC is too loose to make precise predictions that can be put to the test. We may argue there is no such theory, but rather a number of general frameworks that serve to guide research questions, which may be neither precise nor consistent enough to provide exact predictions. Furthermore, a major problem with investigating performance in different curricula is that these involve many different components (e.g., small group teaching vs. lectures, expert vs. non-expert tutors) and variations among these components. The apparent success or failure of a particular curriculum format may be due to one or a combination of these components, rather than the curriculum as a whole. This situation is comparable to the one faced when attempting to contrast theoretical models. Under these circumstances, it is not advisable to use conventional statistical procedures that test whole models because it is impossible to determine which parts of the theoretical models are responsible for the results and which are not. Rather than investigating overall differences between models or between different curricula, one may focus on isolated aspects of performance and examine these in detail. In our research, we have taken this strategy by characterizing performance under different conditions (e.g., describing the reasoning and knowledge integration processes of physicians who have been trained in one curriculum or the other). Based on the differences found, we can then hypothesize what the sources of these differences may be, based on established cognitive research and what we know about the respective curricula. This type of study, in turn, may suggest hypotheses for further investigation.

Cognitive Studies of Curricular Effects

Research is needed that characterizes specific educational components to assess their impact on specific outcomes rather than attempting to compare whole curricula. Patel and colleagues have conducted studies that attempt to specify characteristics of problem solving by students in different medical curricula.9,30 This research focuses on a characterization of the reasoning processes and knowledge used by medical students in PBLC and CC schools. In this research, it has been found that PBLC- and CC-trained subjects differed in the type of knowledge that they rely on and the reasoning skills that they
apply to clinical problems. It is possible that these differences are due to different types of instructional formats involved in the two curricula. Based on cognitive research, we can identify two types of activities that may be the source of difference between housestaff trained in PBLC and CC. The first type of activity is the explicit teaching of problem-solving strategies, sometimes in the form of the hypothetico-deductive method of reasoning that takes place in PBLC schools. The second is the teaching of basic science in the context of clinical problems. It is assumed that since students learn medical knowledge in the context of clinical problems from the beginning of their training, they have many more opportunities to apply this knowledge and practice case-solving skills. A result of this is that students become better diagnosticians. Furthermore, it is suggested that by explicitly applying the diagnostic strategies, such as the hypothetico-deductive method, students learn how to analyze problems and search for explanations, thereby improving their comprehension of clinical problems. With the increasing reliance of clinical medicine on its scientific roots, training in basic science has become increasingly important. Problem-based learning curricula attempts to impart a mastery of basic science to the student by making the relationship between clinical medicine and its scientific basis more explicit. Since teaching of basic science knowledge takes place in the context of clinical problems, students are better able to integrate these two kinds of knowledge in solving clinical problems.

Although these ideas—explicit teaching of the hypothetico-deductive method and teaching basic science in the context of use—are intuitively appealing, their utility for clinical training has come into question. Some researchers have argued that the hypothetico-deductive method may not be the most efficient way of solving clinical problems. Expertise research has suggested that data-driven reasoning, in which clinical data triggers the diagnostic hypothesis, constitutes the most efficient reasoning pattern and is typical of the expert in routine situations. Hypothesis-driven reasoning is involved in the use of the hypothetico-deductive method, as the latter consists of first generating a hypothesis and then searching for confirming evidence for it, which is essentially what hypothesis-driven reasoning is. Indeed, in hypothesis-driven reasoning, the chain of inferences goes from a hypothesis, typically an unobservable process, to the data, a piece of information given as a patient sign or a symptom. We could argue then that the hypothetico-deductive method involves the use of hypothesis-driven reasoning as its main form of generating inferences.

Cognitive research in other domains has shown that specific training in problem solving strategies, such as using a form of hypothetico-deductive reasoning, may have a detrimental effect on performance due to “schema disruption” and “knowledge fragmentation.” Several studies conducted by Sweller and colleagues have shown that giving students problems to solve and training them in the use of problem solving strategies produces a heavy load on cognitive resources and diminishes student ability to focus on any one part of the task. Students in these circumstances allocate some resources to learning problem solving strategies and other resources to learning the content of the material. It appears that to learn a topic well, students must allocate most of their cognitive resources to the content. Dividing the resources results in a failure to construct adequate problem schemata.

Sweller et al. carried out a series of studies to investigate the effects of training in problem solving strategies on problem solving performance in mathematics. They found that when subjects used a strategy based on the use of data-driven reasoning, subjects were more able to acquire a schema for the problem. In addition, other characteristics associated with expert performance were observed, such as a reduced number of steps to reach the solution. However, when subjects used a strategy based on means-ends analysis (a form of the hypothetico-deductive method), their problem solving performance suffered. Sweller et al. explained these findings by hypothesizing that data-driven reasoning forces students to acquire the needed schema for successful performance. In contrast, means-ends analysis impedes the construction of problem schemata, resulting in fragmentary knowledge and a lack of coherence. In this case, the subject has to divide his or her attention between the application of the strategy and forming a schema of the problem. The result is that neither of these efforts yields success.

These findings have been replicated in other domains. Vollmeyer et al., working in the domain of biology, showed that a strategy similar to means-ends analysis does not provide an effective method of inductive learning and may impede the ability to transfer to other problems. In contrast, their results support the hypothesis that means-ends analysis is not satisfactory as a method for learning from problems. Rather, an approach to learning consisting of
the free exploration of the problem with no teaching of particular strategies helps in a better understanding of the problem. The results of a study by Norman et al. appear to contradict the results obtained by Sweller et al. and Vollmeyer et al. In Norman et al.’s study, a group of subjects were encouraged to use data-driven reasoning by instructing them to obtain the data, synthesize it, and then make the diagnosis. A second group (the hypothesis-driven reasoning subjects) were asked first to diagnose the problem and then to find supporting data for their hypotheses. They found that the hypothesis-driven reasoning group was more accurate than the data-driven reasoning group. We could explain such differences in terms of the objectives of the different studies. In the Sweller et al. study, students were actually trying to learn a topic while also learning a method of approaching the problem. This dual learning and its consequent dissociation of cognitive resources to both tasks are responsible for poor performance. We submit that this is what happens in problem-based curricula. It may be, however, that this only works for highly abstract material (i.e., nonvisual), and that as Norman et al. show, it does not apply when using radiological tasks. Alternatively, it may be that these detrimental effects are felt in dual learning situations, that is, when people have to acquire two types of knowledge at the same time, which was not the case in Norman, et al’s study. Another possible explanation stems from Norman et al.’s use of an experimental design in which naïve subjects were asked to generate diagnoses from the data collected, which was termed “forward-directed” reasoning. This is completely different from the use of a forward-directed or data-driven heuristic by experts, which is acquired after many years of experience in a domain. Furthermore, forcing novices to make a diagnosis of a situation that is unfamiliar and ambiguous will automatically generate “hypothesis-directed” reasoning. We do not think that Norman et al.’s study really challenges the finding showing the use of data-driven reasoning by experts or the deficiencies of “dual-learning”; it may only suggest that, in certain situations, providing a hypothesis to naïve subjects may lead to better accuracy than leaving these subjects to their own devices.

