

# Tracking Development of Clinical Reasoning Ability Across Five Medical Schools Using a Progress Test

Reed G. Williams, PhD, Debra L. Klamen, MD, MHPE, Christopher B. White, MD, Emil Petrusa, PhD, Ruth-Marie E. Fincher, MD, Carol F. Whitfield, PhD, John H. Shatzer, PhD, Teresita McCarty, MD, and Bonnie M. Miller, MD

## Abstract

### Purpose

Little is known about the acquisition of clinical reasoning skills in medical school, the development of clinical reasoning over the medical curriculum as a whole, and the impact of various curricular methodologies on these skills. This study investigated (1) whether there are differences in clinical reasoning skills between learners at different years of medical school, and (2) whether there are differences in performance between students at schools with various curricular methodologies.

### Method

Students ( $n = 2,394$ ) who had completed zero to three years of

medical school at five U.S. medical schools participated in a cross-sectional study in 2008. Students took the same diagnostic pattern recognition (DPR) and clinical data interpretation (CDI) tests. Percent correct scores were used to determine performance differences. Data from all schools and students at all levels were aggregated for further analysis.

### Results

Student performance increased substantially as a result of each year of training. Gains in DPR and CDI performance during the third year of medical school were not as great as in previous years across the five schools.

CDI performance and performance gains were lower than DPR performance and gains. Performance gains attributable to training at each of the participating medical schools were more similar than different.

### Conclusions

Years of training accounted for most of the variation in DPR and CDI performance. As a rule, students at higher training levels performed better on both tests, though the expected larger gains during the third year of medical school did not materialize.

*Editor's Note: A commentary on this article appears on page 1073.*

**T**he acquisition of clinical reasoning skills—the ability to organize and use medical knowledge and reasoning ability to diagnose medical problems—is essential for medical students, yet is poorly understood. Even less is known about the development of clinical reasoning over the medical curriculum as a whole. A study by Neufeld et al<sup>1</sup> concluded that the “clinical reasoning process remains relatively constant from medical school entry to practice.” That article noted a paucity of studies examining the evolution of the clinical reasoning process in medical students. In

a more recent study, using a progress testing method,<sup>2–4</sup> Williams et al<sup>5</sup> reported on medical student progress toward achieving the clinical reasoning skills expected of learners by graduation at one institution. As might be expected, years of training had a large effect on clinical reasoning skill performance. Surprisingly, clinical reasoning performance did not accelerate during the third year, when students were intensively immersed in clinical practice. Because this study was conducted at one school, the results may have been unique to that curriculum.

The current study was designed to test whether the pattern of clinical skills acquisition illustrated by Williams et al would be similar at five different medical schools, each with a unique curriculum. The primary purposes of this study were to determine (1) the growth of clinical reasoning across years of training at the five participating schools, (2) whether the improvement was different in the five schools, and (3) whether the rate of improvement aggregated across the five schools was different during the

different years of training (years one, two, and three).

We hypothesized that clinical reasoning skills would improve across years of training at each of the five schools but that, depending on the type of curriculum (e.g., traditional versus problem-based learning), the rates of improvement would differ from school to school. Several articles<sup>6,7</sup> have promoted the idea of teaching clinical skills and basic science knowledge concurrently, noting that this integrated model may be more effective than the traditional approach of teaching basic sciences before proceeding to clinical problems. Thus, we further hypothesized that medical students at those schools with this kind of integrated basic science and clinical curriculum would show higher levels of clinical reasoning skills earlier in their training as compared with students undergoing more traditional forms of instruction. We believed that students in more traditional schools might have less demonstrated clinical reasoning skills in years one and two, but they might “catch up” to their counterparts in integrated curricula with a correspondingly higher

Please see the end of this article for information about the authors.

Correspondence should be addressed to Dr. Klamen, Department of Medical Education, Southern Illinois University School of Medicine, 801 N. Rutledge St., PO Box 19622, Springfield, IL 62794-9622; e-mail: dklamen@siu-med.edu.

