Journal of Experimental Psychology: Applied
2011, Vol. 17, No. 3, 195–209

1076-898X/11/$12.00 DOI: 10.1037/a0024461

Teaching Posttraining: Influencing Diagnostic Strategy With Instructions at Test

Chan Kulatunga-Moruzi, Lee R. Brooks, and Geoffrey R. Norman
McMaster University

It is believed that medical diagnosis involves two complementary processes, analytic and similarity-based. There is considerable debate as to which of these processes defines diagnostic expertise and how best to teach clinical diagnosis and reduce diagnostic errors. The purpose of these studies is to document the use of these strategies in medical students. We shifted the balance in use of these processes and improved diagnostic accuracy with instructions given posttraining at the moment of diagnosis. Analytic processing reflecting the degree to which cases contain the diagnostic rules was indexed by the rate of accuracy on typical versus atypical cases (typicality effect). Similarity-based processing reflecting the degree to which cases resemble previously encountered cases was indexed by the rate of accuracy on similar versus dissimilar cases (similarity effect). Two studies are presented illustrating that diagnosis involves the coordination of analytic and similarity-based processes and that differential instruction given at test shifts the balance in the use of these processes. Study 1 illustrated that participants adopting an analytic strategy exhibit a larger effect of typicality. Participants adopting a similarity-based strategy exhibit a larger effect of similarity. The diagnostic approach of students given no instructions was predominantly analytic. Dual instructions in which participants first employed similarity-based processing followed by the application of rules improved overall accuracy. Study 2 investigated two versions of dual instructions and illustrated that assessing a case with the rules of diagnosis first may inhibit the subsequent use of similarity-based reasoning. The implications for diagnostic expertise and pedagogy are discussed.

Keywords: analytic processing, similarity-based processing, coordination of diagnostic processes, medical cognition, medical education

To accomplish the task of medical diagnosis, information of various types must be considered and integrated. Researchers have advanced various conceptual models of this process and its representation in memory. Expert knowledge has been described in terms of propositional networks (Patel & Groen, 1986), semantic axes (Bordage, 1994), symptom by disease probability matrices (Papa & Elieson, 1993), and concept maps (McGaghie, Boerger, McCrimmon, & Ravitch, 1996). The diagnostic process has been viewed as hypothetico-deductive, in which the practitioner advances multiple hypotheses and tests them by gathering additional data (Elstein, Shulman & Sprafka, 1990; Neufeld, Barrows, Feightner, & Norman, 1981) or as a process of forward reasoning, inferencing from data to diagnosis (Patel, Groen & Arocha, 1990).

These accounts of expert diagnosis assume a single type of reasoning process and a single type of knowledge base. Such a position must be viewed as simplistic. It is possible that in reaching a diagnosis, at least two types of information, *formal analytical knowledge*, including the rules of diagnosis, knowledge of causal mechanisms, and the symptom-disease probabilities, and *episodic memory of specific cases* (implicitly or explicitly) are used. Further, it is conceivable that the balance in the use of these two kinds of information may change with experience. Because novice clinicians have little clinical experience, they would presumably rely more heavily on formal rule-based knowledge. In contrast, experts who can draw on years of experience may rely more on exemplar-based knowledge (Norman, Brooks, Allen, & Rosenthal, 1990; Schmidt, Norman, & Boshuizen, 1990; Schmidt, Boshuizen, & Norman, 1992; Schon, 1983, 1987). The relative balance of these strategies may also depend on the difficulty or nature of the case. Evidence for these conjectures derives from experimental studies of diagnosis in visual domains such as dermatology, electrocardiology, and patient appearance (Brooks, Norman & Allen, 1991; Kulatunga-Moruzi, Norman, and Brooks, 2001; Kulatunga-Moruzi and Brooks, 2004; LeBlanc, Brooks, & Norman, 2002; Regehr, Cline, Norman, & Brooks, 1994) as well as with verbal patient information material (Hatala, Norman & Brooks, 1999a; Mamede

Chan Kulatunga-Moruzi, Department of Family Medicine, McMaster University, Hamilton, Ontario, Canada; Lee R. Brooks, Department of Psychology, Neuroscience, and Behaviour, McMaster University; Geoffrey R. Norman, Department of Clinical Epidemiology and Biostatistics, McMaster University.

This research was supported by a Natural Science and Research Council of Canada predoctoral fellowship and an Ontario Graduate Studies predoctoral scholarship awarded to the first author and by grants to each of the second and third authors awarded by the Natural Engineering Research Council of Canada. The studies presented here were conceived of and conducted by the first author.

It is with deep regret that we acknowledge that Lee Brooks passed away during the preparation of this article.

Correspondence concerning this article should be addressed to Chan Kulatunga-Moruzi, Faculty of Health Sciences, McMaster University, Michael G. DeGroote Centre for Learning and Discovery, room 2209, Hamilton ON, L8N 3Z5 Canada. E-mail: kulatuc@mcmaster.ca

et al., 2010a; Young, Brooks, & Norman, 2007). Emerging from these studies is evidence for two concurrent processes. Diagnosis seems to involve an analytical process, invoking the diagnostic rules and the comparison of the signs and symptoms of the presenting case against them, and a similarity-based process, the comparison, often without awareness, of the similarity of the present case to particular previously encountered cases.

These two hypothesized components have different influences on the diagnostic process. When engaged in analytic processing, the clinician is thought to reach a diagnosis by considering the separate features of the case and comparing these features to well-established diagnostic criteria. Basal cell carcinoma, for example, may be defined by the rules: flesh-colored, translucent, pearly, dome-shaped lesions with telangiectasia on sun exposed skin, with the probability being higher with pale skin and increasing age. Typical cases, those possessing several features of a particular diagnostic category, are diagnosed more efficiently than atypical cases, those containing fewer features of a particular diagnostic category. This *independent cues model* is the basis of several approaches to medical decision-making, including Bayesian (Weinstein & Fineberg, 1980) multiple regression (Wigton, 1988), and prototype models of diagnosis (Bordage & Zacks, 1984), as well as prototype models of categorization (Rosch & Mervis, 1976).

By contrast, similarity-based processing is thought to make use of memory for specific, previously diagnosed cases. If the physical findings and history of the presenting case is similar to a previously encountered case, then it is likely to be of the same diagnosis (Medin, Altom, Edelson, & Frecko, 1982; Brooks, Norman, & Allen, 1991). A critical assumption of this model is that the similarity between cases may arise from contextual aspects of the stimulus in addition to diagnostically relevant features (Allen & Brooks, 1991; Regehr & Brooks, 1994; Hatala, Norman, & Brooks, 1999b). Given the availability of such a repository in memory, it is reasonable to presume that cases that match a prior exemplar are diagnosed more efficiently than dissimilar cases.

Each of these hypothesized components of the diagnostic process has potential positive and negative consequences. A strict analytic strategy whereby all diagnostic data are carefully collected and assessed may reduce the chances of premature closure, that is, the early termination of a patient assessment. However, such an approach may also be an impediment to accurate diagnosis. Norman, Brooks, Colle, and Hatala (2000) found that learners who were instructed to use an analytic strategy by carefully outlining all clinical features before generating any diagnostic hypotheses performed less well than those who simply diagnosed the case and listed features after diagnosis. While participants in the diagnose-then-list-features condition provided a succinct list of clinical features, those in the analytic condition tended to outline irrelevant features, likely producing a confusingly long list of potential diagnoses (i.e., differentials) to choose from. The differential accuracy rate may also be attributable to the strict analytic strategy interfering with the normal balance of combined analytic and similarity-based processes that may have been adopted by those in the diagnose-then-list-features condition (Kulatunga-Moruzi and Brooks, 2001; Kulatunga-Moruzi, Brooks, Norman, and Scully, 2002).

Based on this research, Kulatunga-Moruzi, Brooks, and Norman (2004) examined the performance of experienced clinicians who adopted a strict analytic strategy. Because experienced clinicians typically generate diagnostic hypotheses within seconds of encountering a case (Barrows, Norman, Neufeld, & Feightner, 1982; Elstein, Shulman, & Sprafka, 1977), a strict analytic strategy was enforced by having experienced practitioners consider an accurate, comprehensive feature list before proceeding to examine a photograph of the case. This study showed that the diagnostic performance of experienced clinicians forced into an analytic mode of processing was significantly lower than the performance of a similar group of clinicians who simply diagnosed the case from a photograph alone. Thus, while the careful assessment of features may be prudent and warranted under certain circumstances, the routine analysis of all presenting features may decrease diagnostic performance.

Using a similarity-based strategy may also result in both positive and negative consequences. A potential benefit of the use of similarity is that it allows for the rapid generation of diagnostic hypotheses, possibly resulting from the consideration of features not included in the formal diagnostic rules. While such a strategy may be highly efficient (Hatala, Norman, & Brooks, 1999a; Kulatunga-Moruzi, Brooks, and Norman, 2001), a potential drawback is that it does not provide a way of rejecting an initial incorrect diagnosis. Further, in instances of comorbidity, that is, when more than one medical condition presents in a patient at a given time, the examination of a case may be ended too early as soon as one diagnosis is discovered. This type of premature closure, known as satisfaction of search, is well documented (Ashman, Yu, and Wolfman, 2000; Berbaum et al., 1990) and may be especially so in cases of comorbidity, in which one diagnosis presents atypically and thus is missed as soon as the typically presenting diagnosis is made. Given the evidence for confirmation bias and the ambiguity of medical stimuli (Berbaum et al., 1986, 1988; Brooks, LeBlanc, & Norman, 2000; Hatala, Norman, & Brooks, 1999b; Norman, Brooks, Coblentz, & Babcook, 1992; see Nickerson, 1998 for review), such a strategy may lead to diagnostic error. Many clinical educators (Klein, 2005; Redelemeier, 2001; Croskerry, 2002), therefore, warn against relying on similarity-based processing.