The other claim made by proponents of PBLC—that better learning takes place in the context of use (e.g., that biomedical knowledge should be learned in the specific context of clinical practice)—has also been questioned by cognitive science research. Research has shown that the context-specificity of knowledge depends very much on the type of knowledge being learned. Furthermore, it has been shown that deficiencies in knowledge acquisition are likely to occur when knowledge is too closely aligned with the context in which it was learned, and it is highly improved when abstract instruction takes place before its application in a particular context. In other words, while research has shown that context plays a determining role in learning, as witnessed by the increasing trend toward a situated cognition approach in education, research also shows that not all learning has to be “situated” to be effective.

In the medical domain, Patel and colleagues have shown that teaching basic science within a clinical context may be a disadvantage in that once basic science knowledge is contextualized, it is difficult to separate it from the particular clinical problems into which it has been integrated. By teaching basic science in context, PBLC fails to promote its separation from the particular cases where it has been taught, even when such separation becomes necessary for successful transfer of problem solving. Patel et al. presented evidence for this by showing that students trained in a PBL curriculum failed to separate basic science knowledge from the specific clinical knowledge associated with particular patients. This finding is also in keeping with the results from Vollmeyer et al. regarding transfer of learning. If students have difficulty separating the biomedical knowledge they have learned from the particular clinical cases associated with that knowledge, then it is not surprising that, when given a different problem, they bring to bear on the new problem some irrelevant biomedical knowledge.

In the study by Patel et al., students from the PBLC, using a more hypothesis-driven pattern of reasoning, generated a greater number of elaborations (that is, a detailed chain of information between two main concepts). Elaborations are typically not needed for providing a basic explanation of the case. This was in contrast to the students from the conventional curriculum, who used more data-driven reasoning, fewer elaborations, and showed a higher level of diagnostic accuracy. In terms of knowledge use, students in the PBL curriculum showed a higher reliance on basic science and pathophysiological knowledge, whereas students in the conventional curriculum used more clinical and very little basic science knowledge. The use of data-driven reasoning was closely tied to the organization of clinical information in routine cases. However, in more dif-
difficult cases, subjects used a combination of data-driven and hypothesis-driven reasoning.\(^1\,\text{,}^37\)

In clinical cases where the subjects are presented several segments of information at a time, research has shown that experts generate a small set of highly related hypotheses very early in the case presentation.\(^50\) This is consistent with the findings of other researchers.\(^51\,^52\) However, it is not the number, but the quality of hypotheses that is the distinguishing characteristic of the expert.\(^51\,^53\) The notion that experts have more precise expectation of the findings associated with the diagnostic hypotheses\(^54\) is also in agreement with these results. Novices, on the other hand, typically generate unrelated hypotheses throughout the clinical problem-solving process. In studies by Joseph and Patel\(^53\) and Patel et al.\(^37\), it has been shown that expert subjects are more capable of generating the correct diagnosis early in the problem-solving process than subexperts. Given the differences in approaches to diagnostic problem solving between CC and PBLC students, it is important to understand the extent to which these differences remain beyond medical school training. It seems plausible to assume that once physicians begin their clinical residency training, the differences found between CC and PBLC students would disappear due to the common characteristics of postgraduate training. This study attempts to characterize the use of biomedical and clinical knowledge and different reasoning strategies during problem solving by housestaff who have completed their medical degrees in either conventional or problem-based curricula.

The study presented here is a part of an ongoing research program investigating the effects of instruction and curricular format on reasoning characteristics and knowledge use of students and physicians.\(^9\,^37\) In this paper, we examine how residents who have been trained in different curricula solve clinical problems. This study could be seen as an initial attempt to explore the long-term effects of curriculum type on knowledge organization and reasoning of physicians. Specifically, we examine the cognitive performance of housestaff with undergraduate training from problem-based learning (PBLC) and conventional curricula (CC). We deal with two related issues: the first concerns the directionality of reasoning during clinical problem solving; the second concerns the types of knowledge subjects use when solving various problems. Both issues are investigated in relationship to task difficulty.

### Method

Twenty-one subjects were selected from two areas of medical residency after having graduated from one of two Canadian medical schools: McGill University and McMaster University. At the time the study was conducted, McGill University had a conventional curriculum (CC), where the major characteristic was that the teaching of basic science concepts was carried out before any clinical training. At the time of this study, basic science teaching at McGill was conducted within a lecture format (the curriculum has since changed to include some problem-based-like instructional formats). McMaster University utilizes a problem-based learning curriculum (PBLC) in which basic science is taught in the context of clinical problems.

The subjects were chosen from two areas of medical specialization: internal medicine and family medicine. Six housestaff from internal medicine and three from family medicine volunteered from the CC school. Seven housestaff from internal medicine and five from family medicine volunteered from the PBLC school. In terms of their background, 50 percent of the housestaff in internal medicine at the CC school had a bachelor’s degree in one of the biological sciences, 16.7 percent had obtained a B.A., and 33.3 percent entered the medical school via the Med-P program. The Med-P program allows students coming directly from college to enter the medical training program (rather than obtaining a previous bachelor’s degree). Of McGill housestaff in family medicine, 66.7 percent had a B.Sc., and 33.3 percent came from the Med-P program. Sixty-eight percent of the housestaff in internal medicine at the PBLC school had obtained a B.Sc., 16.7 percent had a degree in veterinary medicine, and 16.7 percent failed to complete the personal history form. Sixty percent of the housestaff in family medicine at McMaster had a B.Sc., 20 percent had a B.A., and 20 percent had a degree in optometry.

### Clinical Material

Two clinical cases were developed as stimulus materials based on actual patient charts. Each case was presented in three segments: the clinical history, the physical examination, and laboratory data. Case 1 depicted a young East Asian male with Hypokalemic Periodic Paralysis with associated Thyrotoxico-
sis. In the first segment of this case, details of the clinical history are presented including two episodes of transient muscle weakness after a large meal and a loss of weight despite increased appetite. Segment 2 of case 1 presents the information obtained from a physical examination of the patient and includes such findings as the presence of a slightly enlarged thyroid, as indications of the underlying pathology. Segment 3 presents the results of laboratory tests. Notable items included a low serum K+ and an elevated T4. The text of case 1 is given below:

Segment 1: A twenty-two-year old Asian male presented with two attacks of transient muscle weakness in the upper region of his legs. The patient fell suddenly while running, thirty minutes after playing squash. The weakness resolved spontaneously five minutes later. The second episode occurred two months later. After sitting for a while, he stood up, took a few steps, and collapsed. This happened about two hours after eating a large meal. Following a night’s rest the weakness had deteriorated to the point where he could not walk nor lift his legs. The weakness was confined to his upper legs. He was subsequently taken to the emergency room. During the episode the patient experienced no loss of consciousness, no dizziness, or severe pain. He had dropped in weight from 120 lbs. to 110 lbs. despite a markedly increased appetite. There was a noticeable increase in thirst and urinary frequency. The patient reported that he had developed a mild heat intolerance. He had experienced sporadic tremors in his hands and fingers during the last six months. He also complained of abnormally dry skin. He had no recent history of illness and was not taking any medication. Past history included a recurrent kidney infection that had been treated successfully several years ago. There was no family history of any thyroid or periodic muscular disorder.