*Acad Med.* 2011;86:1148–1154.  
First published online July 21, 2011  
doi: 10.1097/ACM.0b013e31822631b3

rate of clinical reasoning skills growth in year three.

## Method

We developed two tests to measure medical students' clinical reasoning skills. The first included 70 diagnostic pattern recognition (DPR) items, and the second consisted of 69 clinical data interpretation (CDI) items. Tests of these types have been shown to be reasonable proxies for clinical reasoning.<sup>8-10</sup> Examples of DPR items and CDI items are included in Appendices 1 and 2, respectively.

We designed the DPR test, based on the work of Case and colleagues,<sup>11</sup> to assess students' ability to recognize common patterns of patient signs and symptoms. The CDI test used the script concordance item format of Charlin et al<sup>12</sup> and was designed to test students' ability to interpret the impact of individual items of clinical data on the probability that a given diagnostic hypothesis was correct. Unlike the Charlin approach, we established a correct answer for each item. These methods provide useful educational assessments for medical students as well.<sup>13</sup> D.L.K. and other physicians at the Southern Illinois University (SIU) School of Medicine authored the CDI test items and provided correct answers, which were reviewed for accuracy by other SIU physicians not involved in authoring the items. Details of item generation, scoring, and review, as well as examples of each of these types of questions, have been published elsewhere.<sup>5</sup> We created the version of each test used in this research project by selecting questions with demonstrated desirable characteristics (e.g., item difficulty and discrimination) that had been administered in previous years at SIU alone and had been reviewed and revised by a panel of physicians familiar with working with medical students. The DPR examination covered 20 chief complaints. The CDI examination covered 31 chief complaints. Collectively, the DPR and CDI examinations covered 33 diagnoses.

Five medical schools administered the DPR and CDI examinations to their students in 2008. The five schools participating in this study were, in alphabetical order, Medical College of Georgia, Penn State College of Medicine,

SIU School of Medicine, Vanderbilt School of Medicine, and the University of New Mexico School of Medicine. The institutional review board at each of these institutions approved this study. Three of the medical schools are public, and two are private. Enrollments range from small to large (72–190 per year). Among them, the schools have a variety of curricula and instructional delivery systems, including one with a discipline-based first year and a system-based second year, two that have integrated basic science/clinical preclerkship years, and two that have hybrid problem-based learning curricula.

All students at each school were invited to participate, and participants were grouped according to level of training: medical students just entering medical school (Group 0,  $n = 604$ ), those who had just completed the first year (Group 1,  $n = 600$ ), those who had just completed the second year (Group 2,  $n = 576$ ), and those who had just completed the third year (Group 3,  $n = 614$ ). In all, there were 2,394 participants.

We administered the two paper-and-pencil tests together, and students had 120 minutes to complete both tests. Testing for each group occurred in 2008 at the beginning of the new academic year. For example, students who had just completed their first year of medical school (Group 1) were tested at the orientation to year two of medical school. Students who had completed the fourth year at each school (beginning interns) were not included because they were not available for testing. We told students that their performance would not affect academic progress and that their participation was voluntary. Students were assigned a numeric code to protect their identity.

All answer sheets were mailed to one of the schools (SIU) for scoring, which was overseen by R.G.W. For each test, we computed a percent correct score for each student. Data from all schools and students at all levels were aggregated for further analysis. We computed mean percent correct scores and 95% confidence intervals for student groups from each medical school at each of the four levels of training.

Internal consistency reliability was computed using Cronbach alpha. We performed two-factor (years of training by school) analyses of variance separately

for DPR and for CDI. We also used one-way analyses of variance to determine the effects of years of training separately for each school and to determine the effects of school separately at each level of training. A  $P$  value of  $<.05$  was used to determine statistical significance. We used the Tukey honestly significant difference (HSD) method of multiple comparisons to determine which differences in means were significant when overall analysis of variance results were significant. To determine the growth of student performance from one year to the next within each school, we computed effect sizes ( $d$ ). We computed effect sizes by subtracting the latest year score from the previous year score and dividing by the standard deviation for the previous year. We performed all statistical analyses using SPSS statistical software (SPSS version 16.0, SPSS Inc., Chicago, Illinois).

