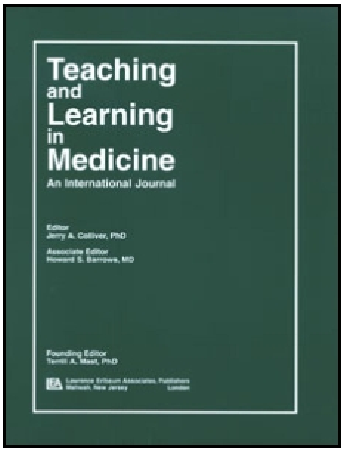


**Teaching and Learning in Medicine**
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775648180

# Medical Student Acquisition of Clinical Working Knowledge

Reed G. Williams [a]; Debra L. Klamen [b]; Rebecca M. Hoffman [c]
[a] Department of Surgery, Southern Illinois University School of Medicine, Springfield, Illinois, USA [b] Department of Medical Education, Southern Illinois University School of Medicine, Springfield, Illinois, USA [c] Department of Family and Community Medicine, Southern Illinois University School of Medicine, Springfield, Illinois, USA

Online Publication Date: 01 January 2008

**To cite this Article** Williams, Reed G., Klamen, Debra L. and Hoffman, Rebecca M.(2008)'Medical Student Acquisition of Clinical Working Knowledge',Teaching and Learning in Medicine,20:1,5 — 10

**To link to this Article: DOI:** 10.1080/10401330701542552

**URL:** http://dx.doi.org/10.1080/10401330701542552

*Teaching and Learning in Medicine*, 20(1), 5–10
Copyright © 2008, Taylor & Francis Group, LLC
ISSN: 1040-1334 print / 1532-8015 online
DOI: 10.1080/10401330701542552

Routledge
Taylor & Francis Group



**APPLIED RESEARCH**

# Medical Student Acquisition of Clinical Working Knowledge

**Reed G. Williams**
*Department of Surgery, Southern Illinois University School of Medicine, Springfield, Illinois, USA*

**Debra L. Klamen**
*Department of Medical Education, Southern Illinois University School of Medicine, Springfield, Illinois, USA*

**Rebecca M. Hoffman**
*Department of Family and Community Medicine, Southern Illinois University School of Medicine, Springfield, Illinois USA*

***Background:*** **Working knowledge of physicians manifests as a combination of diagnostic pattern recognition and clinical data interpretation (analytic fact checking).** ***Purpose:*** **The purpose was to study medical student acquisition of these abilities as a function of years of medical training/experience.** ***Methods:*** **A cross-sectional study involving students who had completed 0, 1, 2, and 3 years of medical school. All students at all levels of training took the same tests of diagnostic pattern recognition and clinical data interpretation. Percent correct scores were calculated and used to estimate learning curves. A cohort of family physicians also took the test to provide a benchmark.** ***Results:*** **Student diagnostic pattern recognition and clinical data interpretation ability demonstrated a steady upward growth curve but leveled off in Year 3. Diagnostic pattern recognition performance was consistently higher than clinical data interpretation performance. The rate of diagnostic performance gain with training and experience was also higher.** ***Conclusions:*** **Medical students acquired diagnostic pattern recognition ability and all years of medical training contributed. The rate of clinical data interpretation performance improvement was slower, and the absolute performance level was lower. What was surprising was the lower rate of improvement in diagnostic pattern recognition and clinical data interpretation performance for students during their 1st year of clinical training. Students' understanding of findings and their relationships to disease processes may be affected by their limited patient experience.**

In a review of current research on clinical reasoning, Eva[1] concludes that the working knowledge of physicians, medical students, and residents manifests itself as a combination of global, pattern detecting, and analytic fact-checking ability. Clinical pattern detecting (which we will call diagnostic pattern recognition [DPR]) is akin to the instant recognition that allows you to identify a cat as a cat because you have seen a hundred of them before and know their distinctive features. It is seemingly unconscious, automatic, and nonanalytic. Regehr and Norman[2] point out that this is similar to the experienced clinician saying that a patient likely has asthma after just the briefest of interactions. If asked to defend this assertion, Regehr and Norman note that the experienced clinician will likely say "because patients who are like this usually have asthma" (p. 998). Eva likens it to knowing that 120 divided by 10 equals 12 without having to do the math.