Because there are potential positive and negative aspects with each of these processes, the coordination of these forms of knowledge, whereby errors generated by one process may be corrected by the other, is likely to improve diagnostic performance. Following this logic, Kulatunga-Moruzi (2005) has shown that the diagnostic accuracy of medical students on dermatology cases was significantly better when instructed to systematically use both analytic and similarity-based processing than when instructed to use a single strategy or when no instructions are given, leaving students to adopt their own strategy. Ark, Brooks, and Eva (2006, 2007) have shown similar findings in the diagnosis of cardiology cases by adopting the same conditions and rationale.

Researchers have also theorized how and when clinicians shift from one mode of processing to another (Schon, 1983, 1987; Dreyfus & Dreyfus, 1986; Bereiter & Scardamalia, 1993) and have uncovered some of the variables that may impact on the utility of each of these diagnostic strategies. Using internal medicine cases, Mamede, Schmidt, Rikers, Penaforte and Coelho-Filho (2007) showed that ambiguous internal medicine cases led residents to switch from a similarity-based, automatic mode of processing to a more analytic, reflective mode of processing. Mamede, Schmidt,

and Penaforte (2008) have also shown that while there was no difference in diagnostic accuracy of routine internal medicine cases between participants who used a quick first impression (likely based on similarity-based processing) and those who were guided to use more analytic processing (based on reflective practice), the diagnosis of more complex cases benefited from a more analytic approach. Further, Mamede and colleagues (2010a) have shown that experienced clinicians using similarity-based processing made more errors consistent with an availability bias and that a reflective analytic strategy countered this bias. These researchers (Mamede et al., 2010b) have also shown that for complex cases, expert performance was improved by the use of analytic processing (conscious deliberation). The performance of novices, however, was improved by similarity-based processing (deliberation-without-attention) for simple cases.

While these studies have provided much insight into the diagnostic process, none of them have systematically indexed the use of both modes of operation. The works of Kulatunga-Moruzi, (2005); Kulatunga-Moruzi, Brooks, and Norman, (2001) and Regehr et al. (1994) are exceptions. Using slides of dermatological lesions and a 2 × 2 factorial design, these researchers experimentally orthogonalized the dimensions of typicality and similarity to examine the separate effects of test-time instructions on analytic and similarity-based processes. The definitions of typicality and similarity in these experiments require some elaboration. In contrast to the psychological literature on prototype theory, typicality in these studies is defined by how well a case displays the characteristics named in the clinical rules. Similarity in these studies refers to the specific similarity of an example previously encountered in the training phase, which may derive from diagnostically relevant and or irrelevant features. It was reasoned that if typical cases are diagnosed more accurately than atypical cases regardless of similarity, then analytic processing is dominant. If, however, similar cases are diagnosed more accurately than dissimilar cases, regardless of typicality, then similarity-based processing is dominant. This orthogonal design revealed a significant component of both analytic and similarity-based processing in both medical students (Kulatunga, Brooks, and Norman, 2001) and residents (Regehr et al., 1994). Further, each of the studies indicated that the instructions given at the moment of test may influence the balance of the two modes of processing. However, the results were inconsistent. While the study with students only showed a positive effect of instruction for the Similarity-based group, the study with residents only showed a positive effect of instruction for the Rule-based group.

The present paper extends the work of Kulatunga-Moruzi, Brooks, and Norman (2001) using a similar orthogonal factorial design. In two studies, we examine the coordinated use of both analytic and similarity-based processes, shifting the balance in the use of these two modes of processing with instructions. Further, we examine how the diagnostic process of learners may be optimized. As with the earlier work, analytic processing is indexed by the difference in performance between typical and atypical cases regardless of similarity. Similarity-based processing is indexed by the difference in performance between similar and dissimilar cases regardless of typicality. The results of the studies indicate the use of both types of processes in reaching diagnosis. Further, Study 1 revealed that the usual diagnostic approach of medical students when no instructions are given is more analytic, resembling the pattern of performance of participants in the rule-based condition. Participants instructed to use similarity to generate diagnostic hypotheses and then to actively test these hypotheses using the clinical rules outperformed participants instructed to use a single strategy. Study 2 was designed to examine whether committing to a diagnosis based on similarity leads to an anchoring effect and to determine whether assessing a case using the clinical rules breaks up a case into feature components such that subsequent similarity-based processing is impeded. This study illustrated that assessing a case with the rules of diagnosis first may inhibit the subsequent use of similarity-based reasoning and that the use of similarity-based processing first is unlikely to result in an anchoring effect.

Experiment 1: Manipulating and Optimizing Diagnostic Strategy

In the Kulatunga-Moruzi, Brooks, and Norman (2001) study, one group of participants was given instructions to foster analytic processes (Rule-based group) while a second group was given instructions to foster similarity-based processes (Similarity-based group). While the instructions at the time of the test differed for the two groups, both groups received the same pretest training. Similarity-based instructions at test significantly shifted the balance toward more similarity-based processing, but the Rule-based group did not exhibit a significantly larger typicality effect than the Similarity-based group. Overall accuracy was low (37%).

A plausible explanation for the lack of typicality by instruction interaction in the Kulatunga-Moruzi, Brooks, and Norman (2001) study may be that the participants were complete novices with no training in dermatology. The lack of formal knowledge of dermatology, as indicated by participants' low overall accuracy rate, would limit their ability to think more analytically and apply the rules when asked. To rectify this issue, all participants in the current studies were given a brief introduction to dermatology and the rules of diagnosis for the disorders presented in the study.

To examine the diagnostic performance of novice clinicians further, two new conditions were added. The first, Similarity + Rule condition, was designed to foster both types of processing, in an attempt to improve overall diagnostic performance. In this potentially optimal condition, participants were first instructed to use similarity between cases to generate a diagnostic hypothesis (Pass 1) and then to reassess this initial diagnosis by carefully applying the learned rules (Pass 2). We reasoned that this condition might take advantage of the positive aspects of each type of processing while minimizing the negative aspects by having the other strategy as a potential check. The second condition, No Instructions, functions as a baseline condition to determine the relative reliance on analytic and similarity-based processing when no explicit instructions are given. Further, it allows us to compare such performance to performance in which instructions are given. Because one of the goals of the Kulatunga, Brooks, and Norman (2001) and Regehr et al. (1994) studies was to shift the relative reliance on the use of these two processes, performance has only been observed under instructions. As such, the normal diagnostic strategy of learners is yet to be determined. By asking participants to focus on a single strategy (rule-based or similarity-based) in previous studies, the potential to adopt both strategies may have been hindered. It might, therefore, be reasonable to expect performance in the absence of instructions to resemble performance

under a dual strategy. Because the rules of diagnosis are strongly emphasized in early medical training, however, we anticipated performance under the No Instruction condition to be similar to performance under the Rule-based condition. Thus, four experimental conditions were run: Rule-based, Similarity-based, Similarity + Rule, and No Instruction. See Figure 1 for schematic of design.

Method

Participants. The participants in the study were 51 medical students at McMaster University in their first three instructional units before clerkship (17–22 weeks completed) in the 2002–2003 academic year. The McMaster undergraduate medical program uses a problem-based approach to learning. The curriculum consists of five organ system units. The program is three years in length and runs 11 months in each year. The first two years largely consist of academic learning, while the third year is primarily dedicated to clinical training. None of the participants had completed any prior training in dermatology. Participants were recruited by the experimenter and paid for their participation. To provide appropriate instruction and feedback, experimental sessions were run in groups of two to five. Participants were informed that their participation was voluntary and that their performance had no bearing on their academic record. Ethics approval for this study was granted by the Faculty of Health Sciences/Hamilton Health Sciences Corporation Research Ethics Board.

Materials. The materials were color slides of the most common North American dermatological conditions consisting of 10 disease quartets similar to the Kulatunga, Brooks, and Norman (2001) study with minor changes plus five filler cases. Each disease quartet consisted of four different cases representing a different disease condition. Within each quartet, two of the cases were typical and two were atypical. The two typical cases (T1 and T2) were similar in overall appearance, and the two atypical cases (A1 and A2) were similar in overall appearance. Each typical pair in a given quartet was dissimilar in overall appearance to the atypical pair. Please see appendix for examples of quartets.

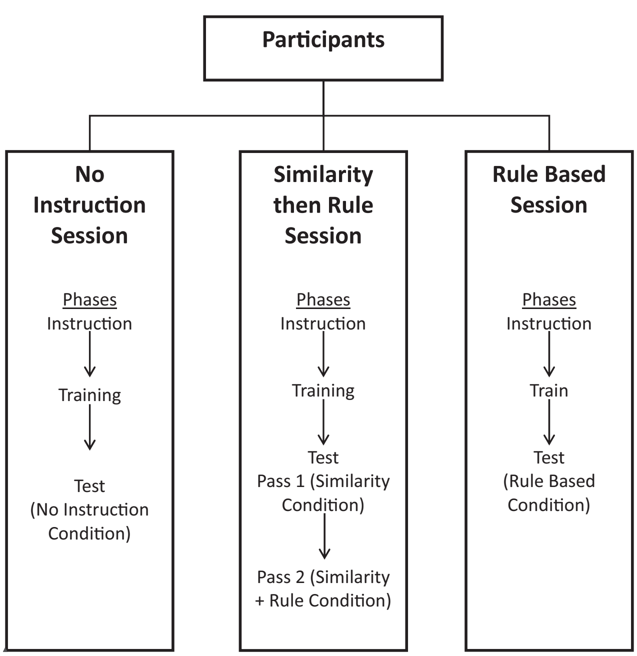


Figure 1. Study 1: Experimental procedure.

Typicality ratings. To verify the dimension of typicality, each slide was rated for typicality by two dermatologists on a seven-point scale (1 = *highly typical, a textbook example* and 7 = *highly atypical, a bizarre manifestation of the disease*). The items were dichotomized based on these ratings: typical (1–3) and atypical (5–7). The mean typicality rating for the typical pairs was 1.6, ranging from 1.0 to 2.5. The mean typicality rating for the atypical pairs was 5.5, ranging from 5.0 to 7.0. As a final check for the dimension of typicality, the two typical slides and the two atypical slides in the each quartet had typicality ratings within one point on a seven-point scale.