## Results

Table 1 provides the internal consistency test reliabilities broken down by school and years of training. The internal consistency reliabilities for the DPR and CDI examinations across all schools and all levels of training were 0.94 and 0.82, respectively. Average reliabilities were highest after two years of training (Group 2), suggesting that student knowledge applicable to these examinations was best integrated in the minds of students at that stage of their training. DPR reliabilities tended to be higher than CDI reliabilities during the first two years. The correlation between DPR and CDI performance was 0.58 ( $P < .05$ ).

Table 2 provides 2008 DPR and CDI performance results for medical students at the five participating medical schools at four levels of training. This table includes the mean percentage of correct answers, the standard deviation, the percentage of students who took the test, and the effect sizes for years of training. Tukey HSD multiple comparison results are included in the last row of the table to designate statistically significant performance differences between schools where applicable. Figures 1 and 2 depict these results as line graphs.

Year of training explains 66% of the variation in DPR scores ( $P < .05$ ) and 25% of variation in CDI scores ( $P < .05$ ). The "medical school influence," which

Table 1

**Internal Consistency Reliabilities for Each Test, Each School, and Each Level of Training\* Represented in a 2008 Study of 2,394 Medical Students' Performance on Two Clinical Reasoning Skills Tests at Five Schools**

School	Diagnostic pattern recognition				Clinical data interpretation			
	Group 0	Group 1	Group 2	Group 3	Group 0	Group 1	Group 2	Group 3
A	0.78	0.63	0.76	0.50	0.62	0.62	0.75	0.75
B	0.78	0.81	0.92	0.72	0.71	0.75	0.85	0.76
C	0.71	0.84	0.75	0.54	0.62	0.69	0.78	0.72
D	0.80	0.82	0.73	0.74	0.63	0.77	0.81	0.77
E	0.80	0.72	0.82	0.81	0.70	0.71	0.80	0.81
All schools, mean	0.77	0.76	0.80	0.66	0.66	0.71	0.80	0.76

\* Level of training is indicated by group: Group 0 = zero years of training, Group 1 = one year of training, Group 2 = two years of training, and Group 3 = three years of training.

includes all medical school elements (e.g., student selection plus curriculum plus instructional delivery system plus faculty) accounted for 4% of variation in DPR scores ( $P < .05$ ) and 2% of variation in CDI scores ( $P < .05$ ). The interaction between medical school and years of training, a statistical indicator that patterns of improvement are somewhat different for the different medical schools, accounted for 2.7% of variation in DPR scores ( $P < .05$ ) and 1.4% of variation in CDI scores ( $P < .05$ ).

Table 2 and Figures 1 and 2 show that, as a rule, student performance increased substantially as a result of each year of training. These results also illustrate that the annual performance gains attributable to training at each of the participating medical schools were more similar than different. CDI performance and performance gains were lower than DPR performance and gains. The figures also depict the smaller DPR and CDI performance gains attributable to year three of medical school training across the five schools. To address the possibility of a ceiling effect on the results, we calculated the amount of gain possible based on performance in the prior year and then determined the percentage of that gain actually achieved in the subsequent year. When viewed this way, students attained 33% of possible DPR gain during year one, 43% during year two, and 35% during year three. Likewise they gained 15% of the amount of CDI gain possible during year one, 12% during year two, and 7% during year three.

Tukey HSD multiple comparison results (Table 2) indicate that school-to-school

group differences are most often not significant and, when they are significant, relative position of schools varies from training year to training year.

### Discussion and Conclusions

On average, among all students, DPR and CDI performance ability improved with each additional year of medical training. The CDI was a more difficult test than the DPR. We believe that the CDI was more difficult because a pattern recognition format (such as is found in the DPR test items) is more in line with how students encounter such data in their studies. That is, students more frequently read about a disease and its constellation of findings. They are less likely to learn about how a single piece of data may be helpful in ruling in or out a list of diseases on a differential diagnosis.

The expected rapid acceleration in third-year performance gains due to the intensive clinical training during that year did not occur. This finding was consistent with results reported previously at a single institution.<sup>5</sup> The performance of students in the participating schools is more similar than different, especially after three years of the curriculum.