Segment 2: The patient is a thin Asian male with widened palpebral fissures, warm clammy tremulous hands. The vital signs were: BP 120/80, heart rate 110/minutes and regular, temperature 37°C. Examination of the eyes revealed mild exophthalmos and positive lid lag. The extraocular movements were full, fundii were normal, and the visual fields were normal to confrontation.

On examination of the neck, the thyroid was slightly enlarged but firm and nontender on palpation. Both lobes were enlarged, the left more so than the right. There were no palpable nodules. A systolic bruit was audible over the thyroid on auscultation. Examination of the chest was normal. Cardiac exam revealed a hyperdynamic precordium. Auscultation was normal except for a grade 2/6 systolic ejection murmur at the left sternal border that did not radiate to the carotids or apex.

Examination of the abdomen was unremarkable. On neurological examination, the cranial nerves 2-12 were normal. Upon motor examination, he had a fine resting tremor of the hands, mild proximal weakness of the gastrocnemus and quadricep muscles bilaterally. Plantar flexion and extension were normal as was the rest of the motor examination. There was no detectable wasting or fasciculation of any muscles. Sensory and cerebellar exams were normal. The deep tendon reflexes were all 2+ and symmetric except for the knee and ankle jerks that were 1+ bilaterally. The plantar cutaneous responses were downgoing and symmetric.

Segment 3: Laboratory tests included a normal CBC. The urinalysis was normal. The SMAC revealed a sodium of 142 (135-145), chloride was 107 (100-110), creatinine was 0.7 (.5-1.5). The serum potassium was 1.5 (3.5-5.0) and the CO2 combining power was 26.8 (25-28). BUN was 13 (10-20) and random sugar was 105 (<180).

Chest X-ray and EKG were normal. Serum CPK was 116 (<140). Urine electrolytes were potassium 13 (15-40), sodium 122 (>15). Thyroid function testing revealed a serum T4 of 23.1 (5-11), a T3RU of 44.5% (24-36%), the free T4 index was 35.1 (4-13).

Case 2, the more difficult of the two, presents a twenty-nine-year-old Caucasian woman with a diagnosis of Acute Hepatic Porphyria. This case is more difficult because it contains a complex set of findings, which are common to other diseases (lead poi-
soning, infection). A longer presentation of this case is discussed in Sculley et al.55 As in case 1, the information is presented in three segments. The clinical history includes details such as recurrent urinary tract infections, seizures, and progressive weakness. Segment 2, the physical examination, includes items such as the patient’s irritability and muscle tenderness. Information such as the dark reddish-orange color of the urine from the laboratory results, as presented in segment 3, are strongly indicative of the diagnosis. The text of the case is presented below:

**Segment 1:** A twenty-nine-year-old woman was admitted to the hospital because of weakness of six days’ duration. She was in her usual state until three weeks prior to admission when she awoke one day with gradually worsening constant pain in the lower back and abdomen that extended into both anterior thighs. She went to an emergency room in Barbados where she was vacationing at the time and was admitted for treatment of a urinary infection. No official case summary from that hospital is available. The patient says she was given intravenous antibiotics. One day after her admission she suddenly lost consciousness. She was told that the sodium in her blood was very low and that she had suffered an attack of epilepsy. She was discharged on Dilantin, 300 mg bid after one week in hospital, and returned to Canada. The above pains persisted unchanged. Six days prior to admission, she first noticed weakness in the right arm upon attempting to lift a cup of coffee. Five days prior to admission, both arms were so weak she needed help to dress herself. Two days prior to admission she needed help to get out of bed because of bilateral leg weakness.

The patient denies any previous seizures, joint pains or swelling, incontinence, double vision, numbness, recent upper respiratory tract infection or diarrhea, fever or chills, head or back trauma, burning or frequent urination, or vaginal discharge. She denies rash, weight loss, goiter, palpitations, or ingestion of raw meat or seafood.

She smokes one package of cigarettes daily and consumes six to twelve bottles of beer and vague amounts of vodka and diazepam daily. She had discontinued alcohol and diazepam at the time of admission to hospital in Barbados. She is a divorced mother of four who works as a technician in a gasoline factory. Family history is unremarkable save for alcoholism in her mother.

**Segment 2:** On examination she was an irritable woman who looked older than her stated age. BP: 140/80, sitting and lying; HR: 90/minute, regular; T: 36.8°C p.o.; RR: 15/minute. There was no rash, no gingival discoloration, no goiter. The lungs, breasts, heart, and abdomen were normal, as were the pelvis and rectum. The joints were all normal. Neurologic examination revealed mild muscle tenderness in the thighs. Strength in the upper extremities was 2/5 in the right arm, 3/5 in the left arm. The more distal muscles in the upper extremities were graded 3-4/5 on the right and 4/5 on the left. Knee flexors and extensors were 5/5 bilaterally. Muscle tone was not increased. All tendon reflexes were 2+ except for the right triceps and biceps which were 1+. Both toes were downgoing on plantar cutaneous stimulation. Sensation, cranial nerves, and mental status were normal. Coordination was difficult to reliably assess because of weakness but seemed grossly normal.

**Segment 3:** The urine was dark reddish-orange, with a specific gravity of 1.020, negative dipstick for protein or blood and a normal sediment. Hematocrit was 38.9%. The white cell count was 6,300, with 58% mature neutrophils, 40% normal-appearing lymphocytes, 1% monocytes and 1% eosinophil. The platelet count was 291,00 and the erythrocyte sedimentation rate was 8mm per hour. BUN:20mg%, glucose: 123mg%; sodium: 125meg/l; potassium: 4.5meg/l; chloride: 92meg/l; creatinine: 1.0mg%; uric acid: 2.0mg%; bicarbonate: 28meg/l; calcium: 9.3mg%; phosphorus: 3.7mg%; bilirubin: 0.3 mg%; total protein: 6.4gm%; albumin: 3.7gm%; SGOT: 31u/ml; LDH: 78u/ml; alkphos: 541u/l; amylase: 10u/ml; cholesterol: 205mg%; triglycerides: 150mg%; serum dilantin level: 10.6mg/ml;
CPK: 13μmol/l; serum ammonia: 80 μg/dl; serum magnesium: 2.4mg/dl. EKG revealed sinus tachycardia at 100/minute but was otherwise within normal limits. Chest x-ray was normal. Spot urine electrolytes revealed a urinary sodium concentration of 45mg/l. Lumbar puncture yielded clear colorless CSF with one lymphocyte per cubic millimeter, a glucose of 80mg% and a protein of 21mg%. VDRL, ANA, and rheumatoid factors were negative. The serum T₄ was 5.9, serum cortisol: 20μg/dl. Electromyographic (EMG) examination revealed on testing proximal muscles, spontaneous flexibilations at rest, decreased recruitment of motor-unit potentials with maximum voluntary effort, and prolonged deviation of high-amplitude motor action potentials, interpreted by the neurologist as compatible with axonal peripheral neuropathy.