Similarity ratings. To verify the dimension of similarity, the similar slides were chosen such that many of the variables visible in the case were the same (e.g., location, size, color, and general appearance of lesion, gender, race, and age of patient, angle of photograph, lighting, etc.). Thus, similarity was strictly governed by visual cues and no patient history accompanied the photographs. Two nonexperts, who had no specific knowledge of dermatology, judged the cases for similarity; picking two typical and two atypical cases that were most similar in overall appearance. A second set of two nonexperts as well as two dermatologists rated the quartet pairs for similarity on a seven-point scale (1 = *not at all alike* and 7 = *highly alike*). The mean similarity rating for the nonexperts for the typical pairs was 6.15, ranging from 5.0 to 7.0. The mean similarity rating for the atypical pairs was 6.1, ranging from 5.0 to 7.0. The similarity ratings of the dermatologists were similar to that of the nonexpert raters. For the dermatologists, the mean similarity rating for the typical pairs was 6.0, ranging from 5.0 to 7.0. The mean similarity rating for the atypical pairs was 6.0, ranging from 5.0 to 7.0.

Five additional slides from four additional disease categories were selected as filler items, yielding 45 possible cases. Three of these filler cases were shown in the training phase and two in the test phase. Materials in the instructional phase consisted of 14 additional color slides of the same 10 diseases represented in the quartets as well as 4 additional disease categories. The instructional cases were carefully chosen so that they were not contextually similar to any cases shown in the training and test phases.

Procedure. The study consisted of an instructional phase followed by a training phase, then a test phase. All participants engaged in each phase of the experiment. In the training phase, one case from each quartet was presented; with five slides representing typical cases (T1) and five slides representing atypical cases (A1) along with three filler cases (13 slides). The test phase consisted of two of the remaining slides, one typical case (T2) and one atypical case (A2) from each quartet resulting in 10 typical cases and 10 atypical cases. Within each of the sets of 10 slides, five were similar to the training slides and five were dissimilar to the training slides. Thus, the test slides consisted of five typical-similar, five typical-dissimilar, five atypical-similar, and five atypical-dissimilar cases. With the addition of two filler cases, there were 22 test cases. To ensure that there was no bias resulting from slide selection during training, four counterbalances were used such that each case was used in both the training and test phases of the study.

Instructional phase. Because the participants had no prior instruction or experience in dermatology, the instructional phase was designed to acquaint them with the basic terminology used in dermatology and 14 of the most common North American dermatological conditions. In this phase, participants were given a booklet developed by one of the authors (C.K.-M.) that described the 14 conditions (representing the 10 disease quartets plus the four filler categories) used in the study. Participants were given as much time as they needed to review the booklet. Participants took approximately 30 minutes to review the booklet and indicated to the experimenter when they were ready to proceed. Participants were then shown typical cases of each of 14 conditions. Using the booklet as a guide, the clinical features of each case were pointed out and discussed. Participants were free to ask questions and discuss the cases. Each case was presented and discussed for approximately five minutes.

Practice phase. Immediately after the instructional phase, all participants engaged in a practice phase in which each slide (one from each of the 10 disease quartets plus three filler cases), was presented twice. To establish that similarity to a past case is often a useful and effective strategy, on the first pass participants were asked to think about whether they had encountered a similar case from personal experience, a family member or friend, textbook, and so forth. They were then asked to examine each case carefully and list all presenting features using the instructional booklet. Finally, they were asked to diagnose each case (primary diagnosis) and list any differentials (other possible diagnoses) they considered. All responses were recorded in a training booklet. Participants were given as much time as they needed with each case before proceeding to the next. Once all 13 cases had been diagnosed, each case was presented again. On the second pass, participants were invited to volunteer their diagnoses. For each case, feedback was given with respect to the correct diagnosis, its typicality (typical/atypical), as well as the presenting clinical features. For atypical cases exhibiting few clinical features, participants were reminded of the rules of diagnoses. On the second pass, each case was discussed for three to five minutes.

The practice phase was intended to invoke both types of processing: analytic processing by encouraging the systematic evaluation of clinical features and similarity-based processing by providing prior experience with particular case and giving a rationale for using similarity and was identical to the practice phase of the Kulatunga-Moruzi, Brooks, and Norman (2001) study. A key factor of the practice phase is that it allows for the manipulation of similarity by providing participants with a particular example from each quartet that is similar to one that is used at test.

Test phase. Immediately after training, participants engaged in the test phase. Each subject was randomly allocated to one of three experimental *sessions*: No Instruction, Rule-based, or Similarity then Rules. Participants in the No Instruction session were simply asked to diagnose each case without instruction on how to approach the task (NI condition). Participants in the Rule-Based session were asked to carefully assess each case by identifying all clinical features present in the case before providing a diagnosis (RB condition). Participants in the Similarity then Rule session were presented with each test case twice. On the first pass, participants were instructed to provide a quick diagnosis by giving the diagnosis that first came to mind (SB condition). On the second pass, participants were instructed to reexamine their initial diagnosis thoroughly using the rules of diagnosis (S + R condition). Participants were explicitly told that they were welcome to keep or change their initial diagnosis as necessary. Participants in all three conditions took approximately 45 minutes to complete the test phase of the study. Thus, three sets of comparisons may be made: 1) RB to SB, 2) RB to S + R and SB to S + R, and 3) NI to S + R, SB, and RB. In both the training and test phases, if the participant listed the correct diagnosis as their primary diagnosis, it was scored as correct.

Comparing RB and SB conditions. The analyses and comparison of performance under the RB and SB conditions allows for replication and allows for some interesting comparisons on case type. Based on the previous research, we anticipated effects of both typicality and similarity. If the type of information used at the moment of diagnosis is amenable to instruction, we should observe a larger effect of typicality and a smaller effect of similarity for the RB group and a larger effect of similarity and a smaller effect of typicality for the SB group. Because typical-similar cases can be diagnosed by using either the rules of diagnosis or by drawing an analogy to the similar case in the practice phase, we anticipated similar rates of relatively high diagnostic accuracy for both the RB and SB groups. Following the same logic, because neither the rules of diagnosis nor similarity to a case examined in the practice phase is helpful in diagnosing atypical-dissimilar cases, we anticipated similar rates of relatively low accuracy for both instructional groups. The critical comparisons indicating a shift in the balance of the type of information used at the moment of diagnosis arise from the examination of typical-dissimilar and atypical-similar cases. Specifically, because typical-dissimilar cases can be more adequately assessed with the rules of diagnosis, we anticipated the RB group to have an advantage over the SB group for typical-dissimilar cases. For atypical-similar items, however, we anticipated the SB group to have an advantage over the RB group because these cases contain few of the diagnostic features but are reminiscent of cases seen in training.

Comparing RB to S + R conditions and SB to S + R conditions. The comparison of performance under the S + R conditions to both of the conditions emphasizing only one mode of operation, RB and SB conditions, allows us to examine whether instructions fostering both types of processing would improve diagnostic accuracy over the use of a single diagnostic strategy. We reasoned that the two-stage procedure would have the most impact on typical-dissimilar cases, where increased attention to the rules on the second pass might correct errors arising from similarity-based reasoning on the first pass.

Comparing the NI condition to S + R, SB, and RB conditions. The comparison of performance under the NI condition to performance under each of the other conditions, RB, SB, and S + R conditions, allows us to determine the usual strategy adopted by novice clinicians, whether the pattern of performance under the NI condition resembles any of the other strategies and which of these strategies results in the greatest diagnostic accuracy. It may be that the “normal strategy” is one that incorporates the use of both similarity and typicality, such that the pattern of performance under the NI condition is similar to that of the S + R condition. Because there is a heavy emphasis on the rules of diagnosis in medical training and these participants have had little opportunity to acquire and encode examples, we anticipated that performance

under the NI condition would be similar to performance under the RB condition.

Results and Discussion

Table 1 and Figure 2 depict diagnostic accuracy for participants in each of the four instructional conditions: RB, SB, S + R, and NI on each of the four case types (typical-similar, typical-dissimilar, atypical-similar, and atypical-dissimilar). Although participants were asked to list differentials in the experimental phase, for the majority of cases only one diagnosis was listed. Therefore, a case was scored correct if the participant listed the correct diagnosis as their primary diagnosis. Differentials were not analyzed. Although a direct comparison cannot be made between the Kulatunga, Brooks, and Norman (2001) study owing to small differences in experimental materials, the overall accuracy of participants in the current study (52%) is substantially higher than that of participants in the previously reported study (37%), confirming that students had adequately grasped the diagnostic rules covered in the instructional phase.

An analysis of the training data indicated that diagnostic performance on typical cases (50%) was significantly better than on atypical cases (40%), providing a check of our manipulation of typicality, $t(51) = 2.13$, $p < .05$, $d = 0.28$.

Comparing RB and SB conditions. To assess the impact of instruction given at the moment of test, the proportion of correct responses for each of the four types of cases (typical-similar, typical-dissimilar, atypical-similar, and atypical-dissimilar) was used to compare the two instructional test groups using a three-way mixed analysis of variance with instructional test condition (RB, SB) as a between-subjects factor and typicality (typical, atypical) and similarity (similar, dissimilar) as within-subject factors.

The results of the analysis revealed no difference in overall accuracy between the RB (50%) and SB (51%) conditions, $F(1, 37) = 0.049$, MSe = 0.050, $p = .83$. The effect of typicality was significant, $F(1, 37) = 58.7$, MSe = 0.031, $p < .001$, $\eta^2 = .19$ with typical cases (61%) being diagnosed more accurately than atypical cases (40%). The effect of similarity $F(1, 37) = 140.7$, MSe = 0.051, $p < .001$, $\eta^2 = .74$ was also significant with similar cases (73%) being diagnosed more accurately than dissimilar cases (29%). The typicality by instruction interaction, $F(1, 37) = 10.14$, MSe = 0.031, $p < .001$ was significant, as was the similarity by instruction interaction, $F(1, 37) = 7.88$, MSe = 0.051, $p < .05$. Post hoc analyses (Newman–Keuls) revealed that participants in the SB group (78%) diagnosed similar cases significantly more accurately than participants in the RB group (67%), $p < .05$, and that participants in the RB group (66%) diagnosed typical cases significantly more accurately than participants in the SB group (57%), $p < .05$. No other interactions were significant.