Analytic fact-checking ability, on the other hand, is a logical, systematic, conscious, carefully controlled process of considering individual pieces of clinical information and their fit with diagnostic possibilities under consideration. Eva notes that models of clinical reasoning that use Bayes's theorem or regression analyses best represent this systematic, conscious, analytic process that we will call clinical data interpretation (CDI). Mamede and Schmidt[3] provide a supportive view of this process. They propose that the thinking activities of reflective practitioners have three components. The first component, deliberate induction, is described as the generation of possible alternative explanations for the patient's problem. The second step, deliberate deduction, is determining the observable expression of these clinical hypotheses, that is, the signs and symptoms the patient will likely have if any of these hypotheses are true. The third step, hypothesis verification, is the testing of observed patient findings against the expected findings for each hypothesis.

Creating a clinically useful knowledge base is usually a multistage process. Medical students initially acquire and store new information to support passing anticipated basic science tests.[2] To be clinically useful, it must be reorganized into a network of concepts and examples that support clinical functions. Such a

Final revision received May 26, 2007.
Address correspondence to Reed G. Williams, Southern Illinois University School of Medicine, Department of Surgery, 800 North Rutledge Street, P.O. Box 19638, Springfield, IL 62794-9638. E-mail: Rwilliams@siumed.edu

knowledge base allows a physician to recognize features from a new patient case and to activate related information, concepts, or examples from memory. This information in turn activates other relevant information, concepts, or examples. Concept maps[4] provide representations of the nature of these networks of concepts and examples and their relationships. This knowledge reorganization process is thought to occur during the intensive clinical portion of medical school training (Year 3 in the United States). Faculty members speculate that this reorganization process occurs at different rates for different students.

For this study we built a test that directly measures these two applications of clinically useful working knowledge (DPR and CDI). Our purpose was to study the acquisition of these two abilities as a function of medical training and experience and to report the learning curves for each. Our specific research objectives were to determine whether the learning curves for the two abilities are similar or different and to determine the relationship between activities in various phases of the curriculum and the learning curve for students. This was accomplished by directly measuring changes in performance for cohorts of students at different levels of training. These examinations are viewed as a complement to, but not a replacement for, current medical school course examinations. They test the second phase of knowledge organization and integration (clinical working knowledge), whereas course examinations test initial knowledge acquisition and integration.

## METHODS

### Participants

Participants were medical students enrolled in one medical school (73 medical students who had just entered medical school [Group 0], those who had just completed the 1st year [Group 1, $n = 73$], 2nd year [Group 2, $n = 69$], and 3rd year [Group 3, $n = 70$] of medical training). In North America, medical school typically takes 4 years, with the 3rd year being the first intensive opportunity for clinical practice in the various specialties of medicine. Students who had completed the 4th year were not included in this study for two reasons. First, the experiences and training for 4th-year students is idiosyncratic. Second, students who had completed the 4th year of medical school (graduates) were not reliably available for testing.

A group of 20 practicing family physicians, including family medicine residency faculty and community physicians, also took the test to provide benchmark scores regarding the average performance of practicing physicians for comparison purposes. These physicians had been in practice an average of 17 years (range = 2–30 years).

Students who had just entered this medical school were told that they would be tested in a similar manner each year and that their progress in developing DPR and CDI ability would be taken into consideration when making promotion decisions in their 4th year. The other students were told that they were required to take the progress examination but that their performance would not influence promotion. These students were urged to do their best as they would be given performance test results that would allow them to compare their performance to that of students at all levels of training, and their performance would help establish norms for entering students and subsequent classes for whom performance would count. All participants appeared to be receptive to and engaged in the task throughout.

