

Giving Learners the Best of Both Worlds: Do Clinical Teachers Need to Guard Against Teaching Pattern Recognition to Novices?

Tavinder K. Ark, MSc, Lee R. Brooks, PhD, and Kevin W. Eva, PhD

Abstract

Purpose

There has been much debate in the medical education literature regarding the extent to which feature-driven and nonanalytic (similarity-based) reasoning strategies define expertise, but the relative value of teaching these strategies, together or in isolation, remains uncertain. The purpose of this study was to compare the diagnostic accuracy achieved upon receiving instruction to use each strategy in isolation to that of a combined approach.

Method

In 2003–04, 48 undergraduate

psychology students from McMaster University in Ontario, Canada, were taught to diagnose ten cardiac disorders (including normal) via electrocardiogram (ECG) presentation. Twelve students were instructed to carefully identify all features present before assigning a diagnosis (feature first). Twelve were given the same instruction with notice that some test ECGs had been seen during training (implicit combined). Twelve were simply instructed to trust familiarity and diagnose based on this impression (similarity-based). Finally, 12 students were given feature first and similarity-based instructions in combination (explicit combined).

Results

No difference in diagnostic accuracy was observed between the groups given the feature first (42%) and first impression (41%) instructions ($p > .4$), but the groups instructed to use both strategies (explicitly or implicitly) performed significantly better (56% and 53%, respectively; $p < .01$).

Conclusions

The results support an additive model of clinical reasoning in which instructions to be feature oriented and to trust similarity improve performance in novice diagnosticians.

Acad Med. 2006; 81:405–409.

To determine if a patient has suffered an anteroseptal myocardial infarction requires the identification of certain diagnostic features, such as chest discomfort consistent with angina and ST elevations in leads V1 to V4 in an electrocardiogram (ECG). Research in both cognitive psychology and medical education proposes two classes of mechanisms for performing this task: analytic and nonanalytic. Both of these classes have been operationalized in many ways and should not be considered mutually exclusive.^{1,2} In fact, we recently

reviewed the literature on clinical reasoning and were led to the conclusion that clinical teachers can best serve their students by viewing expertise as the ability to use numerous approaches to solving diagnostic problems.¹ In this article, we provide direct evidence in support of this conclusion and direction to clinical teachers regarding instructional tactics that can enhance the diagnostic accuracy of their learners.

Analytic processes are those that entail controlled, systematic consideration of features and their relation to potential diagnoses. This is the form of clinical reasoning that has traditionally been endorsed by educators charged with teaching medical students the diagnostic process. Analytic approaches tend to be invoked with admonitions to carefully identify and consider all clinical features before generating a diagnostic hypothesis, or to follow a specific diagnostic algorithm when considering a novel case. Such instructions arise from concern about premature closure (i.e., failure to consider all diagnostic possibilities), the need to provide students with a reliable diagnostic strategy, and a desire to emphasize the evidence-based nature of medical care. Broadly, analytic processes

are believed to reduce biases that can arise upon considering a case with a specific diagnosis in mind.

In the last two decades, however, medical educators have empirically demonstrated the nonanalytic basis of clinical reasoning.³ Central to this view is the argument that rapid processes such as pattern recognition provide a valid alternative mechanism to informing diagnostic decision making. That is, one might automatically recognize the correct diagnosis simply because the current case is similar to one that has been seen in the past. Such activation typically occurs unconsciously.⁴

It is now broadly recognized that this form of reasoning nicely describes much (but certainly not all) of the activity in which both experts (having acquired a vast repertoire of cases) and novices engage, process differences having failed to capture the essence of expertise.^{5,6} Still, uncertainty remains regarding how, and if, nonanalytic processes can be operationalized and invoked to the benefit of learners.⁷ The study we report in this article was designed to assess the potential strengths/weaknesses of adopting a multifaceted approach to

Ms. Ark is research associate, Program for Educational Research and Development, McMaster University, Hamilton, Ontario, Canada. At the time of this study, Ms. Ark was an undergraduate thesis student in the Department of Psychology, McMaster University, working under the supervision of Dr. Eva.

Dr. Brooks is professor, Department of Psychology, McMaster University, Hamilton, Ontario, Canada.

Dr. Eva is assistant professor and associate chair, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada.

Correspondence should be addressed to Dr. Eva, Program for Educational Research and Development, MDCL 3522, McMaster University, 1200 Main Street West, Hamilton, Ontario, Canada, L8N 3Z5; telephone: (905) 525-9140 x27241; fax: (905) 572-7099; e-mail: (evakw@mcmaster.ca).

clinical instruction when training absolute novices.

Addressing this question is important, because while it is well documented that reasoning from a diagnosis can bias one's interpretation of the features present in a clinical case,⁸ it must also be recognized that there are drawbacks to reviewing cases in a more systematic and nonbiased manner. Norman et al.⁹ for example, taught undergraduate psychology students to diagnose ECGs in a feature-driven manner by having them systematically identify features before considering diagnostic possibilities. Students in this condition identified more features than did those who were told to generate a diagnostic hypothesis prior to their feature search, but the additional features that were generated tended to be irrelevant to the correct diagnosis. As a result, the diagnostic performance of participants that were told to delay making a diagnosis was 10% to 20% less accurate than that of participants who biased themselves by following instructions to generate a diagnosis before systematically considering the features.