These results indicate that both the rule-based and similarity-based instructions given at the moment of test are effective in altering the diagnostic strategies used. Planned comparisons of diagnostic performance for each of the four types of test cases are consistent with this interpretation. Specifically, for typical-similar cases, diagnostic performance under the RB condition (81%) was similar to that of the SB condition (84%), $t(37) = 0.52$, $p = .74$. Likewise, there was no difference in diagnostic accuracy for atypical-different cases (Rule-based 18%, Similarity-based 18%), $t(37) = .06$, $p = 68$. However, for typical-dissimilar cases, the accuracy of those adopting a RB strategy (50%) was significantly greater than those adopting a SB strategy (31%), $t(37) = 2.44$, $p < .05$, $d = .91$. Conversely, for atypical-similar cases, those adopting a SB strategy (71%) significantly outperformed those adopting a RB approach (52%) $t(37) = 2.97$, $p < .01$, $d = .93$.

Comparing RB and S + R conditions and SB and S + R conditions. To assess whether systematic instruction to use both specific similarity and the diagnostic rules lead to superior performance than when instructed to use a single strategy, two separate planned comparisons were done. The first compared overall diagnostic performance under the S + R condition to performance under the RB condition, and the second compared performance under the S + R condition to the SB condition. An independent sample t test comparing overall performance under the RB condition (50%) to the S + R condition (57%) indicated that performance under the dual strategy (S + R) was significantly better than performance under the RB strategy, $t(37) = 3.84$, $p < .001$, $d = .55$. A paired t test indicated that performance under the S + R condition (57%) was also significantly better than under the SB condition (51%), $t(37) = 3.50$, $p < .001$, $d = .76$. Students were able to gain accuracy on the second pass by reassessing each case using the rules of diagnosis.

To examine the improvement in diagnostic accuracy under the dual instructions more closely, any change in primary diagnosis from Pass 1 (SB condition) to Pass 2 (S + R condition) was noted. A score of + 1 was assigned for a change to the correct diagnosis, and a score of −1 was assigned to a change to an incorrect diagnosis. Thus, each participant's score could potentially range from −20 (Pass 1 all correct, Pass 2 all incorrect) to + 20 (Pass 1 all incorrect, Pass 2 all correct). Analysis of the average change score between typical (0.38) and atypical (0.10) cases approached

Table 1
Study 1 Mean Diagnostic Accuracy (and Standard Deviation) as a Function of Case Type and Instructional Conditions

				Test				
		Training		Typical		Atypical		
Instruction	# S	Typical	Atypical	Similar	Dissimilar	Similar	Dissimilar	All cases
Rule-based	18	0.47 (0.27)	0.43 (0.12)	0.81 (0.16)	0.50 (0.25)	0.52 (0.24)	0.18 (0.15)	0.50 (0.13)
Similarity-based pass 1	21	0.50 (0.27)	0.42 (0.17)	0.84 (0.16)	0.31 (0.24)	0.71 (0.16)	0.18 (0.18)	0.51 (0.10)
Similarity + rule pass 2	21			0.90 (0.14)	0.42 (0.23)	0.71 (0.20)	0.24 (0.16)	0.57 (0.10)
No instruction	12	0.50 (0.14)	0.32 (0.18)	0.80 (0.15)	0.48 (0.28)	0.44 (0.11)	0.24 (0.14)	0.49 (0.07)

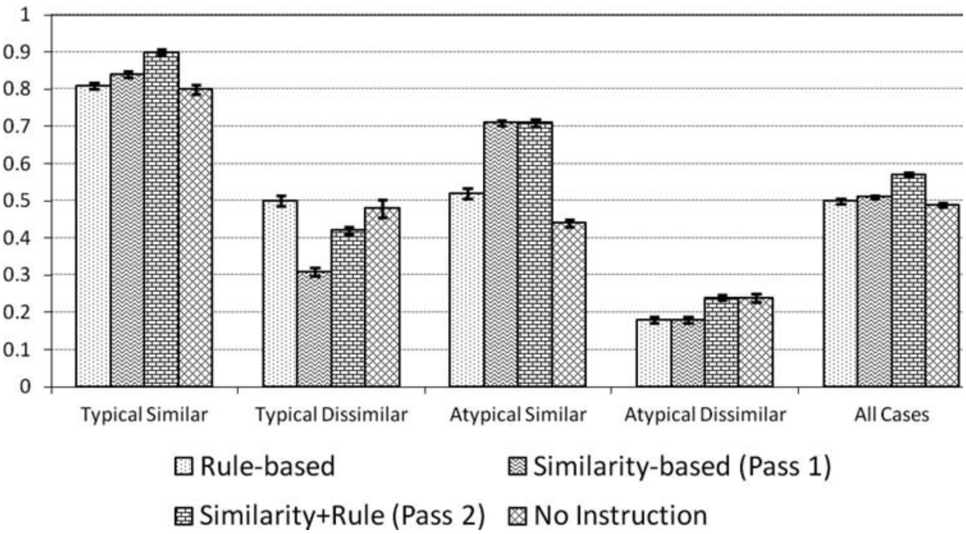


Figure 2. Study 1: Mean diagnostic accuracy as a function of case type and instructional conditions.

significance $t(41) = 1.48$, $p = .07$. Consistent with the hypothesis that analytic processing on the second pass would primarily benefit typical cases, a planned comparison of the average change score on typical different (0.43) and atypical-similar (0.0) cases indicated that significantly more changes were made for typical different cases $t(20) = 1.75$, $p < .05$.

Comparing the NI condition to RB, SB, and S + R conditions. The pattern of performance under the NI condition was remarkably similar to that of the RB condition, particularly for typical cases, indicating that the clinical decision making of students relies heavily on diagnostic rules. A three-way mixed analysis of variance with instructional group (RB, NI) as a between-subjects factor and typicality (typical, atypical) and similarity (similar, dissimilar) as within subject factors yielded only significant effects of typicality $F(1, 28) = 79.27$, MSe = .03, $p < .0001$, $\eta^2 = .51$ and similarity $F(1, 28) = 51.44$, MSe = 0.05, $p < .0001$, $\eta^2 = .48$, confirming that performance under NI condition (49%) is similar to that of performance under the RB condition (50%). While it may be argued that the instructional phase of the experiment may have primed participants in the NI condition to respond using an analytic approach, producing the observed results, the instructional phase is analogous to the type of instruction that medical students receive. Thus, the manner in which students approach the diagnostic task in this experiment is similar to how they would approach any diagnostic task.

Because overall diagnostic accuracy under the RB (50%) and SB (51%) conditions was similar, we can assume that overall accuracy under the SB (51%) and NI (49%) conditions to also be similar. An independent sample *t* test confirmed this conclusion $t(31) = 0.66$, $p = .26$.

An independent sample *t* test on overall diagnostic accuracy confirmed the benefit of dual instructions. The difference in performance between the dual (57%) and NI (49%) conditions approached statistical significance, $t(31) = 2.26$, $p = .05$, $d = .86$. Thus, systematic instruction to use both types of information, case similarity and the rules of diagnosis resulted in greater diagnostic accuracy (an absolute increase of 6–8% or a relative increase of 12–16%) than when instructed to use only one type of information.

The results of this study, consistent with previous studies (Kulatunga-Moruzi, Brooks, and Norman, 2004; Regehr et al., 1994), provide evidence for a coordinated use of analytic and similarity-based processes in reaching a diagnosis. Further, the results illustrate that, keeping training constant, instruction given at the moment of diagnosis can shift the balance of analytic and similarity-based processes. The finding that the performance of preclinical medical students was predominantly analytic is worth documenting, as it is often believed that errors in diagnosis are more likely to occur because of the use of similarity-based reasoning (Croskerry, 2002, 2006). Instructions aimed to capitalize on both types of processes resulted in increased diagnostic performance.

Experiment 2: Fine-Tuning Dual Strategies

A potential problem with the S + R dual strategy used in Study 1 is that although participants were free to change their original diagnosis on the second pass, writing down their primary diagnosis on the first pass may have created an anchoring effect, whereby an incorrect diagnosis arrived at based on similarity is maintained even when the assessment of the case with the rules of diagnosis on the second pass may lead to the correct diagnosis. If this is the case, their primary diagnosis and search for and interpretation of features may be affected (LeBlanc, Brooks, & Norman 2002; Hatala, Norman, and Brooks, 1999a; Norman, Brooks, Coblentz, & Babcook, 1992). Thus, the sequential nature in which the diagnostic task was performed under the S + R dual condition of Study 1 may be less optimal than one in which participants are encouraged to use both types of information concurrently. The order of information used by participants in the dual strategy condition may also be critical. It is possible, for example, that the careful assessment of clinical features "breaks up" the case in such a way that it impedes the ability to subsequently use a similarity-based strategy. Based on previous research (Kulatunga-Moruzi, Brooks, & Norman, 2004) we hypothesized that this would be the case.

The purpose of Study 2 was to examine two different versions of a dual strategy, Similarity & Rule and Rule + Similarity, to address the possible issues of similarity-based processing resulting in an anchoring effect and rule-based processing "breaking-up" the

case such that it is not possible to subsequently use similarity-based processing. Participants in the Rule + Similarity condition were asked to approach the task using the diagnostic rules on Pass 1 and then to reassess this diagnosis using similarity to a prior case on Pass 2. Participants in the Similarity & Rules condition were asked to consider their diagnosis with both strategies.