### Assessment Instrument

The 111-item examination was composed of 59 DPR items and 52 CDI items. The DPR test was based on the work of Case and colleagues[5] and was designed to assess student recognition of common patterns of patient signs and symptoms. Our operationalization of DPR ability may not correspond perfectly with the conception described earlier, because, in a paper-and-pencil test, test takers could use an analytic strategy for considering individual data elements when formulating an answer to the question. We believe, however, that these DPR items are the best available paper-and-pencil measures of DPR ability. The CDI test used the script concordance item format of Charlin and his associates[6] and was designed to test students' ability to interpret the impact of individual bits of clinical data on diagnostic possibilities. Unlike the Charlin approach, we established correct answers for each item in the traditional multiple-choice manner. The correct answer was provided by the item author and was reviewed by other physicians. Examples of each of these two types of examination questions follow:

Diagnostic pattern recognition question:

**Panic and Anxiety**

Diagnostic Possibilities:

A: Acute medical condition

B: Anxiety secondary to a general medical condition

C: Substance-induced anxiety disorder

D: Intoxication/withdrawal state

E: Panic disorder

F: Generalized anxiety disorder

G: Somatoform Disorder

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.

___1) A 23-year-old college student comes to the Emergency Department with the complaint of "feeling as if I am going to die." The patient complains of intense anxiety, palpitations, parethesias and feeling as if she is smothering. This has happened twice before in the past week, each episode lasting about 10 minutes. Toxicology screen is negative and physical exam is within normal limits.

___2 A 64-year-old man is brought to the Emergency Department after complaining of intense anxiety, rapid heart beat, and chest pain. The patient has a prior history of coronary artery disease and angina.

Clinical Data Interpretation Items

**Coughing and Wheezing**

Clinical Vignette: An adult patient comes in complaining of coughing

| Item Number | If you were thinking of: | And then the patient or you find on clinical or laboratory examination: | This diagnostic hypothesis becomes: | | | | |
|---|---|---|---|---|---|---|---|
| 65 | Gastro-esophageal reflux disease | No history of heartburn | A | B | C | D | E |
| 66 | Chronic obstructive pulmonary disease | White blood cell count normal | A | B | C | D | E |
| 67 | Thyroid mass | Voice changes, hoarseness | A | B | C | D | E |

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.

The examination covered 17 chief complaints and 6-9 underlying diagnoses for each. The CDI and DPR items for each chief complaint were created by the same physician author and were selected to provide broad coverage of the chief complaints and underlying diagnoses.



### Testing Method

Testing occurred at the beginning of the academic year. The test was administered during that year's orientation for Groups 0, 1 and 2. It was administered for most Group 3 students at the end of the school's required clinical comprehensive examination. This examination occurred at the beginning of the 4th year of medical school. Group 3 students were allowed to take the progress examination at a different time within a 2-week window if they preferred. Seventeen students chose this option. Family physicians who agreed to take the test took the same exam at their convenience over a 3-month period.

### Test Scoring and Data Analysis

Although administered together, the DPR and CDI tests were considered separate tests for analysis purposes.

Statistical item analyses were performed to identify poorly performing items. These items subsequently were reviewed by a panel of physicians. Based on this review, some items were eliminated from scoring and others had the correct answer changed. Participants' tests were then rescored using the revised scoring protocol.

For each test, percent correct scores were computed by dividing the number of items answered correctly by the best possible score (items correct).

Internal consistency reliability was computed using the formula developed by Cronbach (Cronbach's alpha). This measure was used to determine whether the tests we created included an adequate sample of items to estimate each aspect of clinical performance ability.

Means and confidence intervals (CIs) were computed to determine differences in DPR and CDI performance for student groups at the four levels of training and experience. Linear regression analysis was used to determine the nature and extent of the relationship between training/experience and DPR/CDI performance. Both a linear and quadratic model were used to determine whether the relationship was best described as a linear one or whether there was a curvilinear component to the relationship (which would be present if the relationship looked like a typical asymptotic learning curve function).

Box and whisker plots were also created to provide more information about the distribution of scores. In box and whisker plots, the dark horizontal line in the box represents the median score. The box itself incorporates the middle half of the score distribution (the 25th to the 75th percentile). The ends of the lines (whiskers) represent the 10th and 90th percentiles, and the circles represent the few extreme scores. These box and whisker plots illustrate the degree to which the distributions of scores for the various training groups overlap.