Although neither of the two cases is common in hospital settings, it was expected that the subjects would be able to provide a reasonable interpretation for both cases. Indeed, case 1 had been used previously in another study, where it was solved by expert physicians and housestaff. Both cases were reviewed by an expert physician to make sure that they were not out of the reach of most physicians. Furthermore, the diseases involved are frequently covered in clinical textbooks.

Procedure

Each subject was given a booklet containing the two cases. The cases were presented in their conventional format; that is, the clinical history was given first, followed by the physical examination, and finally the laboratory test results. After reading the cases, the subjects were requested to write a differential diagnosis and explain the underlying pathophysiology of the case. This procedure was identical for both cases.

The instructions given to the subjects were the following:

This package contains two (2) clinical problems. For each problem, please read one segment of patient information at a time. After reading each segment, evaluate the problem by giving in writing your thoughts and reasons supporting your evaluation. Explain the underlying pathophysiology of the problem providing as much detail as you can.

This procedure has been used in past research where subjects are given clinical cases and are then asked to write their pathophysiological explanations. In our past research with written and “think-aloud” (e.g., verbal explanations) protocols, we found that regardless of what procedure was used, students did not demonstrate fundamental differences in terms of their knowledge use and the hypothesized cognitive strategies. The major difference is that verbal protocols are typically longer than written protocols but are also more redundant.

Methods of Analysis

The data were analyzed using techniques of cognitive protocol analysis. These consist of segmenting the written response protocols, the exact transcript of the written responses as well as the stimulus texts, into idea units, or propositions, and examining the relation between these propositions. A scheme for the identification and classification of propositions was used based on previous work. Each protocol was segmented and analyzed by two research assistants trained in propositional analysis techniques. The research assistants worked independently. Interrater reliability calculation was performed (number of agreements/total number of propositions = 0.91). Most disagreements were resolved through discussion. In case of any major disagreement, a third person was consulted to resolve the disagreement propositions. These disagreements involved differences in judgments between the coders and were not critical for the final reporting of the results. An example is coding “membrane” as a part (PRT:) of a cell or as a location (LOC:).

The pathophysiological explanation of each case produced by the housestaff was analyzed for the source of information, whether given in the text or inferred by the housestaff. The directionality of the reasoning in the explanation was assessed in relation to the information source. The method, described in Patel et al., was used to identify the point of origin of each proposition in the pathophysiological explanation. This consists of taking each proposition or idea unit from the explanation and comparing it to those produced from the propositional analysis of each of the three segments of the clinical case: the clinical history, physical examination, and
laboratory data. If a proposition from the pathophysiological explanation was not derived from any of the segments, it was coded as an inference. This method was used to identify two different forms of inferences: those concerning a) the diagnoses (diagnostic inferences), and b) biomedical and pathophysiologically based inferences.

The directionality of the pathophysiological explanations was assessed by representing the propositions as semantic networks.39 In these networks, the propositions themselves are represented by “nodes” that are connected by “links” to represent the relations present among the propositions. The type of link used is indicative of the relation between the proposition depicted. A causal link indicates a causal relation between the propositions (X causes Y), and a conditional link indicates a conditional relation between the propositions (if X then Y). Since a causal rule generally moves away from a diagnosis toward a clinical symptom and a conditional rule leads toward a diagnosis from a clinical symptom, the directionality of the relations is important. A network in which all of the relations between propositions move from observed facts (e.g., symptoms) to a diagnosis is described as generated by a process of pure data-driven reasoning. A network in which all of the relations move in the opposite direction, namely from diagnosis to the observable signs and symptoms, is described as being generated by pure hypothesis-driven reasoning. Links that were neither causal nor conditional were coded as “other,” which largely consist of elaborations. Elaborations include statements that refer to information such as location, LOC (e.g., the patient felt some weakness in her arms), temporal relation, TEM (e.g., the patient stopped drinking three weeks ago), or attribute information, ATT (e.g., the patient looked pale). Due to the varying lengths of the subject protocols, data are reported in mean percentages rather than in raw numbers.

The diagnoses were characterized by their degree of accuracy using a two-point scale based on Patel et al.50 The method of coding in Case 1, Hypokalemic periodic paralysis with associated Thyrotoxicosis, can be decomposed into two components: hypokalemia and hyperthyroidism. The correct diagnosis was rated as 2.0. Mention of both these components without connecting the two warranted a score of 1.5 out of 2.0. If the final diagnosis was given as Diabetes Mellitus, Pheochromocytoma, Myasthenia Gravis, Diabetes Insipidus, Hysterical Conversion Reaction, or Guillain-Barre, the subject was rated 0.5. The reason for this score is that these diagnoses form a set of competing diagnoses to the final diagnosis and cannot be completely excluded from consideration. Any other diagnoses, which were unrelated to the final diagnosis, were considered incorrect and thus were rated 0. The correct diagnosis for case 2 was Acute Hepatic Porphyria. This diagnosis was rated 2.0. Any other form of Porphyria warranted a score of 1.5. The undifferentiated diagnosis of Porphyria, without specifying the type, rated 1.0. If the subject generated one of the diagnoses from the competing set, such as Syndrome of Inappropriate Anti-Diuretic Hormone Secretion, Demyelinating Neuropathy, Acute Rhabdomyolysis with Alcoholism, or Acute Myopathy, they were assigned a score of 0.5. Any other diagnosis was given a score of 0. The accuracy rating for each subject was assessed, and the group score was calculated.

In the analysis of the pathophysiological explanation, we used the number of subcomponents to assess coherence. An explanation with many different subcomponents suggests some understanding of specific aspects of the problems, but little understanding of the overall problem. Coherence is defined in terms of two notions: connectivity, that is, the extent to which any concept in an explanation can be derived from any other concepts; and non-contradiction, that is, the extent to which an explanation is devoid of concepts or propositions that contradict one another. Coherence can be either global or local. Global coherence refers to an explanation where all the concepts are either directly or indirectly connected to each other, with no “loose ends.” Local coherence refers to an explanation where some concepts are connected into clusters, but the clusters are not connected among themselves.