We observed that students in Group 0, who were just starting medical school, answered nearly half the items correctly, but this is not so surprising when one considers that by chance alone students should score 20% of the CDI items (five options for each question) and 12.5% to 20% of the DPR items (five to eight options for each question) correctly. In addition, students are not blank slates on

entry to medical school. Many have already had extensive mentoring, shadowing, and volunteering experiences with physicians and patients. At entry, school-to-school differences amounted to an average of four items on the DPR and the CDI test. The largest school-to-school differences after years one and two amounted to a difference of only five correct answers. After year three, school-to-school differences were two correct answers on the DPR exam and no appreciable school-to-school differences in CDI performance. The DPR standard deviation decreased in year three, indicating that students' performance became more homogenous. This was not true for CDI. In summary, DPR and CDI performance differences among school cohorts were relatively small at the time of admission, as would be expected, and were also small by the end of year three.

The lack of accelerated progress in the third year is puzzling. One explanation for this finding is a ceiling effect that makes further improvement difficult after a certain level of achievement. Our calculations regarding percentage of possible gain indicate that ceiling effect may be a partial explanation for the findings—especially the DPR findings. Although it is also possible that portions of the test may have been too advanced for the level of learner, especially in the third year, all questions were created specifically for medical students by the test creators, and they were reviewed by other physicians also familiar with teaching medical students. Another plausible explanation for the lack of more rapid acceleration in performance gain

**Table 2**  
**Diagnostic Pattern Recognition (DPR) and Clinical Data Interpretation (CDI) Scores**  
**for 2,394 Medical Students from Five Schools at Four Levels of Training,\* 2008**

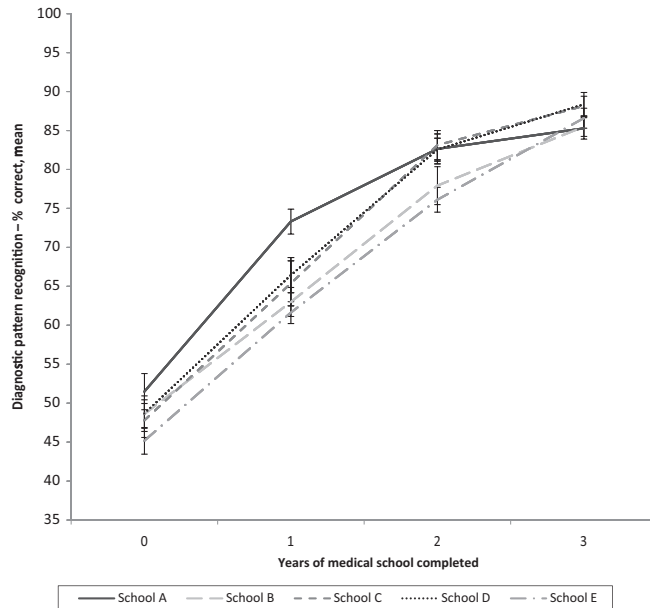
School	DPR				CDI					
	Group 0	Group 1	Group 2	Group 3	All groups, mean	Group 0	Group 1	Group 2	Group 3	All groups, mean
<b>A</b>										
% Correct, mean	51.49	73.32	82.64	85.28	71.74	39.83	54.84	60.05	58.65	52.56
Standard deviation (SD)	11.14	7.60	8.29	5.55	16.00	9.12	9.10	10.88	10.49	12.81
% Who took test <sup>†</sup>	100	100	92.6	100	98.2	100	100	92.6	100	98.2
Effect size for training year, d	N/A	1.96	1.23	0.32	0.96	N/A	1.65	0.57	-0.13	0.78
<b>B</b>										
% Correct, mean	48.61	63.00	77.93	85.50	68.41	41.41	47.21	53.50	57.84	49.81
SD	11.02	11.41	14.91	7.15	18.17	10.47	11.14	14.29	11.20	13.37
% Who took test <sup>†</sup>	97	97	98	98	97.5	97	97	98	98	97.5
Effect size for training year, d	N/A	1.31	1.31	0.51	1.04	N/A	0.55	0.56	0.30	0.47
<b>C</b>										
% Correct, mean	47.79	65.38	83.12	88.15	73.63	44.12	51.11	56.58	57.65	53.37
SD	9.61	12.30	7.94	5.20	17.02	9.10	10.05	11.85	10.31	11.38
% Who took test <sup>†</sup>	100	100	100	99	99.8	100	100	100	99	99.8
Effect size for training year, d	N/A	1.83	1.44	0.63	1.3	N/A	0.77	0.54	0.09	0.47
<b>D</b>										
% Correct, mean	48.66	66.44	82.50	88.37	70.79	41.48	51.33	56.85	59.60	51.99
SD	11.61	11.51	7.80	6.77	18.10	9.23	11.47	12.56	11.25	13.12
% Who took test <sup>†</sup>	97	99	100	83	94.8	97	99	100	83	94.8
Effect size for training year, d	N/A	1.53	1.40	0.75	1.23	N/A	1.07	0.48	0.22	0.59
<b>E</b>										
% Correct, mean	45.09	61.66	76.12	86.57	66.16	38.33	47.51	53.67	57.76	48.76
SD	11.75	9.64	10.50	8.06	18.57	10.10	10.42	12.25	12.02	13.33
% Who took test <sup>†</sup>	98	97	96	83	93.5	98	97	96	83	93.5
Effect size for training year, d	N/A	1.41	1.50	1	1.30	N/A	0.91	0.59	0.33	0.61
<b>All schools, mean</b>										
% Correct, mean	47.83	64.98	79.53	86.62	68.84	40.55	49.61	55.43	58.18	50.55
SD	11.38	11.18	11.23	7.05	18.14	9.93	10.89	12.82	11.24	13.11
No. (%) who took test	604 (98.4)	600 (98.6)	576 (97.3)	614 (92.6)	2394 (96.7)	603 (98.4)	600 (98.6)	575 (97.3)	613 (92.6)	2391 (96.7)
Effect size for training year, d	N/A	1.61	1.38	0.51	1.17	N/A	0.99	0.55	0.16	0.57