Examination of the Rule + Similarity condition. If approaching a case with the rules of diagnosis first breaks up a case such that it cannot be assessed anew with similarity, we should find an inability to gain accuracy from Pass 1 to pass 2 in the Rule + Similarity condition. This would be particularly so for atypical-similar cases where one would expect to find the most gain (within study comparison).

Examination of the Similarity & Rules condition. If the act of committing to a diagnosis on paper based on similarity alone before checking that diagnosis with the rules, results in an anchoring effect, we should find performance under Similarity & Rules condition to be significantly better than the performance of participants in the dual strategy (S + R) condition in Study 1 (between study comparison).

Examination of Similarity & Rule and Rule + Similarity conditions. If how analytic and similarity-based processing is combined has an effect on the diagnostic task, we should find the pattern of performance across the four types of cases be different. Thus, the two different versions of a dual strategy, Similarity & Rule and Rule + Similarity, were compared. Based on the previous study, we anticipated the overall rate of accuracy to be similar but the pattern of performance to be different (within study comparison).

Method

Participants. The participants in the study were 22 medical students at McMaster University in their second instructional unit (17–20 weeks completed) in the 2003–2004 academic year. Participation was voluntary. Students were paid for their participation. Experimental sessions were run in groups of two to four students. All participants were informed of their rights as experimental participants and assured that their performance in the experiment would have no bearing on their academic profile. Ethics approval for this study was granted by the Faculty of Health Sciences/Hamilton Health Sciences Corporation Research Ethics Board.

Although the participants in Study 2 were from the next academic year, effort was made to ensure that participants in Study 2 were comparable to those in Study 1. The participants were at a similar point in the curriculum and had not covered any dermatology. The admissions selection process for the undergraduate medical program and the recruitment process for the study remained the same. Further, all details of the study, the experimenter, location of study, recruitment process, and study procedure were held constant. Thus, there is no reason to believe that this cohort differed significantly from the previous cohort, allowing us to compare performance across the two studies.

Materials. The materials used in the study were the same dermatological cases used in the previous study.

Procedure. The design of the study was identical to that of the previous study. Participants first engaged in an instructional phase aimed at acquainting them with the basics of dermatology, followed immediately by a training phase and transfer phase. The materials, procedures, and expectations of participants in the instructional and training phases of the study were also identical to that of the previous study. In the transfer phase, participants were randomly assigned to one of two experimental groups, Rule + Similarity or Similarity & Rule. Participants in the Rule + Similarity condition were asked to diagnose each case twice. On the first pass, they were asked to approach the task using the diagnostic rules listing all clinical features, and then list their primary diagnosis and any differentials being considered. On the second pass, they were asked to reconsider their original primary diagnosis using similarity between cases to reach their final diagnosis. They were reminded that similarity is usually beneficial as similar cases are often in the same diagnostic category. Participants in the Similarity & Rules condition were told that both similarity to previously seen cases and the rules of diagnosis are integral to the diagnostic process and to use whichever they feel is necessary. They were told that for some cases similarity may be particularly effective, while for others, rules may be more effective and to consider their diagnosis with both strategies. Overall time on task was similar for the two conditions. Participants in each experimental condition completed the test phase of the study in approximately 45 minutes. Similar to the first study, a diagnosis was scored as correct if it was listed as the primary diagnosis.

Results and Discussion

Table 2 and Figure 3 depict the mean proportion of correct responses of participants in the two dual instruction conditions of the second study, Rule + Similarity, and Similarity & Rules.

Rule + Similarity condition. A repeated measures ANOVA on accuracy for the Rule + Similarity condition with pass (1, 2), typicality (typical, atypical), and similarity (similar, dissimilar) only yielded a main effect of typicality, $F(1, 7) = 17.00$, MSe = 0.01, $p = .001$, $\eta^2 = .21$ and similarity, $F(1, 7) = 19.33$, MSe = 1.06, $p = .01$, $\eta^2 = .98$. No effect of pass was found, $F(1, 7) = 1.34$, MSe = 0.004, $p = .19$, indicating that once a case has been assessed with the

Table 2
Study 2 Mean Diagnostic Accuracy (and Standard Deviation) as a Function of Case Type and Instructional Conditions

				Test				
		Training		Typical		Atypical		
Instruction	# S	Typical	Atypical	Similar	Dissimilar	Similar	Dissimilar	All cases
Rule + similarity pass 1	8	0.53 (0.26)	0.40 (0.15)	0.75 (0.12)	0.18 (0.21)	0.45 (0.27)	0.40 (0.14)	0.44 (0.16)
Rule + similarity pass 2	8			0.75 (0.12)	0.40 (0.21)	0.43 (0.27)	0.38 (0.13)	0.49 (0.14)
Similarity & rule	14	0.54 (0.18)	0.39 (0.17)	0.86 (0.14)	0.67 (0.21)	0.31 (0.27)	0.39 (0.12)	0.56 (0.08)

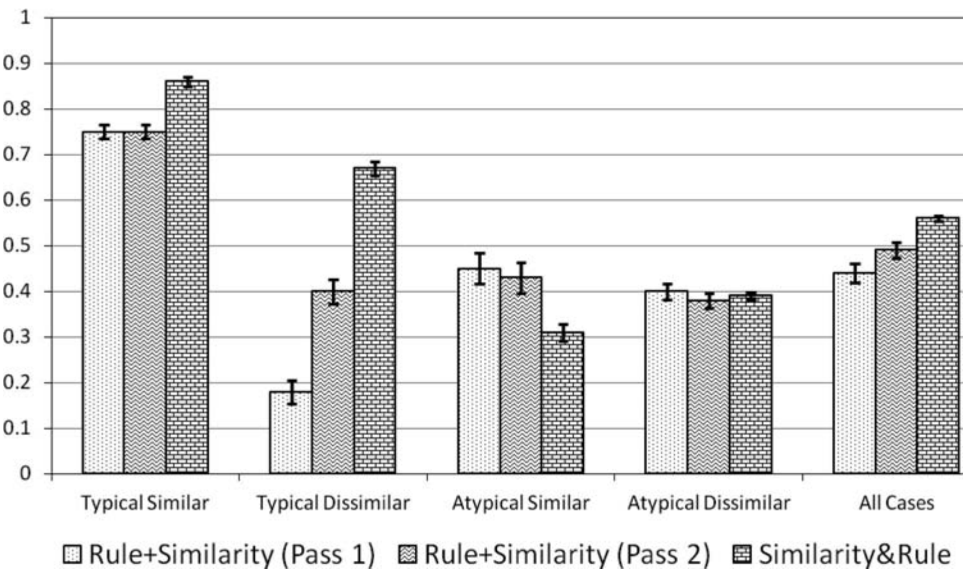


Figure 3. Study 2: Mean diagnostic accuracy as a function of case type and instructional conditions.

clinical rules, it is likely not possible to assess it anew based on similarity. In an analysis of the average change score for atypical-similar cases (pass 1 = 0.45, pass 2 = 0.43) where one might expect to have the greatest gain by reassessing with similarity, no such gain was found $t(7) = 1.0$, $p = .35$.

Similarity & Rules condition. To determine whether committing to a diagnosis on paper based on similarity alone results in an anchoring effect, the diagnostic accuracy under the S + R condition of Study 1 was compared with diagnostic accuracy under Similarity & Rules condition. A three-way mixed analysis of variance with test condition (S + R, Similarity & Rules) as a between-subjects factor and typicality (typical, atypical) and similarity (similar, dissimilar) as within-subject factors did not reveal a main effect of condition $F(1, 33) = 0.09$, MSe = 0.04, $p = .77$, indicating similar rates of accuracy in the two conditions. Thus, using similarity to generate a diagnosis, which may be checked with the rules of diagnosis concurrently, does not appear to result in an anchoring effect. This analysis highlighted interesting differences between participants in the two conditions in performance on the different types of cases. The analysis revealed not only a main effect of typicality, $F(1, 33) = 9.36$, MSe = 0.03, $p < .001$, $\eta^2 = .39$ and similarity, $F(1, 33) = 59.18$, MSe = 0.04, $p < .001$, but also significant two-way interactions of condition by typicality $F(1, 33) = 14.38$, MSe = 0.03, $p < .001$, $\eta^2 = .32$ and condition by similarity $F(1, 33) = 5.5$, MSe = 0.03, $p < .05$ as well as a three-way interaction of condition by typicality by similarity $F(1, 33) = 5.5$, MSe = 0.3, $p < .05$. Post hoc analyses (Newman–Keuls) indicated that while participants in the Similarity & Rules condition (76%) performed significantly better on typical cases than participants in the S + R condition (66%), $p < .05$, participants in the S + R condition (48%) performed significantly better on atypical cases than participants in the Similarity & Rules condition (35%), $p < .01$. Conversely, while participants in the S + R condition (80%) performed significantly better on similar items than their peers in the Similarity & Rules condition (59%), $p < .001$, on dissimilar items, those in the Similarity & Rules condition (53%) performed significantly better than those in the S + R condition (33%). The pattern of performance indicates that the S + R strategy resulted in greater reliance on similarity-based processing, while the Similarity & Rules strategy resulted in greater reliance on rule-based processing. As with Study 1, while the instructions did not differentially affect the overall diagnostic accuracy, it did differentially affect the type of information used to reach diagnosis. This finding is also consistent with the results of the first study in which novice clinicians who were given no instructions at test adopted a more rule-based approach to diagnosis.