Our way of measuring coherence differs from the one used by Hmelo,10 who measured it in terms of the number of relational operators that are chained in an explanation, i.e., connecting findings and relational concepts. In Hmelo’s analysis, an explanation with higher number of operators (a longer chain) is considered more coherent than an explanation with fewer operators (shorter chain). In our analysis, the number of operators (e.g., chains of inferences linking concepts) is irrelevant to the measure of coherence, where we focus on the number of separate components, in a semantic network (the higher the number of components the less coherent the explanation is). In turn, components of a semantic network are determined by argument overlap in a propo-
sitional analysis. An explanation is coherent if there are no breaks in the inference chaining, regardless of the number of concepts that form the chain. The reason for choosing the connectivity of the explanation is to allow various levels of abstraction in the explanation. For instance, experts are known to use shorter chains of inference at a higher level of abstraction than novices. These shorter chains of inference have greater explanatory power than those explanations used by novices. This difference in our notions of coherence may explain the differences between Hmelo’s results and those obtained by Patel et al.9 Actually, Hmelo’s notion of coherence is closer to our notion of elaboration, which is defined in terms of the number of concepts and inference steps in an explanation. Thus, the longer the explanatory steps, the higher the number of elaborations.

Finally, the number of hypotheses generated as a function of information presented (after each of the three segments) was calculated. Hypotheses were coded only the first time they occurred (new hypothesis). These hypotheses were also evaluated for accuracy with the help of a domain expert. Diagnostic hypotheses refer to specific medical disorders, and can be distinguished from prediagnostic hypotheses (e.g., endocrine problem).

Results

The results are presented in terms of the two different undergraduate training backgrounds of the housestaff: CC and PBLC, divided into their areas of specialization: internal and family medicine. The analyzed data was for the sources and use of knowledge, directionality of reasoning, coherence and accuracy, and finally hypothesis generation. Given that our main purpose was to characterize the subjects’ cognitive processes rather than to test hypotheses, we analyzed all the dimensions (e.g., source and use of knowledge, directionality, coherence and accuracy, and hypothesis generation) separately.

Source and Use of Knowledge

This analysis (descriptive statistics are presented in Table 1) looked at the type of information used by the subjects in their explanations. In particular, we looked at the source of the information used by all groups. We calculated the mean number of text propositions (history, physical exam, and laboratory) and inferences (diagnostic, biomedical, and other) from the pathophysiological explanations of Case 1 and Case 2 by internal and family medicine housestaff from conventional and problem-based curricula. Housestaff in both curricula generated more inferences than text propositions. Specifically, the housestaff from the CC generated a mean of 21.67 inferences compared to 5.66 text propositions in Case 1, and a mean of 12.84 inferences compared to 3.84 text propositions in Case 2. Similarly, the housestaff from the PBLC generated a mean of 22.37 inferences compared to 5.11 text propositions in Case 2, and a mean of 20.91 inferences compared to 2.66 text propositions in Case 2. However, overall no statistically significant differences were found between the

| Table 1. Mean and standard deviation for source of text information used and inferred from clinical text in pathophysiological explanations by internal and family medicine residents from conventional and problem-based curricula for both cases 1 and 2 |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                 | History         | Physical Exam   | Laboratory      | Text Inferences |                  |
|                 | M   | SD  | M   | SD  | M   | SD  | M   | SD  | M  | SD  | M  | SD  |
| **CASE 1**      |     |     |     |     |     |     |     |     |     |     |     |     |
| CC School       |     |     |     |     |     |     |     |     |     |     |     |     |
| Internal Medicine | 0.50| 0.84| 1.83| 2.56| 1.67| 1.86|
| Family Medicine  | 1.00| 1.00| 0.33| 0.58| 0.33| 0.58|
| PBLC School     |     |     |     |     |     |     |     |     |     |     |     |     |
| Internal Medicine | 0.71| 1.11| 1.14| 1.77| 0.86| 0.90|
| Family Medicine  | 0.40| 0.89| 1.00| 1.00| 1.0 | 1.23|
| **CASE 2**      |     |     |     |     |     |     |     |     |     |     |     |     |
| CC School       |     |     |     |     |     |     |     |     |     |     |     |     |
| Internal Medicine | 0.67| 1.21| 0.67| 0.82| 0.50| 0.84|
| Family Medicine  | 1.50| 2.12| 0.00| 0.00| 0.50| 0.71|
| PBLC School     |     |     |     |     |     |     |     |     |     |     |     |     |
| Internal Medicine | 0.83| 1.17| 1.83| 2.14| 0.0 | 0.0 |
| Family Medicine  | 0.00| 0.00| 0.00| 0.00| 0.00| 0.00|
|                  |     |     |     |     |     |     |     |     |     |     |     |     |
| **Diagnosis**    |     |     |     |     |     |     |     |     |     |     |     |     |
| **Biomedical**   |     |     |     |     |     |     |     |     |     |     |     |     |
| **& Other**      |     |     |     |     |     |     |     |     |     |     |     |     |

1208 Journal of Dental Education • Volume 65, No. 11
two schools for the use of text propositions in Case 1 F(1,21) = .63, p > .05 or in Case 2 F(1,18) = .09, p > .05. No significant differences were found between the two schools for the generation of inferences for Case 1 F(1,21) = .60, p > .05 and in Case 2 F(1,18) = 2.54, p > .05. All subjects used more biomedically based inferences than diagnostic inferences, which may reflect two factors: the relative difficulty of the cases and/or the level of expertise of the subjects. Housestaff from the CC generated slightly more diagnostic inferences, whereas PBLC housestaff generated more biomedically based inferences. The housestaff from family medicine in both CC and PBLC appeared to have had more difficulty with case 2. None of these differences were statistically significant.

**Directionality of Reasoning**

This analysis looked at the directional pattern (data ♦ hypothesis vs. hypothesis ♦ data) of the subjects. Previous research has shown that CC students use somewhat more data-driven reasoning than PBLC students. The question we ask is: Is this the case also for residents? The results depicted in Table 2 show that CC housestaff consistently generated the highest number of data-driven links in their pathophysiological explanations, with an average of 2.37 data-driven links per network, as opposed to PBLC housestaff who generated an average of 1.5 data-driven links. This was similar for both cases. In contrast, CC housestaff generated fewer hypothesis-driven links (with an average of about 3.75 links) than PBLC housestaff (with an average of about 5 links). However, the CC housestaff consistently generated fewer elaborations than the PBLC housestaff (3.5 vs. 4.3, respectively).

A multivariate analysis of the differences between schools across cases showed significant differences for the number of data-driven links (F[3,31] = 7.248; P < .001) and for elaborations (F[3,31] = 4.415; P < .01). On further analysis, using the Tukey test, the amount of data-driven reasoning for the CC internal medicine housestaff was found to be significantly different from PBLC internal medicine (F[3,35] = 7.864; P < .001), PBLC family medicine (F[3,35] = 7.864; P < .002), and for CC family housestaff (F[3,35] = 7.864; P < .05). The number of elaborations present in the explanations of CC internal medicine housestaff was found to be significantly different from PBLC internal medicine housestaff (F[3,35] = 3.846; P < .01). There was also an overall significant interaction between schools and cases for the number of elaborations (F[3,31] = 3.846; P < .05). Further analyses did not show statistically significant results. The differences in data-driven and hypothesis-driven inferences are illustrated in Figures 1 and 2.