(Continues)

Table 2  
(Continued)

School	DPR					CDI				
	Group 0	Group 1	Group 2	Group 3	All groups, mean	Group 0	Group 1	Group 2	Group 3	All groups, mean
Statistically significant between-school differences*	A > E	A > B, E; D > E	A, C, D > B, E	E > A	N/A	C, D > A, E	A > B, E; D > C	A > B, E	None	N/A

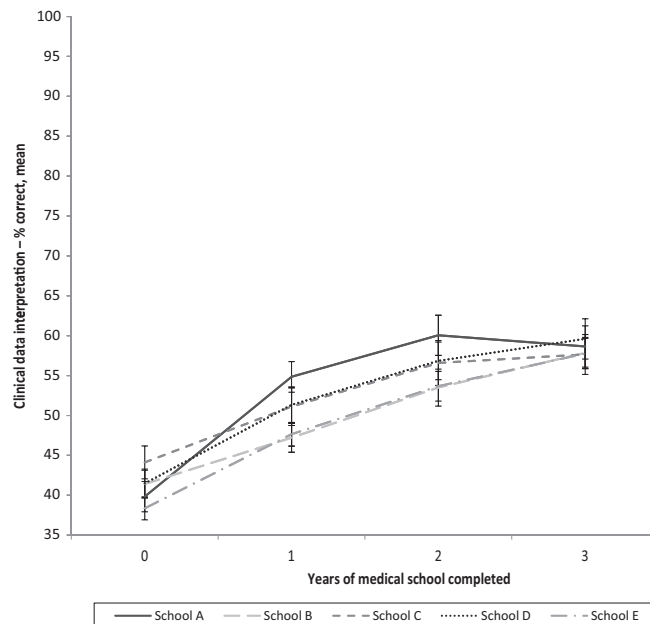
\* Level of training is indicated by group: Group 0 = zero years of training, Group 1 = one year of training, Group 2 = two years of training, and Group 3 = three years of training.  
 † Raw numbers of participants have been omitted to maintain confidentiality and protect the identities of the five schools.  
 ‡ Determined by Tukey honestly significant difference, multiple comparisons.