Similarity & Rules and Rule + Similarity conditions. Lastly, the performance of participants in the two conditions introduced in Study 2 was compared. A three-way mixed analysis of variance with test condition (Rules + Similarity, Similarity & Rules) as a between-subjects factor and typicality (typical, atypical) and similarity (similar, dissimilar) as within-subject factors resulted in a main effect of typicality $F(1, 20) = 54.22$, MSe = 0.03, $p < .0001$, $\eta^2 = .59$, a main effect of similarity $F(1, 2) = 9.70$, MSe = 0.03, $p < .01$, $\eta^2 = .11$ and a typicality by similarity interaction $F(1, 20) = 13.90$, MSe = .03, $p < .01$. No other effects were significant. Again, although there was no difference in overall accuracy, the pattern of performance across the four types of cases differed significantly between the two conditions. Post hoc analyses (Newman–Keuls) found that for typical cases participants in the Similarity & Rules condition (0.76) did significantly better than participants in the Rules + Similarity condition (0.58%), $p < .001$. Although this finding may, at first, be surprising, there may be an explanation. The participants in the Similarity & Rules were free to choose whichever strategy they deemed appropriate. Preclinical medical students who are given no instructions were shown to adopt a primarily analytic approach. Thus, if given the choice, as is the case in the Similarity & Rules condition, participants adopted a more rule-based approach. This explanation, however, does not address why participants in the Rule + Similarity condition did not use typicality to a greater extent. One

possibility is that their use of the rules of diagnosis may have been tempered by knowledge that they will be using similarity on the second pass.

The Rule + Similarity condition contained only eight participants, thus the null effects observed may be attributable to a lack of statistical power. A rough power analysis using MSe of 0.015 suggests that we would be able to detect a difference in accuracy of the order of 9%. Further study with a larger student sample may provide more definitive conclusions. The second possible shortcoming of this study is the cross-study comparisons of the S + R condition of Study 1 with the Similarity & Rules condition of Study 2. Because of the violation of random allocation, one could argue that the participants in the two studies may have differed on one or more dimensions. However, effort was taken to ensure that the participants in the two studies were similar. Further, all other details about the execution of study were held constant. The results of Study 2, though tentative, indicate that how analytic and similarity-based processing is combined may have an effect on the task of diagnosis in medical students.

Summary and Concluding Discussion

Implications

This research illustrates that instruction given to learners at the time of test can shift the balance of both analytic and similarity-based processing, a finding with some practical consequences. Much of the literature in medical education focuses on ways to improve clinical performance through designing and implementing more effective training (e.g., Fraser, 2003; Hatala, Brooks, Norman, 2003; Ogrinc et al., 2003) and is reminiscent of earlier research in psychology focusing on the impact of differential instruction and training on subsequent performance (e.g., Medin, Alton, & Murphy, 1984; Whittlesea, 1987). In both the psychology and medical education literature, little attention has been given to instruction that may be given *after* training to improve performance. If instruction is given, it has usually taken the form of analytic instructions. Medical educators and those writing professional practice guidelines (Eisenberg, 1986), for example, have traditionally espoused an exhaustive search of all relevant information. In the present studies, training was held constant and only instructions given at test were varied. The results indicate that such instruction has an impact on the manner in which students approached specific cases; however, there was no overall difference. Those who were given a rationale for using similarity performed at par with their peers who were asked to use a rule-based strategy. This finding is consistent with previous research (Ark, Brooks, & Eva, 2006; Kulatunga-Moruzi, 2005; Kulatunga-Moruzi, Brooks, and Norman, 2001; Regehr et al., 1994) and extends several studies illustrating the strong impact of similarity in the diagnostic process in both visual (Allen, Norman, & Brooks, 1992) and nonvisual domains (Hatala, Norman, & Brooks, 1999a, 1999b; Young, Brooks, & Norman, 2007). Further, this finding is consistent with observational studies directly comparing the use of similarity-based processing against other modes of processing that found similarity-based processing to be effective at several levels of expertise (Coderre, Mandin, Harsyn, & Fick, 2003). Because of the pseudoexperimental design of that study, however, it was not possible to compare diagnostic strategies on identical cases. Diagnostic strategy may be confounded with case difficulty (or case type, typical/atypical) leading to the observed results. The experimental design of the studies reported here allow for a more direct comparison of these two processes, suggesting that similarity-based processing should not simply be viewed pejoratively as rudimentary and error-prone pattern recognition (cf. Croskerry, 2002, 2006; Klein, 2005; Redelmeier, 2001; Redelmeier, Shafir, & Aujla, 2001; Redelmeier, Tan, & Booth, 1998).

The role of similarity-based processing is not well addressed in current medical pedagogy. When learning about various diseases, the primary focus is on the identification of clinical feature in cases that present with the diagnostic rules. As such, the discussion of a disease condition is often accompanied by one or, at best, a few representative examples. Given the essential role of similarity-based processing in clinical diagnosis, early medical training could be carefully structured to expose students to a variety of exemplars, both typical and atypical. Using a variety of presentations of the same disease condition will allow learners to begin building a repertoire of cases early in their medical training, which will prove useful in clinical settings. A large and more diverse set of exemplars in training has been shown to increase transfer (Homa & Cultice, 1984). Moreover, such presentation will allow students to learn the variability in presentation, at both the global and feature levels (see Kulatunga-Moruzi, Brooks, & Norman, 2004), which may provide a source of perceptual matching (see Brooks & Hannah, 2006; Hannah & Brooks, 2006) and allow for contrastive learning (Ark, Brooks, & Eva, 2007; Coderre et al., 2003). Given that similarity may be a powerful initial source of hypotheses, which may then be assessed analytically, the role of each of these processes must be taken into consideration in medical pedagogy.

A critical finding of this research that may inform medical training is that preclinical medical students who use a dual strategy in which similarity is used first or concurrently with the rules of diagnosis resulted in a consistent improvement over either single strategy. Performance under a dual strategy in which the rules of diagnosis are considered first, however, did not improve students' diagnostic performance. Although further research is necessary to definitively indicate why this may be so, the pattern of performance across the four different types of cases indicates that it may be attributable to the inability to use similarity once a strict rule-based mode of operation has been adopted.

One conjecture based on these results is that similarity may be particularly critical in the early stages of diagnostic reasoning, during hypothesis generation. While the two modes of operation are not mutually exclusive in that perceptual matching may be done on clinically relevant features, examination of the use of similarity on atypical cases, which contain few features of their diagnostic category, indicates matching on nondiagnostic but contextually correlated features. The generation of a diagnostic hypothesis based on a clinically relevant feature may be arrived at by either the use of similarity or the rules of diagnosis. If similarity is used, the rules may be critical in testing this initial hypothesis, generating alternate hypotheses as needed. Similarly, an initial diagnosis based on the rules of diagnosis may be checked by the deployment of similarity. Although there is little evidence for the latter in our study, this possibility is logical and perhaps may be seen more readily in experienced clinicians who have a larger repertoire of cases upon which to base a diagnosis. For the experienced diagnostician, then, similarity may be a way of reducing

the oversimplification inherent in causal models. This conception of the diagnostic process may be viewed as an interplay of the two processes, similar to the "top-down, bottom up" processes seen in visual attention and perception. This more flexible conception of the diagnostic process that uses the generation and testing of diagnostic hypotheses by using both similarity between cases and the rules of clinical practice is contrary to that of linear forms of diagnosis (e.g., Bayesian models, forward reasoning models) and is consistent with other research (Kulatunga-Moruzi, Brooks, & Norman, 2004; Norman et al., 2000). These results suggest that neither exemplar-based nor analytic processes can be considered a "fallback" approach that is only used when the other process is deemed insufficient for the purposes of diagnosis. Rather, both processes seem to provide concurrent support for diagnostic decisions.

Limitations

A possible concern for the generalizability of these studies stems from the fact that the participants in our study were preclinical students with no instruction or experience in dermatology, which calls into question whether diagnostic strategy may be as readily influenced in more experienced clinicians. The research reported by Regehr and colleagues (1994) using the same experimental paradigm with family and internal medicine residents illustrated the use of both types of processes and provided some indication of the effect of test-time instruction on the balance of analytic and similarity-based processes. Further, while it may be difficult to influence diagnostic strategy as readily with experienced clinicians, who promptly generate diagnostic hypotheses upon case presentation, the performance of both family practitioners and dermatologists has been shown to be influenced by forcing a more analytic strategy than they would normally adopt (Kulatunga-Moruzi, Brooks, & Norman, 2004).

Future Directions

While this research contributes to the body of literature aimed at understanding the complexity of the diagnostic process and a number of variables have been identified that influence the relative use of these two processes (see McLaughlin, Rikers, & Schmid, 2008 for review), a question that needs to be addressed is how and when a clinician knows that his or her "usual balance" of rules and similarity may have resulted in an incorrect diagnosis. A fascinating topic of investigation, then, is this transition. A true expert as defined by Bereiter and Scardemalia (1993) is one who engages "in the integration and coordination of similarity-based, automatic resources and more effortful reflective processes" (Moulton, Regehr, Mylopolous & MacRae, 2007 p. 110) and can make the transition with ease, as the task requires it. Some thoughtful research has been pursued documenting this transition in surgery (Moulton, Regehr, Lingard, Merritt, and MacRae, 2009, 2010a, 2010b). Studying this shift across other disciplines in medicine will not only yield important aspects of diagnostic expertise but is also likely to yield the most benefit in understanding diagnostic errors.

Another issue that needs to be addressed is the definition of similarity-based processing. In these studies we define it as the retrieval of and analogy to a similar case in training, where similarity is based on both clinically relevant and irrelevant features. The detection of similarity, here, is more likely conscious. For experienced clinicians, however, similarity-based processing may take on a different meaning, namely the unconscious analogy to a previously encountered case. It is possible that both of these types of similarity-based processes are likely to play a significant role in diagnosis. Further, the fact that matching may be done on relevant features blurs the distinction between what is traditionally thought of as similarity-based and analytic processing. The current medical cognition literature does not address these issues. Teasing apart these aspects of similarity-based processing is also likely to yield benefits in understanding the true nature of medical expertise and diagnostic errors.

References

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General, 120,* 3–19.

Allen, S. W., Norman, G. R., & Brooks, L. R. (1992). Experimental studies of learning dermatologic diagnosis: The impact of examples. *Teaching and Learning in Medicine, 4,* 35–44.

Ark, T., Brooks, L. R., & Eva, K. W. (2006). Giving learners the best of both worlds: Do clinical teachers need to guard against teaching pattern recognition? *Academic Medicine, 81,* 405–409.