Figure 1 gives the explanation protocol generated for case 1 by one of the housestaff (#4) from the conventional curriculum, together with its semantic representation. The housestaff uses given clinical information (large meal) to generate the hypothesis of hypokalemia, using data-driven reasoning. In addition, the diagnostic hypothesis (hypothyroidism) is used to explain muscle weakness, which is given information, using hypothesis-driven inferences.
"Hyperthyroidism can result in shifts of potassium as well as loss of potassium, which can lead to the hypokalemic state. Severe hypokalemia can lead to muscle weakness and periodic paralysis. After the large meal there may have been a surge of insulin shifting potassium intracellularly, which lead to hypokalemia and the resulting paralysis."

**Coherence and Accuracy**

This analysis looked at the way explanations provided by the subjects tie together the different

```
Figure 2 gives the explanation protocol generated by one of the housestaff (#4) from the PBL school, for the same case, together with its semantic representation. This housestaff uses the state of hypermetabolism as a diagnostic hypothesis and goes on to explain the mechanisms by which such a state may be produced in much detail. Most of the explanation is either causal in nature or composed of elaborations.

Figure 2: Explanation protocol and semantic representation of the pathophysiological explanation for case 1 by a housestaff (4) from the problem-based learning curriculum. Square boxes indicate given information; rounded boxes indicate diagnostic components; CAU: indicates a causal relation; COND: indicates a conditional relation; RSLT: indicates resultive relation. Arrows indicate directionality.
```

"Hypermetabolic state: Abnormality caused by thyroid simulating antibodies, which activate thyroid adenylate cyclase and compete with TSH for receptors on thyroid cell membranes. The increased stimulation of TSH receptors causes release of T₃ (peripheral T₃) with increased metabolism. Also suppression of TSH through negative feedback."

**Thyroid stimulating antibodies**

```
CAU: Activation of adenylate cyclase
   | Competes for TSH receptor site in thyroid cell membrane
   | Suppression of TSH by negative feedback
   | Release of T₃
   | RSLT: Increased metabolism

Figure 2: Explanation protocol and semantic representation of the pathophysiological explanation for case 1 by a housestaff (4) from the problem-based learning curriculum. Square boxes indicate given information; rounded boxes indicate diagnostic components; CAU: indicates a causal relation; COND: indicates a conditional relation; RSLT: indicates resultive relation. Arrows indicate directionality.
```
concepts they used, using a measure of explanatory coherence based on concept linking. Coherence is measured by the number of subcomponents in the pathophysiological explanation provided by the housestaff. This is presented in Table 3, along with the ratings of diagnostic accuracy. The number of subcomponents in the explanation of internal medicine housestaff were lower than those of family medicine housestaff in both the clinical cases. The only exception was the PBLC family medicine housestaff in case 2, whose protocols had the lowest number of subcomponents. However, overall diagnostic accuracy for housestaff from CC was significantly higher than for housestaff from PBLC (F[2,30] = 4.434; P < .05). On further analysis, Tukey tests on accuracy showed CC internal medicine housestaff to be significantly different from CC family medicine housestaff (F[3,38] = 8.023; P < .01), from PBLC internal medicine housestaff (F[3,38] = 8.023; P < .001), and from PBLC family medicine housestaff (F[3,38] = 8.023; P < .01). Furthermore, an overall difference between cases was also found to be significant (F[1,30] = 61.735; P < .001). The results show a relationship between an increase in diagnostic accuracy with a decrease in the number of subcomponents. For obvious reasons, if a physician knows what the underlying problem of a case is, its explanation will likely be more coherent.

The results are consistent with the findings from previous research in that the subjects from the CC used more data-driven reasoning as compared to hypothesis-driven reasoning. In addition, subjects from the CC used fewer elaborations than PBLC students in their explanations of the problems. The lower degree of accuracy for the PBLC may reflect the fact that they generated more elaborations and hence longer explanations. It could be argued that this should increase the likelihood of making factual mistakes and, as a result, generating more diagnostic errors.

Figure 3 gives a semantic representation of the explanation protocol generated by a PBL housestaff (48) for case 1 as an illustration of a number of subcomponents in the explanation. There are four subcomponents, and each one is locally coherent with elaborations, without apparent contradictions. The network shows a lack of global coherence, however, since these subcomponents are not connected. This protocol is representative of the types of explanations frequently observed in the PBL housestaff participating in this study. In contrast, Figures 1 and 2 provide examples of coherent representations, where there are only single connected components. The nature of the protocol in Figure 1 is representative of the explanation generated by housestaff from the conventional curriculum who participated in this study.

### Hypothesis Generation

An analysis of the time course production of new diagnostic hypotheses focused on the differences between the two schools in terms of (a) the total number of new hypotheses produced and (b) the number of accurate new hypotheses produced with presentation of each segment of the case description by problem type. Figure 4 gives the mean new hypotheses produced and the mean new accurate hypotheses generated, with each segment of information for housestaff in the two schools for case 1.

There were no overall differences in the pattern of new hypotheses generated between the two schools. However, there were significant differences between the two schools with respect to accuracy of the generated hypotheses. This was true only at the level of the initial evaluation of the patient history (F[1,38] = 16.862; p < .001).

<table>
<thead>
<tr>
<th></th>
<th>Number of Subcomponents</th>
<th>Diagnostic Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>CASE 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC School</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal Medicine (n=6)</td>
<td>1.83</td>
<td>0.75</td>
</tr>
<tr>
<td>Family Medicine (n=3)</td>
<td>2.67</td>
<td>0.57</td>
</tr>
<tr>
<td>PBLC School</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal Medicine (n=7)</td>
<td>2.00</td>
<td>1.41</td>
</tr>
<tr>
<td>Family Medicine (n=5)</td>
<td>2.20</td>
<td>1.31</td>
</tr>
<tr>
<td>CASE 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC School</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal Medicine (n=6)</td>
<td>1.17</td>
<td>0.41</td>
</tr>
<tr>
<td>Family Medicine (n=3)</td>
<td>2.50</td>
<td>0.71</td>
</tr>
<tr>
<td>PBLC School</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal Medicine (n=7)</td>
<td>2.43</td>
<td>2.15</td>
</tr>
<tr>
<td>Family Medicine (n=5)</td>
<td>1.25</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Similar results were obtained for case 2. Figure 5 gives the mean new hypotheses produced and mean accurate hypotheses generated, with each segment of information for housestaff in the two schools for case 2. There were significant differences between the two cases with respect to the number of accurate new diagnostic hypotheses generated. Once again this was true only at the level of evaluating the patient history ($F[1,38]=6.616; P<.05$).