**Figure 1** Mean performance (with 95% confidence intervals) on a 2008 test measuring diagnostic pattern recognition for 2,394 students with zero, one, two, or three years of training in five medical schools.

during the third year is that acquisition of diagnostic reasoning ability is not the primary goal of clinical training during the third year. Students may be concentrating on other patient-care-related activities (e.g., learning to function in clinical environments, patient management, choosing a specialty). Another plausible though unexplored explanation is that student conceptions of disease presentations

(based on classic presentation of diseases in years one and two) may be disrupted by the complexity and ambiguity of real patient presentations. Students, who have spent the third year being confronted by patients who do not look like they look in a textbook, may have been more hesitant and unsure when answering clinical reasoning questions on the progress test.



**Figure 2** Mean performance (with 95% confidence intervals) on a 2008 test measuring clinical data interpretation for students with zero, one, two, or three years of training in five medical schools.

We hypothesized that schools with early, integrated clinical and basic science teaching would produce sustained superior gains in clinical reasoning, but this did not occur. These findings are consistent with those reported by others<sup>14,15</sup> regarding the impact of medical school selection processes, curricula, and curriculum delivery systems on United States Medical Licensing Examination test performance. The variations in North American medical school selection policies and procedures, curricula, and curriculum delivery systems seem to result in relatively small differences in measured student performance capabilities at the end of undergraduate medical education training. Perceived inadequacies in medical education result in frequent calls for curriculum change, and schools seem to go through regular cycles of reinvention. Although we do not wish to discourage consideration and implementation of curricular changes, we believe that our results and those of Hecker and Violato<sup>14</sup> should be considered before investing time and effort in curriculum revision to improve knowledge or diagnostic reasoning skills. Although changes in curricula or delivery systems might influence other features of student performance not measured by these tests, changes in student knowledge and diagnostic ability are likely to be modest.

National data regarding how practicing primary care physicians and residents perform on these tests would add context for interpreting the results presented here. Williams et al<sup>5</sup> provided some data about DPR and CDI performance of practicing physicians and residents, but the data do not adequately represent the population of North American primary care physicians and residents. Such data would allow comparisons of medical student performance to practicing physician performance.

This study has limitations worth noting. The cross-sectional design of the study (as opposed to a longitudinal follow-up of the same cohort of students over time) makes it difficult to conclude that the gains seen were the result of acquisition of clinical reasoning skills rather than cohort differences attributable to differences in student selection during the years involved. However, it is unlikely

that all five participating schools would have experienced similar change in student selection, a circumstance which would have been necessary to attribute these observed differences to selection. We plan to report longitudinal results for the participating schools in the future.

In summary, our results demonstrate that medical students achieve substantial gains in clinical reasoning ability with each year of medical school training and experience. However, these students do not seem to experience larger gains in clinical reasoning ability after the clinically intensive third year of medical school. Further, the gains observed are similar across the spectrum of medical schools that participated in this research project.

**Dr. Williams** is J. Roland Folse Professor of Surgical Education Emeritus, Department of Surgery, Southern Illinois University School of Medicine, Springfield, Illinois.

**Dr. Klamen** is associate dean for education and curriculum and professor, Department of Medical Education, Southern Illinois University School of Medicine, Springfield, Illinois.

**Dr. White** is associate dean for faculty development and educational achievement, Medical College of Georgia School of Medicine, Augusta, Georgia.

**Dr. Petrusa** is director, Office for Teaching and Learning in Medicine, director, Center for Outcomes Research in Education, and professor of medical education and administration, Vanderbilt School of Medicine, Nashville, Tennessee.

**Dr. Fincher** is professor, Department of Medicine, and vice dean for academic affairs and codirector, Education Discovery Institute, Medical College of Georgia School of Medicine, Augusta, Georgia.