A one-way analysis of variance (ANOVA) was used to determine whether there were statistically significant differences in group DPR and CDI performances that could be attributed to years of training. When the one-way ANOVA was significant, Tukey's honestly significant difference tests were used to determine which of the training groups differed in performance compared to others.

A repeated measures ANOVA was performed with DPR and CDI performance as the within-group measure and years of training as the between-groups measure. The purpose for this analysis was to determine whether the rate of performance growth was different for DPR and CDI. This relationship is represented by the Dependent Measure × Years of Training interaction term in the repeated measures analysis.

Paired sample $t$ tests were used to compare DPR and CDI performance at each level of medical training. Effect sizes for these comparisons were computed using d. d determines the magnitude of the performance differences between two measures in pooled standard deviation terms. d is defined as the mean of DPR performance minus the mean of CDI performance divided by the pooled standard deviation of the two measures.

All statistical analyses were computed using SPSS statistical software (SPSS version 13.0, SPSS, Inc., Chicago). This research was approved by the local Institutional Review Board.

## RESULTS

Table 1 provides the internal consistency reliabilities for the four student cohorts and for all students combined. Reliabilities for individual cohorts of students are all less than the 0.80 standard advocated for high-stakes examinations, indicating that larger numbers of items are needed to optimally sample these two domains of clinical working knowledge. DPR reliability

TABLE 1
Internal consistency reliability by student cohort

| Test | Beginning Students | Students who Completed Year 1 | Students who Completed Year 2 | Students who Completed Year 3 | All Students |
|---|---|---|---|---|---|
| Diagnostic Pattern Recognition | 0.70 | 0.73 | 0.63 | 0.58 | 0.92 |
| Clinical Data Interpretation | 0.63 | 0.57 | 0.65 | 0.69 | 0.75 |

for all students was 0.92, which is comparable to that observed in such examinations as the United States Medical Licensure examination (USMLE). DPR reliabilities are higher than CDI reliabilities for beginning students and those who have completed Year 1. They are lower than CDI reliabilities for students who have completed Years 2 and 3.

Table 2 provides the mean percent correct score, the standard deviation, and the 95% CIs for each performance measure (DPR and CDI) for each cohort of students. Table 2 also provides the effect size measures comparing the magnitude of the differences in DPR and CDI performance at each level of training. The effect size represents the difference in the two performance measure scores in pooled standard deviation units at each level of training.

Inspection of Table 2 and Figure 1 indicates that both DPR and CDI performance increase rapidly and linearly from the beginning of medical school to the completion of Year 2. The rate of performance improvement then tapers off through the 1st year of clinical training (Year 3). By the end of the 3rd year of medical training, average DPR percent correct performance has increased by 87%, whereas growth in average CDI performance has increased by only 44%.

Curve fitting analyses documented that the observed quadratic component of each relationship was statistically significant ($p < .000$ for DPR and $p = .047$ for CDI) but small by comparison to the linear components. The linear component of the relationship between years of training and DPR alone accounted for 75% of the score variance. When the quadratic component was added the model accounts for 77% of the total variance. The linear component of the relationship between years of training and CDI performance accounts for 31% of CDI score variance. Adding the quadratic component resulted in a model, which accounted for 32% of the variance in CDI scores.

A one-way ANOVA demonstrated that student increases in DPR and CDI performance at different levels of training/experience were statistically significant, $F_{\text{DPR}} = 323.62$, $p < .000$; $F_{CDI} = 44.41$, $p < .000$. Post hoc analyses using Tukey's honestly significant difference test indicated that all group means are significantly different from each other except the CDI performance differences between Groups 2 and 3. Partial eta-squared analyses indicated that the level of training/experience accounted for 78% of variance in DPR performance and 32% of the variance in CDI performance. Inspection of the CIs included in Table 2 shows that the CDI lower bound 95% CI for the students who have completed year 3 and the upper bound 95% CI for the students who have completed Year 2 overlap. That is the only case where the 95% CIs for student groups have any overlap.