Ark, T., Brooks, L. R., & Eva, K. W. (2007). The benefits of flexibility; the pedagogical value of instructions to adopt multifaceted diagnostic reasoning. *Medical Education, 41,* 281–287.

Ashman, C. J., Yu, J. S., & Wolfman, D. (2000). Satisfaction of search in osteoradiology. *American Journal of Roentgenology, 175,* 541–544.

Barrows, H. S., Norman, G. R., Neufeld, V. R., & Feightner, J. W. (1982). The clinical reasoning of randomly selected physicians in general medical practice. *Clinical and Investigative Medicine, 5,* 49–55.

Berbaum, K. S., El-Khoury, G. Y., Franken, E. A., Kathol, M., Montgomery, W. J., & Hesson, W. (1988). Impact of clinical history on fracture detection with radiography. *Radiology, 168,* 507–511.

Berbaum, K. S., Franken, E. A., Dorfman, D. D., Barloon, T., Ell, S. R., Lu, C. H., . . . Abu-Yousef, M. M. (1986). Tentative diagnoses facilitate the detection of diverse lesions in chest radiographs. *Investigative Radiology, 21,* 532–539.

Berbaum, K. S., Franken E. A., Jr., Dorfman, D. D., Rooholmanini, S. A., Kathol, M. H., Barloon, T. J., . . . El Khoury, G. Y. (1990). Satisfaction of search in diagnostic radiology. *Investigative Radiology*, 25, 133–140.

Bereiter, C., & Scardamalia, M. (1993). The need to understand expertise. In surpassing ourselves: An enquiry into the nature and implications of expertise. Chicago: Open Court; 1–24.

Bordage, G. E. (1994). Elaborated knowledge: A key to successful diagnostic thinking. *Academic Medicine, 69,* 883–885.

Bordage, G. E., & Zacks, R. (1984). The structure of medical knowledge in the memories of medical students and general practitioners: Categories and prototypes. *Medical Education, 18,* 406–416.

Brooks, L. R., & Hannah, S. D. (2006). Instantiated features and the use of "rules." *Journal of Experimental Psychology: General, 135,* 133–151.

Brooks, L. R., LeBlanc, V. R., & Norman, G. R. (2000). On the difficulty of noticing obvious features in patient appearance. *Psychological Science, 11,* 112–117.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General, 120,* 278–287.

Coderre, S., Mandin, H., Harasym, P. H., & Fick, G. H. (2003). Diagnostic reasoning strategies and diagnostic success. *Medical Education, 37,* 695–703.

Croskerry, P. (2002). Achieving quality in clinical decision making: Cognitive strategies and detection of bias. *Academic Emergency Medicine, 9,* 1184–1204.

Croskerry, P. (2006). Critical thinking and decision making: Avoiding the perils of thin-slicing. *Annals Emergency Medicine, 48,* 720–722.

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine*. New York; Free Press.

Eisenberg, J. M. (1986). *Doctor's decisions and the cost of medical care: The reasons for doctors' practice patters and how to change them*. Ann Arbor (MI): Health Administration Press.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1990). Medical problem solving: A ten year retrospective. *Evaluation and the Health Professions, 13,* 5–36.

Fraser, J. (2003). Teaching practical procedures in general practice: A primer for supervisors of medical students and registrars. *Australian Family Physician, 32,* 540–543.

Hannah, S. D., & Brooks, L. R. (2006). Producing diagnostic biasing with unambiguous stimuli: The role of instantiated features. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32, 1416*–1423.

Hatala, R., Brooks, L. R., & Norman, G. R. (2003). Practice makes perfect: The critical role of mixed practice in the acquisition of ECG interpretation skills. *Advances in Health Sciences Education Theory Practice, 8,* 17–26.

Hatala, R. M., Norman, G. R., & Brooks, L. R. (1999a). The impact of a clinical scenario upon electrocardiographic interpretation. *Journal of General Internal Medicine, 14,* 126–129.

Hatala, R. M., Norman, G. R., & Brooks, L. R. (1999b). Influence of a single example upon subsequent electrocardiogram interpretation. *Teaching and Learning in Medicine, 11,* 110–117.

Homa, D., & Cultice, J. (1984). Role of feedback, category size and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 83–94.

Klein, J. G. (2005). Five pitfalls in decisions about diagnosis and prescribing. *British Medical Journal, 330,* 78–784.

Kulatunga-Moruzi, C. (2005). Investigating the diagnostic process. Ph.D. Dissertation. *ETD Collection for McMaster University*. Paper AAINR36087.

Kulatunga-Moruzi, C., & Brooks, L. R. (2001). *Clinical teaching: A picture is worth a thousand words.* Paper presentation at American Educational Research Association Seattle, WA.

Kulatunga-Moruzi, C., Brooks, L. R., & Norman, G. R. (2004). Using comprehensive feature lists to bias medical diagnosis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 563–572.

Kulatunga-Moruzi, C., Brooks, L. R., Norman, G. R., & Scully, K. (2002, July). The good and bad: The utility of verbal information. Paper presentation at the 10th Ottawa Conference, Ottawa, Ontario.

Kulatunga-Moruzi, C., Brooks, L. R., & Norman, G. R. (2001). Coordination of analytical and similarity based processing strategies and expertise in dermatological diagnosis. *Teaching and Learning in Medicine, 13,* 110–116.

LeBlanc, V. R., Brooks, L. R., & Norman, G. R. (2002). Believing is seeing: The influence of a diagnostic hypothesis on the interpretation of clinical features. *Academic Medicine, 77*(10), S67–S69.

Mamede, S., Schmidt, H. G., & Penaforte, J. C. (2008). Effects of reflective practice on the accuracy of medial diagnoses. *Medical Education, 42,* 468–475.

Mamede, S., Schmidt, H. G., Penaforte, J. C., & Coelho-Filho, J. M. (2007). Breaking down automaticity: Case ambiguity and the shift to reflective approaches in clinical reasoning. *Medical Education, 41,* 1185–1192.

Mamede, S., Schmidt, H. G., Rikers, R. M., Custers, E. J., Spliner, T. A., & van Saase, J. L. (2010a). Conscious thought beats deliberation without attention in diagnostic decision-making: At least when you are an expert. *Psychological Research, 74,* 586–592.

Mamede, S., van Got, T., Rikers, R. M., van Saase., J. L., van Guldener, C., & Schmidt, H. G. (2010b). Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *Journal of the American Medical Association, 304,* 1198–1203.

McGaghie, W. C., Boerger, R. L., McCrimmon, D. R., & Ravitch, M. M. (1996). Learning pulmonary physiology: Comparison of student and faculty knowledge structures. *Academic Medicine, 71,* S13–S15.

McLaughlin, K., Rikers, R. E., & Schmidt, H. (2008). Is analytic information processing a feature of expertise in medicine? *Advances in Health Sciences Education, 13,* 123–128.

Medin, D. L., Altom, M. W., Edelson, S. M., & Frecko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 37–50.

Medin, D. L., Altom, M. W., & Murphy, T. D. (1984). Given versus induced category representations: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 333–352.

Moulton, C., Regehr, G., Lingard, L., Merritt, C., & MacRae, H. (2009). Operating from the other side of the table: Control dynamics and surgeon educator. *Journal of American Colleges of Surgeons, 210,* 79–86.

Moulton, C., Regehr, G., Lingard, L., Merritt, C., & MacRae, H. (2010a). Slowing down when you should: Initiators and the influences of the transition from the routine to the effortful. *Journal of Gastrointestinal Surgery, 14,* 1019–1026.

Moulton, C., Regehr, G., Lingard, L., Merritt, C., & MacRae, H. (2010b). Slowing down to stay out of trouble in the operating room: Remaining attentive in automaticity. *Academic Medicine, 85,* 1571–1577.

Moulton, C. E., Regehr, G., Mylopolous, M., & MacRae, H. M. (2007). Slowing down when you should: A new model of expert judgement. *Academic Medicine, 82,* S109–S116.

Neufeld, V. R., Norman, G. R., Barrows, H. S., & Feightner, J. W. (1981). Clinical problem solving by medical students: A longitudinal and cross-sectional analysis. *Medical Education, 15,* 315–322.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2,* 175–220.

Norman, G. R., Brooks, L. R., Allen, S. W., & Rosenthal, D. (1990). Sources of observer variation in dermatologic diagnosis. *Academic Medicine, 65,* S19–S20.

Norman, G. R., Brooks, L. R., Coblentz, C. K., & Babcook, C. J. (1992). The interdependence of feature and category identification in diagnostic radiology. *Memory & Cognition, 4,* 344–355.

Norman, G. R., Brooks, L. R., Colle, C. L., & Hatala, R. M. (2000). The benefit of diagnostic hypotheses in clinical reasoning: Experimental study of an instructional intervention for forward and backward reasoning. *Cognition and Instruction, 17,* 433–448.

Ogrinc, G., Headrick, L. A., Mutha, A. S., Coleman, M. T., O'Donnell, J., & Miles, P. V. (2003). A framework for teaching medical students and residents about practice-based learning and improvement, synthesized from a literature review. *Academic Medicine, 78,* 748–756.

Papa, F. J., & Elieson, B. (1993). Diagnostic accuracy as a function of case prototypicality. *Academic Medicine, 68,* S58–S60.

Patel, V. L., & Groen, G. J. (1986). Knowledge-based solution strategies in medical reasoning. *Cognitive Science, 10,* 91–116.

Patel, V. L., Groen, G. J., & Arocha, J. F. (1990). Medical expertise and a function of task difficulty. *Memory & Cognition, 18,* 394–406.

Redelmeier, D. A., Ferris. L. E., Hux, J. E., & Schull, M. J. (2001). Problems for clinical judgment: Introducing cognitive psychology as one more basic science. *Canadian Medical Association Journal, 164,* 358–360.

Redelmeier, D. A., Shafir, E., & Aujla, P. (2001). The beguiling pursuit of more information. *Medical Decision Making, 21,* 376–381.

Redelmeier, D. A., Tan, S. H., & Booth, G. L. (1998). The treatment of unrelated disorders in patients with chronic medical diseases. *New England Journal of Medicine, 338,* 1516–1520.