**Dr. Whitfield** is associate dean for preclinical curriculum, Penn State College of Medicine, Hershey, Pennsylvania.

**Dr. Shatzer** is director, Center for Experiential Learning and Assessment, and associate professor of medical education and administration, Vanderbilt School of Medicine, Nashville, Tennessee.

**Dr. McCarty** is associate professor, Department of Psychiatry, and assistant dean, Simulation and Assessment, University of New Mexico School of Medicine, Albuquerque, New Mexico.

**Dr. Miller** is senior associate dean for health sciences education, Vanderbilt School of Medicine, Nashville, Tennessee.

**Acknowledgments:** The authors would like to acknowledge the ongoing work of the North American Progress Testing in Medical Education Group: Medical College of Georgia, Pennsylvania State University, Southern Illinois University, Texas Tech at El Paso, University of New Mexico, Vanderbilt University, and the University of Minnesota. They would also like to thank Randall Robbs at Southern Illinois University School of Medicine for scoring the tests, and Robert

Gowin, also at Southern Illinois University, for preparing the test score reports.

**Funding/Support:** None.

**Other disclosures:** None.

**Ethical approval:** Ethical approval has been granted by the IRB committees at each of the participating institutions: Medical College of Georgia, Penn State College of Medicine, Southern Illinois University School of Medicine, Vanderbilt School of Medicine, and the University of New Mexico School of Medicine.

## References

- 1 Neufeld VR, Norman GR, Feightner JW, Barrows HS. Clinical problem solving by medical students: A cross-sectional and longitudinal analysis. *Med Educ.* 1981;15:315–322.
- 2 Van der Vleuten C, Verwijnen GM, Wijnen W. Fifteen years of experience with progress testing in a problem-based curriculum. *Med Teach.* 1996;18:103–109.
- 3 Blake JM, Norman GR, Keane DR, Mueller CB, Cunningham J, Didyk N. Introducing progress testing in McMaster University's problem-based medical curriculum: Psychometric properties and effect on learning. *Acad Med.* 1996;71:1002–1007. [http://journals.lww.com/academicmedicine/Abstract/1996/09000/Introducing\\_progress\\_testing\\_in\\_McMaster.16.aspx](http://journals.lww.com/academicmedicine/Abstract/1996/09000/Introducing_progress_testing_in_McMaster.16.aspx). Accessed May 25, 2011.
- 4 Arnold L, Willoughby TL. The quarterly profile examination. *Acad Med.* 1990;65:515–516. [http://journals.lww.com/academicmedicine/Abstract/1990/08000/The\\_quarterly\\_profile\\_examination.5.aspx](http://journals.lww.com/academicmedicine/Abstract/1990/08000/The_quarterly_profile_examination.5.aspx). Accessed May 25, 2011.
- 5 Williams RG, Klamen DL, Hoffman RM. Medical student acquisition of clinical working knowledge. *Teach Learn Med.* 2008;20:5–10.
- 6 Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ.* 2004;39:98–106.
- 7 Charlin B, Tardif J, Boshuizen HPA. Scripts and medical diagnostic knowledge: Theory and applications for clinical reasoning instruction and research. *Acad Med.* 2000;75:182–190. [http://journals.lww.com/academicmedicine/Fulltext/2000/02000/Scripts\\_and\\_Medical\\_Diagnostic\\_Knowledge\\_Theory.20.aspx](http://journals.lww.com/academicmedicine/Fulltext/2000/02000/Scripts_and_Medical_Diagnostic_Knowledge_Theory.20.aspx). Accessed May 25, 2011.
- 8 Case SM, Swanson DB. Extended-matching items: A practical alternative to free-response questions. *Teach Learn Med.* 1993;5:107–115.
- 9 Carriere B, Gagnon R, Charlin B, Downing S, Bordage G. Assessing clinical reasoning in pediatric emergency medicine: Validity evidence for a script concordance test. *Ann Emerg Med.* 2009;53:647–652.
- 10 Boutros M, Nouh T, Reid S, et al. The script concordance test as a measure of clinical reasoning: A cross-Canada validation study. *Med Educ.* 2010;44(suppl 2):1.
- 11 Case SM, Swanson DB, Stillman PL. Evaluating diagnostic pattern recognition: The psychometric characteristics of a new item format. *Proc Annu Conf Res Med Educ.* 1988;27:3–8.