Rates of student improvement in DPR and CDI performance across years of experience are different. This difference is statistically significant as indicated by the interaction term in the repeated measures ANOVA ($F = 52.95$, $p < .05$). Inspection of the effect sizes in Table 2 and the plots in Figure 1 document the magnitude and nature of these differences. DPR and CDI performance for students entering medical school were similar but significantly different. DPR performance accelerated at a more rapid rate than did CDI performance over the ensuing years of training and experience. Inspection of Table 2 and Figure 1 also indicates that the performance of medical students on these tests begins to approach that of practicing family physicians.

TABLE 2
Mean diagnostic pattern recognition and clinical data interpretation percent correct score (standard deviation) and [95% confidence intervals] by years of medical training

| | Years of Medical Training | | | | |
|---|---|---|---|---|---|
| Test | 0[a] | 1[b] | 2[c] | 3[d] | Family Medicine Physicians[e] |
| DPR | 47 (10.00) [45–50] | 64 (10.30) [62–67] | 81 (7.26) [80–83] | 88 (5.85) [87–89] | 91 (4.95) [89–93] |
| CDI | 41 (10.30) [39–43] | 49 (9.56) [47–51] | 56 (10.50) [54–59] | 59 (10.89) [57–62] | 66 (9.15) [62–70] |
| d | .60* | 1.52* | 2.79* | 3.32* | 3.40* |

[a]$n = 73$. [b]$n = 73$. [c]$n = 69$. [d]$n = 70$. [e]$n = 20$. DPR = diagnostic pattern recognition; CDI = clinical data interpretation; d = effect size comparing DPR and CDI performance.

*Paired sample $t$ test results significant at $p < .05$.