The results show that the housestaff from the CC generated a larger number of accurate diagnoses early in the problem encounter and narrowed it quickly, as further information was made available. The housestaff from PBLC entertained a wider range of hypotheses, both accurate and inaccurate, early in the problem encounter and selected the accurate hypothesis as further information was made available. The larger number of hypotheses generated at the
first segment of information by the PBLC subjects suggests that they must have considered the clinical problems within a much larger problem space. This seems to be the case at least for the less complex problem (case 1), but less so for the more complex problem (case 2). Each of the hypotheses generated was classified as belonging to one of nineteen disease categories. This classification shows that for clinical problem 1, CC subjects generated hypotheses from ten categories of disease, whereas PBLC subjects generated hypotheses from twenty categories of diseases. For clinical problem 2, however, CC subjects generated hypotheses from twenty-one disease categories and PBLC subjects from eighteen categories.

In summary, the results from this study show that housestaff who had their medical training in either conventional or problem-based curricula interpreted and explained the patient problems differently. Clinical concepts were used more by those housestaff trained in a CC school and biomedical concepts used more by those housestaff trained in a PBLC school. Differences were also found in the directionality of the subjects’ reasoning. Consistent with the type of concepts generated, CC housestaff displayed a greater use of a pattern of data-driven reasoning than PBLC subjects. In turn, PBL housestaff showed more hypothesis-driven reasoning and elaborations than CC housestaff. This pattern of results is similar to previously reported results from the final year students in the CC and PBL schools and suggests possible long-term effects of curricular exposure.

General Discussion

We find it remarkable that subjects trained in PBLC do not seem to have acquired data-driven reasoning, even in routine problem-solving tasks. As has been suggested elsewhere, data-driven reasoning patterns are developed through the specific organization of clinical knowledge, which typically occurs in the context of practice. Studies by Patel et al. showed that students from the CC schools focus on...
clinical information from the given texts and use little biomedical information, except when they encounter a difficult problem. In contrast, students in the PBLC were shown to generate biomedical inferences, rather than categorizing specific signs and symptoms related to the problem. Little development of data-driven reasoning relative to hypothesis-driven reasoning was seen in these PBLC students. In the current study, this pattern of reasoning also occurred in the housestaff trained in the PBLC, suggesting that the organization of knowledge and the directionality of reasoning acquired during the undergraduate medical training persists in residency training programs. However, we cannot conclude that the findings are due to particular aspects of their respective training (e.g., use of small-group sessions, teaching of basic science in context). We can only advance the hypothesis of the relationship between training and later cognitive performance. Furthermore, we have made no attempt to determine if there is any other effect that PBLC or CC may have on housestaff performance.

Another hypothesis is that the results reflect pre-existing differences between the groups; for instance, it could be argued that the differences could exist before the subjects entered their respective medical schools. However, although this is certainly a possibility, the subjects in our study were already advanced in their careers as physicians, and we find it unlikely that differences existing before they entered medical school—several years before the study was conducted—could have such a long-lasting effect. Furthermore, the fact that we have found a great similarity between the patterns of reasoning between PBLC medical students and PBLC housestaff suggests that there may exist a relation between their performance during residency and their medical school training. However, recognizing the limitations of our study, we acknowledge that a more direct investigation of this issue is needed.

A third possible reason for the results is that they simply reflect differences in the schools themselves (McMaster vs. McGill), rather than differences between PBLC and CC. The previous study by Patel and colleagues was also conducted at these two universities. However, we found a similar pattern of results, as reported here, in a technical report that included students from other PBLC and CC schools in the provinces of Ontario and Quebec in Canada. In this study, diagnostic explanations of three groups of medical students were compared. One group of subjects was from a PBLC school, a second group was from a CC school, and the third group was from a hybrid school (with both PBL and CC components). It was found, as before, that PBLC students used more hypothesis-driven reasoning, CC students used more data-driven reasoning, and students from the hybrid school used a similar amount of both.

Another factor to take into account when considering the implications of this study is that we only used two clinical cases. This constitutes a limitation since it has been shown that medical students and physicians vary a great deal in the way they solve different clinical cases. The sample sizes for this project were small, and we acknowledge that this limits the generalizability of the results. However, the aim of this study and much of cognitive research is to characterize in detail the cognitive processes used by the subjects and not just to search for statistical differences. Certainly large-scale studies are needed; hopefully, this and other research will provide the platform for these more ambitious studies.

Although all the above are plausible factors involved in explaining our results, we believe that they alone cannot explain why the particular patterns of reasoning and knowledge integration were exhibited by our subjects. To account for these patterns, more specific theory-based explanations are needed. We can envision several hypotheses that account for this difference in performance. The first is that the knowledge of PBLC subjects is organized in such a way that data-driven reasoning cannot be easily developed. Pure data-driven reasoning is facilitated by a hierarchically organized knowledge base that has been developed over many years of experience, but cannot be explicitly taught. This expert-like strategy is triggered by specific clinical signs and symptoms related to the diagnostic problem under consideration and has been found even in studies, which unlike the present one, give only partial data to the subjects, forcing them to produce preliminary hypotheses. When expert physicians were presented with a case in segments, they generated general hypotheses at the beginning of the case, which served to limit the possible diagnoses. Once the hypothesis was generated, the experts simply interpreted further data as confirming their previously generated diagnosis. This is an instance of data-driven reasoning, as they used the data to generate and refine their first generated hypothesis (e.g., Findings 1, 2, 3 are consistent with

* For information about this study, readers are requested to contact the corresponding author.
the diagnosis), rather than generating a diagnosis and inferring expected findings from it (e.g., If the diagnosis is correct then look for finding 1, finding 2, finding 3). Subexperts, in contrast, generated several hypotheses at the same time and explored each hypothesis in detail, even after they had generated the correct diagnosis.

In this discussion, we simply provide an explanation as to why PBLC housestaff employ little data-driven reasoning. This is not to say hypothesis-driven reasoning is not useful. On the contrary, it is a powerful and useful strategy for dealing with complex cases. The strategies used by the housestaff are probably linked to the quality and quantity of knowledge they possess. The PBLC students and housestaff use hypothesis-driven reasoning to integrate clinical data with basic science information. This integration, which is one of the fundamental assumptions of the PBLC, actually impedes the development of the reasoning pattern used by experts. Since data-driven reasoning is highly automated, it promotes efficiency and accuracy when used in the appropriate context (i.e., when one’s knowledge base is adequate and when the problem is routine). This brings into question the necessity of fostering integration of biomedical and clinical knowledge as one of the principle goals of medical curricula. The necessity of biomedical knowledge, however, varies with the field and the task. Biomedical knowledge, for instance, seems to play a more central role in perceptual domains such as dermatology and radiology where extensive knowledge of anatomical structures is critical in detecting abnormalities. This would explain why Norman et al. found that fostering the consideration of the hypothesis produced better diagnostic performance. The use of biomedical knowledge is less pervasive in routine medical problem solving in other subdomains of medicine such as cardiology and endocrinology.