12 Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C. The script concordance test: A tool to assess the reflective clinician. *Teach Learn Med.* 2000;12:189–195.

13 Bland AC, Kreiter CD, Gordon JA. The psychometric properties of five scoring methods applied to the script concordance test. *Acad Med.* 2005;80:395–399. [http://journals.lww.com/academicmedicine/Fulltext/2005/04000/The\\_Psychometric\\_Properties\\_of\\_Five\\_Scoring.19.aspx](http://journals.lww.com/academicmedicine/Fulltext/2005/04000/The_Psychometric_Properties_of_Five_Scoring.19.aspx). Accessed May 25, 2011.

14 Hecker K, Violato C. How much do differences in medical schools influence student performance? A longitudinal study employing hierarchical linear modeling. *Teach Learn Med.* 2008;20:104–113.

15 Ripkey DR, Swanson DB, Case SM. School-to-school differences in Step 1 performance as a function of curriculum type and use of Step 1 in promotion/graduation requirements. *Acad Med.* 1998;73(10 suppl):S16–S18. [http://journals.lww.com/academicmedicine/Citation/1998/10000/ALTERNATIVE\\_APPROACHES\\_TO\\_PROGRAM\\_EVALUATION\\_32.aspx](http://journals.lww.com/academicmedicine/Citation/1998/10000/ALTERNATIVE_APPROACHES_TO_PROGRAM_EVALUATION_32.aspx). Accessed May 25, 2011.

## Appendix 1

### Examples of Questions and Answers on a Diagnostic Pattern Recognition Test Administered to 2,394 Medical Students With Zero, One, Two, or Three Years of Training at Five Schools, 2008

#### Mental Status Change: Diagnostic Possibilities

- A. Dementia
- B. Delirium
- C. Pseudodementia secondary to major depression
- D. Chronic psychosis
- E. Mental retardation/developmental disabilities
- F. Intracranial process
- G. Intoxication

For each patient, select the most likely diagnosis from the list above. Each option may be used once, more than once, or not at all.

     B 1) A 62-year-old woman with no prior psychiatric history is hospitalized for pneumonia. She has a history of diabetes mellitus, hypertension, coronary artery disease, and degenerative joint disease and she is on multiple medications. Her second night in the hospital she becomes agitated, yelling and screaming that she sees demons “trying to tear her apart.” She is not oriented to date, believing it to be 1908. The patient is put in posey restraints for her own protection. The symptoms resolve by morning when her family visits.

     F 2) A 47-year-old man is admitted to the hospital, status post myocardial infarction. The first night in the ICU he has a sudden onset of unintelligible speech, and disorientation without agitation. This is accompanied by right arm and leg weakness.

     A 3) A 78-year-old man with no prior history of psychiatric treatment is brought in to the doctor by his wife. He has a history of hypertension and hypercholesterolemia. Over the past 6–7 months the wife notes that the patient has had increasing memory loss, episodes of irritability and agitation, increasingly garbled speech, and has been unable to plan or organize.

## Appendix 2

### Examples of Questions and Answers on a Clinical Data Interpretation Test Administered to 2,394 Medical Students With Zero, One, Two, or Three Years of Training at Five Schools, 2008

#### Skin Lesions

*Clinical Vignette: A 58-year-old female presents with a chief complaint of a skin rash.*

Item number	If you were thinking of:	And then the patient reports or you find on clinical or laboratory examination:	This diagnostic hypothesis becomes:					Correct answer
134	Drug eruption	Recent changes in blood pressure medication	A	B	C	D	E	D
135	Urticaria	Transient erythematous patches	A	B	C	D	E	E
136	Herpes zoster	Asymptomatic blistering rash on chest	A	B	C	D	E	A

A = the hypothesis is almost eliminated, B = the hypothesis becomes less probable, C = the information has virtually no effect on the hypothesis, D = the hypothesis becomes more probable, E = the hypothesis is almost certainly correct.