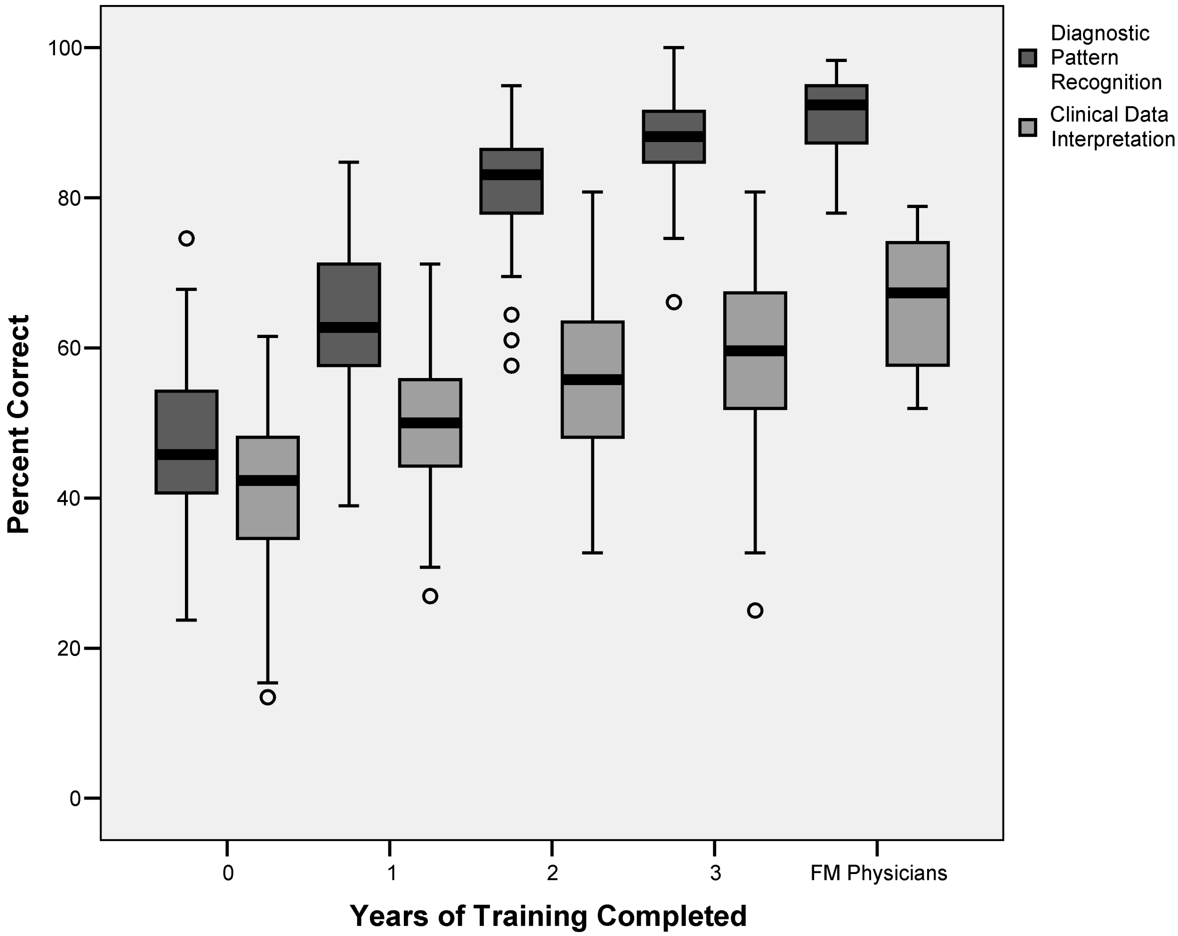


FIG. 1. Box and whiskers depiction of the diagnostic pattern recognition and clinical data interpretation score distributions. *Note*. Horizontal line indicates the median score; ends of each box depict the 25th and 75th percentiles; ends of the vertical lines (whiskers) depict the 10th and 90th percentiles; circles depict extreme scores.

## CONCLUSIONS

Internal consistency reliabilities indicated that longer DPR and CDI tests are required to achieve the 0.80 reliability level needed if these test results are to be used for individual progress decisions and to test the two domains of interest optimally.

Years of training and experience have a large effect on both DPR and CDI performance. The rate of DPR performance gain is greater than that for CDI performance. By the end of the 3rd year of medical training, average DPR performance had increased by 87%, whereas CDI performance had increased by only 44%.

What is most dramatic is the absolute difference in DPR and CDI performance at each training level. DPR performance is higher than CDI performance when students enter medical school and it stays considerably higher throughout medical school. The consistently higher DPR performance is likely a function of two contributing factors: (a) the nature of teaching and learning occurring in medical school and (b) the inherent difficulty of the tasks. As regards the nature of teaching and learning, students see similar DPR items in basic science and clinical textbooks and test items and on medical programs on commercial television networks. The phrasing of these questions is familiar as is the linear thought process required. The same is not true for CDI items. Although students learn this information during teaching encounters using clinical interactions, it is rarely presented in this form in textbooks and the test format is unfamiliar. Thus, the demands that CDI items place on medical students for reorganizing information may be greater than those of DPR items.

As regards the lower performance on the CDI test relative to the DPR test, these differences are relatively small for entering students and increase with training. This suggests that the differences are mainly attributable to training rather than item features unrelated to medical knowledge. We think there are two major factors at work. First is the ability to compensate for knowledge deficits. In DPR items, the student may not be aware of the relationship between a specific symptom or sign and a disease but will be able to compensate for this deficit and answer the question correctly if the relationship with the remaining symptoms is known. In CDI items if one does not know the predictive probability of the target symptom or sign, it is not possible to answer the question correctly based on other known information. The second factor that contributes to the inherent difficulty of CDI items is that the student must have specific knowledge about the target finding and the specific disorder *and* must know the link between the two. If that link has not been made, the student will be unable to arrive at the correct conclusion.

A third possible explanation for lower CDI performance is an artifact of the test design. The CDI items were the last items on the test. This could have differentially affected CDI test performance if examinees ran out of time or became fatigued. The only pertinent evidence we have suggests this is a minor factor if it is a factor at all. There were 14 instances where a student did not answer a CDI question, and these were mainly near the end of the test. There was only 1 instance where a student did not answer a DPR question. However the number of nonresponses to CDI questions was very small compared to the 14,806 total responses to CDI questions.