The highly organized knowledge base of the expert has been found to account for an observed trend in hypothesis generation and evaluation processes. Furthermore, experts also have the ability to organize and manage large amounts of information in a coherent way in order to arrive at a resolution of the problem. The study presented in this paper shows that housestaff from the CC school generate a large number of new and accurate diagnostic hypotheses early in the problem encounter. This is unlike the housestaff from the PBLC, who also generated a large number of diagnostic hypotheses but with only 23 percent of accuracy. The generation of accurate and inaccurate hypotheses by PBLC housestaff is suggestive of the “intermediate level” phenomenon observed in many studies of expertise. It is likely that this pattern of results is related to the hypothetico-deductive method of teaching in PBLC. The students in this curricular format are specifically asked to generate many alternative hypotheses, which eventually are tested against new evidence.

We argue that because clinical problem solving demands the coordination of multiple tasks and goals, the ability to organize and communicate observations in a coherent form is an absolute prerequisite for optimal medical practice. The suggestion is that the integration of basic biomedical knowledge and clinical knowledge in medical school may, for the purpose of solving diagnostic problems, be an inappropriate approach, given the differences between the two bodies of knowledge. We should note that coherence does not necessarily imply accuracy. A line of reasoning can be perfectly coherent, but may be based on inaccurate assumptions, leading to inaccurate conclusions.

The acquisition of problem-solving skills is perceived as being very important in problem-based education. Educational approaches have been designed based on the idea that problem-centered education is superior to conventional education. It is argued that students who are taught problem-solving skills, in particular through the use of the hypothetico-deductive method coupled with problems to practice those skills, learn in a more meaningful way. On the other hand, as Elstein has pointed out, several researchers have been skeptical of the utility of training in the use of such methods. The reasons for skepticism vary, but the one that concerns us most here is that both knowledge organization and schema acquisition seem to be more important for the development of expertise than the use of particular methods of problem solving. In this regard, cognitive research has shown that in order to achieve expertise in a domain, learners must acquire the necessary schemata that allow them to meaningfully and efficiently interpret information and identify the problem structure. Schemata accomplish this by guiding the selection of relevant information and the screening out of irrelevant information.

Based on the results from our study and those from other researchers, we can add another reason for the skepticism regarding the use of hypothetico-deductive methods as a means of learning. Mounting evidence suggests that the use of goal-directed
heuristics, such as the hypothetico-deductive method, and their repeated application to problems could prevent learners from acquiring the schemata necessary for comprehending and solving problems. The evidence shows that when subjects are asked to solve problems either by means-ends analysis (a form of hypothetico-deduction method) or by inducing a rule based on the problem, those subjects who use the first approach fail to understand the structure of the problem. Furthermore, when subjects are prevented from using a means-ends analysis strategy, their learning of the problem structure is more efficient and rapid than when they are allowed to use such a strategy.

This finding can be accounted for by the effects of dividing attentional resources and increased working memory load on schema acquisition during problem solving. In solving clinical problems, subjects must attend to the current diagnostic hypothesis, the data in the problem presented to them, and any intermediate hypothesis between the diagnosis and the patient data (e.g., a pathophysiological process underlying the signs and symptoms). If we consider that more than one hypothesis has been generated, the cognitive resources needed for maintaining this information in working memory must be such that few cognitive resources are left for acquiring the problem schema. Although problems can be solved successfully using the hypothetico-deductive method, the scarcity of attentional and memory resources may cause difficulties for students in adequately acquiring problem schemata. We hypothesize that one of the reasons for the failure of PBLC subjects to acquire a data-driven reasoning style, as found in this study, may be due to the use of problem-solving strategies, such as the hypothetico-deductive method, as a learning strategy.

Schema acquisition and automatization play a major role in achieving expert reasoning and problem-solving skills. These processes help experts circumvent their limited working memory capacity. The consequence of the application of hypothetico-deductive methods during learning is that the cognitive resources of the subjects are directed away from processing the content of the problems, but towards processing the strategic aspects involved in the application of the particular method they have been taught. The present study, together with previous research by Patel et al., suggests that this hypothesis may be a possible explanation for the results concerning the cognitive effects of problem-based and conventional curricula.

Implications for Professional Education

There is increasing awareness among educators and educational researchers of the necessity to modify current educational practices in view of the failure of traditional education to achieve its goals and the great societal changes currently taking place, such as growing competitiveness and education costs. Changes in health professional education are taking place without considering whether or not the pertinent psychological research supports such changes. In recent decades, numerous studies have shown how seemingly sensible and philosophically well-supported educational approaches fail when tested in real world situations (e.g., discovery learning). These approaches were put into place without a thorough and detailed examination of their cognitive foundations. Without such research, one is at a loss to explain the reasons for such failures, given the complexity of educational approaches. Thus, investigation of the processes underlying any educational change is needed before such changes are carried out. Assessing learning and improving learning methods require research in contexts that are consistent with the skills under investigation.

As Anderson et al. have argued, the extent to which learning is associated with any context depends on the body of knowledge being acquired. If the knowledge is to be used in a single context, contextualization is beneficial, since it helps recall by providing extended elaboration. However, there has to be some independence of knowledge from the context of use for knowledge to be transferable. As is the case with basic biomedical knowledge, this provides explanations for many different clinical presentations occurring in diverse contexts.

We believe research on the cognitive processes underlying performance in educationally relevant tasks may provide insights useful to dental education. Although the nature of dentistry (e.g., the issues involved, the type of training) is somewhat different than that of medicine, there is in both curricula the need for a balance between abstract and practical knowledge. Although practical, hands-on knowledge is probably the most important for any profession, we need to make sure that such knowledge is grounded in strongly theoretical knowledge. This theoretical knowledge may not be used in most ordinary situations, but it is indispensable when problems or impasses in practical knowledge occur. Al-
though a practice-based education, such as that found in PBLC, may provide useful insights into teachings that motivate students to learn on their own and to acquire high-level inquiry skills, it does not seem to provide all the answers. Both PBLC and CC have benefits and pitfalls. Since there is a widespread adoption of PBLC in medical and dental schools, it would be worthwhile to recognize the problems in PBLC and to explore ways to correct some of the flaws. Aspects of traditional instruction have a place in current health science education. We need detailed investigations of the cognitive tasks that physicians and dentists perform; that information will enable us to design instructional procedures that balance the advantages of both forms of instruction to make knowledge transfer possible. There would appear to be much to be gained from building a stronger cognitive basis for problem-based education.

Acknowledgments

The research reported in this paper was supported in part by a grant from Social Sciences and Humanities Research Council of Canada (9410-92-1535) to Vimla L. Patel. The authors would like to thank the housestaff from McGill and McMaster Medical Schools for their cooperation with this study. We acknowledge Geoffrey Norman for his assistance with the data collected from the McMaster subjects. We also wish to acknowledge Earl Hunt, Alan Lesgold, and David Kaufman for their valuable comments on earlier drafts of the paper and Timothy Branch for assistance with the final editing.

REFERENCES