Regehr, G., & Brooks, L. R. (1994). Perceptual manifestations of an

analytic structure: The priority of holistic individuation. *Journal of Experimental Psychology: General, 122,* 92–114.

Regehr, G., Cline, D. J., Norman, G. R., & Brooks, L. R. (1994). Effect of processing strategy on diagnostic skill in dermatology. *Academic Medicine, 69,* S34–S36.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 3,* 382–439.

Schmidt, H. G., Boshuizen, H. P. A., & Norman, G. R. (1992). Reflections on the nature of expertise in medicine. In E. Keravnou (Ed.), *Deep models for medical knowledge engineering.* Elsevier Science Publishers B. V.

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

Schon, D. A. (1983). *Educating the reflective practitioner: How professionals think in action*. New York: Basic Books.

Schon, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. San Francisco, CA: Jossy-Bass.

Weinstein, M. D., & Fineberg, H. U. (1980). *Clinical decision analysis.* Philadelphia, PA: Saunders.

Whittlesea, B. W. A. (1987). Preservation of specific experiences in the representation of general knowledge. *Journal of Experimental Psychology: Learning Memory & Cognition, 13,* 3–17.

Wigton, R. S. (1988). The use of linear models to analyse physicians' decisions. *Medical Decision Making, 8,* 241–252.

Young, M., Brooks, L., & Norman, G. (2007). Lost in translation: The impact of familiar descriptions on diagnosis in novices. *Medical Education, 41,* 146–1151.

Appendix A

Example of Instructional Phase Material

Lichen Planus

Lichen planus (LP) is a pruritic (itchy), papular (raised flat lesions) eruption characterized by its violaceous color; polygonal shape; often with overlying white lines (known as Wickham's striae) and fine scale. The lesions occur in linear, annular, or confluent groups.

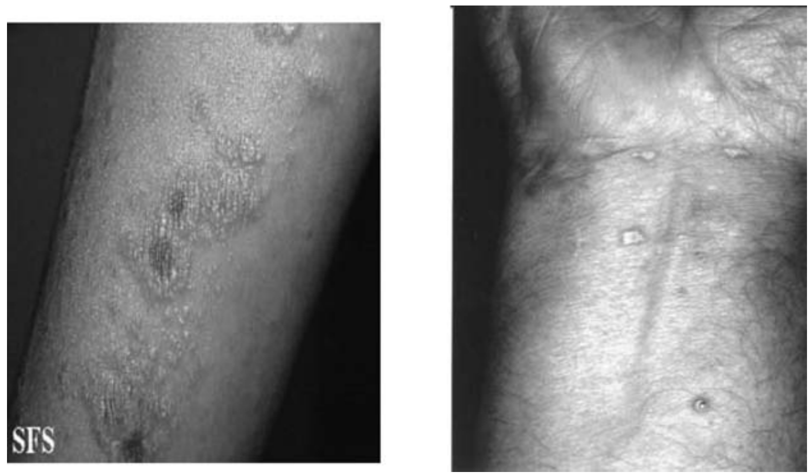
It is most commonly found on the flexor (elbows, wrists, knees, ankles), lumbar region (lower back), shins, genitalia, and on the mucous membranes.

Lesions of the mucous membranes appear as whitish, reticulated, lacy plaques.

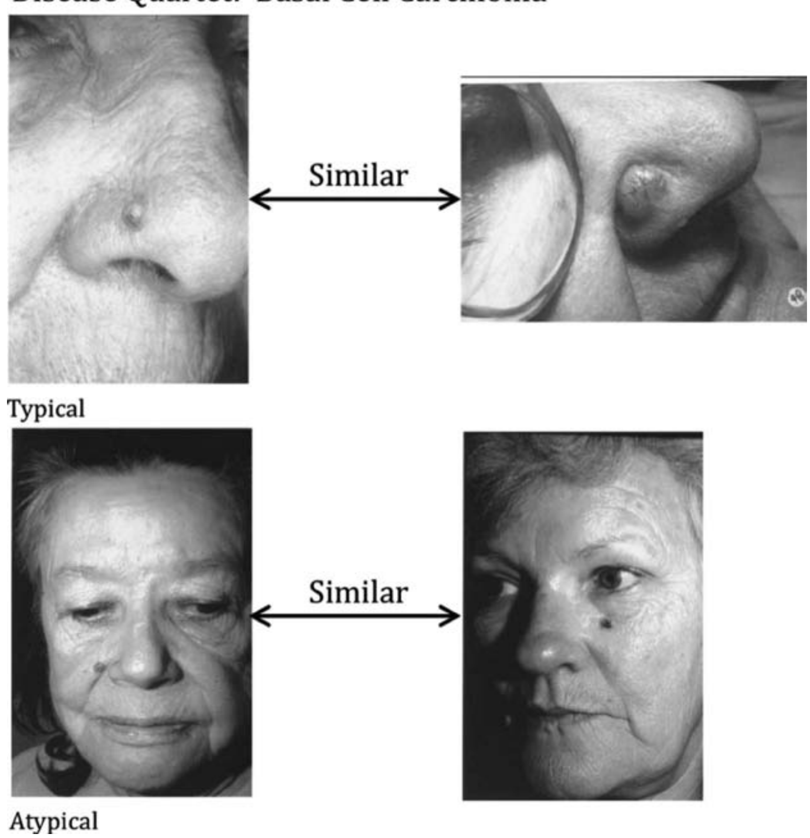
Females are more frequently affected than males, and the age range is approximately 30 to 60 years of age.

The cause of lichen planus is unknown. It is most likely an immunologically mediated reaction.

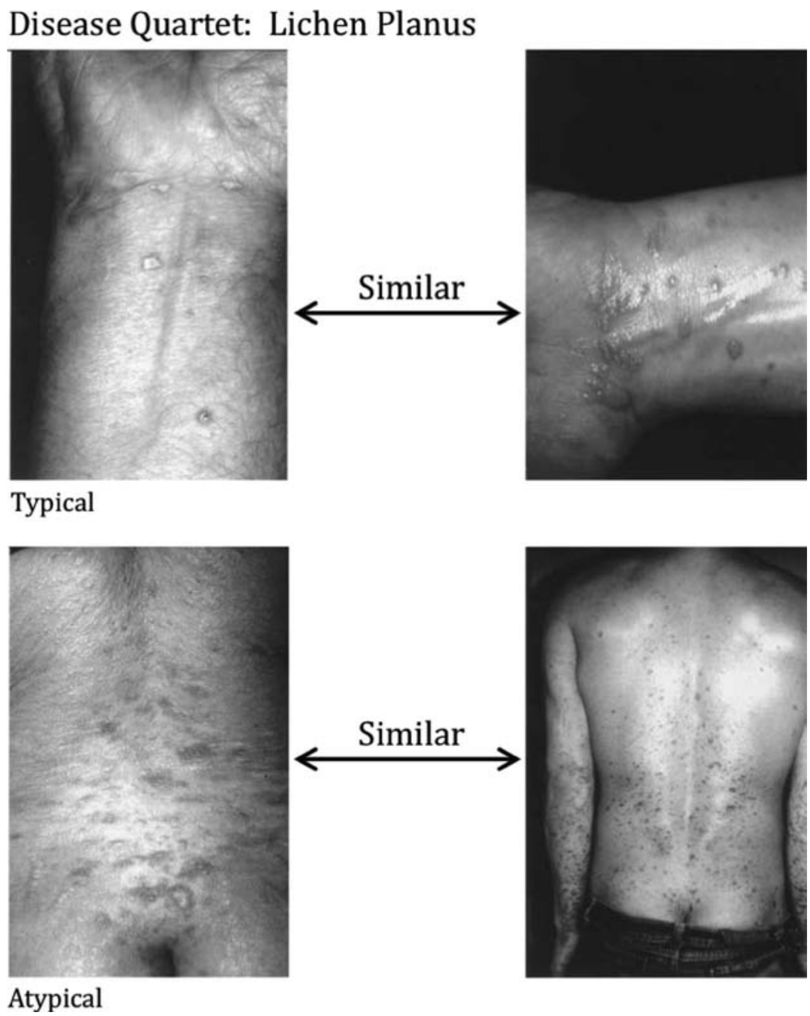
(*Appendix continues*)

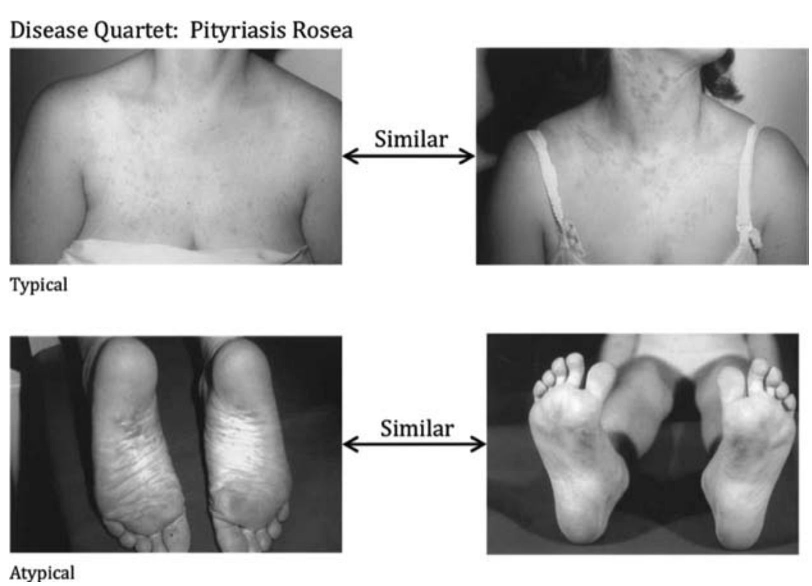


Disease Quartet: Basal Cell Carcinoma



(*Appendix continues*)





Received May 19, 2010
Revision received March 29, 2011
Accepted April 5, 2011 ■