What is surprising is the decreased rate of increase in DPR and especially CDI performance for those students who have just completed the 3rd year of medical school. This training year is an intensive and systematic immersion in clinical practice across the various medical specialties. One would expect to see an associated spike in DPR and CDI performance improvement. This did not occur. In fact, the rate of performance improvement is less in the 3rd year than in other years. Regarding DPR performance, this lack of a rapid rate of change might be due to a ceiling effect because of the high absolute performance scores. This is less likely to be the case with CDI performance, because the majority of the observed CDI scores are in the 50s and 60s and there is plenty of room for score improvement.

Another possible explanation for the Group 3 lowered rate of performance improvement during the 3rd year would be differences in extrinsic motivation to perform on the test. These students had no school imposed extrinsic incentive to perform well on the test. However, students in Groups 1 and 2 performed under those same incentive conditions, and their observed rate of improvement is higher than that of Group 3. It is possible that these extrinsic incentive conditions affected students in Group 3 differently than those in Groups 1 and 2, but we cannot confirm or disconfirm this with existing data.

The final alternative explanation could be a function of the cross-sectional design of this study. It is possible that students in the class that comprised Group 3 had a different achievement profile due to selection, or training. To investigate this, we compared the USMLE Step 1 scores of this class with those of the three classes preceding them. The average USMLE Step 1 score for this class was not significantly different than that of the preceding three classes, suggesting that the lower level of gain in CDI and DPR performance for Group 3 was not likely to be a function of the achievement profile of the class.

This lack of a spike in DPR and especially CDI performance by Group 3 members suggests that the process of medical students converting information from a form that was functional for passing basic science multiple-choice tests into a form that is functional for retrieval and use in clinical settings[2] may take more than a single year of clinical training and experience to accomplish. Regehr and Norman stated that medical education strategies should be directed at three goals: (a) to enhance meaning of information, (b) to reduce dependence on specific contexts (case features), and (c) to provide repeated practice in retrieving information useful in particular clinical situations. All three of these goals are heavily dependent on repeated and varied practice across a range of clinical presentations and situations. It may well be that the experience gained during the 3rd year of medical school disrupts student concepts of classic disease presentations learned during the first 2 years. Students' views of findings and their relationships to a disease process may be distorted because their experience is limited to a very few patients. This is analogous to believing that all cats are black and white because the only two cats seen in your experience have been that color. The performance of the family physicians with considerable patient experience was higher than that of medical students on both the DPR and CDI tests.

In summary, we agree with Eva[1] that both DPR and analytic fact-checking (CDI) ability are necessary for the efficient and effective practice of clinical diagnosis. Our results suggest that medical students are acquiring the necessary clinical pattern detection ability and that all years of medical training appear to be contributing to development of this ability. As for analytic fact checking, medical student performance is improving with each year of training, but the rate of improvement is slower and the absolute level of performance is low even after the 1st clinical year of training. The relatively low absolute level of CDI performance by practicing family medicine physicians suggests that this finding could simply reflect the lack of agreement in medicine regarding the strength of relationships between individual findings and diagnoses. Nonetheless, it seems prudent to investigate what changes in instruction and experience will increase the rate of student acquisition of analytic fact-checking ability.

## REFERENCES

1. Eva KW. What every teacher needs to know about clinical reasoning. *Medical Education* 2005;39:98–106.
2. Regehr G, Norman GR. Issues in cognitive psychology: Implications for professional education. *Academic Medicine* 1996;71:988–1001.
3. Mamede S, Schmidt HG. The structure of reflective practice in medicine. *Medical Education* 2004;38:1302–8.
4. West DC, Pomeroy JR, Park JK, Gerstenberger EA, Sandoval J. Critical thinking in graduate medical education: A role for concept mapping assessment? *Journal of the American Medical Association* 2000;284:1105–10.
5. Case SM, Swanson DB, Stillman PL. Evaluating diagnostic pattern recognition: The psychometric characteristics of a new item format. *Proceedings of the Annual Conference on Research in Medical Education* 1988;27:3–8.
6. Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C. The Script Concordance test: A tool to assess the reflective clinician. *Teaching and Learning in Medicine* 2000;12:189–95.