Norman et al.'s study provided a first opportunity to assess the advantage of instruction to adopt a multifaceted (i.e., combined reasoning strategies) approach because the authors essentially compared instruction in which learners were told to be nonanalytic (i.e., derive a Gestalt impression of the diagnosis) and systematically (i.e., analytically) consider the features present, to an approach in which the analytic instructions were given in isolation. We built our study on this result in two ways. First, we created a more purely nonanalytic instruction condition to determine whether combined instruction truly provides diagnostic advantage in this domain relative to instruction to adopt either an analytic or nonanalytic strategy in isolation. That is, can absolute novices take advantage of the best that both worlds have to offer? Second, we operationalized the combined instruction three different ways:

- by having participants complete a nonanalytic and analytic consideration of each ECG sequentially (i.e., providing a diagnostic decision before carrying out the more analytic search);

- by explicitly telling learners to trust feelings of familiarity in addition to carefully considering features; and
- by implicitly informing participants about the benefits of similarity by warning them that some test cases were seen during training.

Differences across these conditions, we anticipated, would provide guidance regarding how educational prescriptions should be provided should a multifaceted instruction strategy prove beneficial.

Method

Participants

Forty-eight undergraduate psychology students enrolled at McMaster University in Ontario, Canada in 2003–04 participated in this study for course credit. We selected this population because the debate in the medical education literature has focused on the role different diagnostic approaches should play in the training of novices. Using psychology students ensured that none of the participants had had previous experience with ECGs. The literature on which our work builds has routinely involved both medical trainees and lay participants with good reason to believe that findings are generalizable between the two populations.¹⁰ We received ethics approval from the Hamilton Health Sciences Research Ethics Board.

Design and procedure

Training phase. Learning took place during one-on-one training sessions with one of the authors (TKA). All participants were presented with general information regarding the 12 leads in an ECG and were taught to read a normal ECG waveform using materials created for teaching medical students.⁹ TKA taught participants each of ten diagnostic categories (including normal) using a feature list containing the key features for each diagnosis. For example, Right Bundle Branch Block was presented as typically having RSR' (rabbit ears) present in V1 and V2 and a widened QRS complex.¹¹

For each diagnostic category, participants were then presented with four examples in sequence. TKA identified the key features using the feature list for the first two ECGs in the category. Participants were then asked to identify the relevant

features on the next pair of ECGs before moving onto the next diagnostic category.

Practice phase. At the end of the training phase, participants were given a practice booklet of ten ECGs. During this phase, participants were randomly assigned to one of four experimental conditions:

- *Group 1: Similarity-based training.* To emphasize a nonanalytic reasoning strategy, participants in the similarity-based training condition were given the following instructions:

For each ECG, assign a diagnosis using similarity as a guide. New ECGs often look like ECGs that have been seen before (i.e., during training). Trust this sense of familiarity. Use the disease list to help you.

Once they had diagnosed all ten ECGs, they were asked to work through the ECGs a second time (Pass 2), this time identifying the features present in each ECG using a provided checklist. Participants were then asked to rediagnose each ECG, keeping their initial diagnosis in mind, but changing it if necessary.

- *Group 2: Feature-first training.* To emphasize an analytic reasoning strategy, participants assigned to the feature first training condition were given the following instructions:

For each ECG, work down the response sheet list and indicate all features that can be seen.

Only after this task was completed were participants asked to assign a diagnosis.

- *Group 3: Implicit combined.* Participants in the implicit combined training condition were given the same instructions as the feature first group. Distinct from the feature first instructions, participants in the implicit combined group were told that the ten ECGs in the practice phase were drawn randomly from the training book. This was an indirect (i.e., implicit) way to highlight the usefulness of similarity.

- *Group 4: Explicit combined.* Participants in the explicit combined training condition were given both the first impression and the feature first instructions. They were instructed:

For each ECG, assign a diagnosis using similarity as a guide. New ECGs often

look like ECGs that have been seen before (i.e., during training). Trust this sense of familiarity, but realize that basing decisions solely on similarity can lead to diagnostic errors. So, don't "jump the gun." Use the Response Sheet List to indicate the features that can be seen. Use the disease list to help you.

That is, participants were explicitly told to use similarity while simultaneously performing a careful consideration of the features present.

All participants were allowed to view the ECGs when assigning their diagnosis. Participants were told the correct answer after diagnosing each ECG (during Pass 2 for the similarity-based group). In addition, whenever an incorrect diagnosis was assigned, immediate feedback was given that reinforced the instructions provided.

Test phase. During the test phase, participants were asked to diagnose 20 ECGs, ten of which were novel and ten that had been seen during training. Participants were not given any feedback during the test phase, but otherwise, the procedure and reasoning instructions were identical to those of the practice phase for each participant.

Results

Diagnostic accuracy

Participants' mean diagnostic accuracy is reported in Table 1 by training condition and novelty of the stimuli. A repeated measures analysis of variance (ANOVA) was performed on these data with condition included as the between-subjects grouping factor. ECG (nested

within old/new) and old/new were included as within-subjects repeated measures. There was a significant main effect of training condition ($F_{3,792} = 4.98$, $p < .001$). Post hoc analyses revealed that the explicit combined group (mean = 56%) and the implicit combined group (53%) significantly outperformed both the feature first (42%) and the similarity-based (41%) groups ($F_{1,396} > 9.0$, $p < .01$ in all cases). There was no difference between the feature first and the similarity-based conditions, nor between the implicit combined and explicit combined conditions ($F_{1,396} < 1$, $p > .4$ in both cases).

Old ECGs (those that had been seen during training) were diagnosed more accurately, on average, than were new ECGs (those that participants had never seen) ($F_{1,792} = 46.94$, $p < .001$). The interaction between training condition and new/old cases was nonsignificant ($F_{3,792} < 1$, $p > .3$).

When the similarity-based group was given the opportunity to revise their diagnoses after completing the feature identification task (i.e., after completing a "sequential combined" task during Pass 2), their diagnostic accuracy increased to 58.3% for old ECGs and 42.5% for novel ECGs (50.4% overall) and was no longer significantly different from the explicit or implicit combined conditions.

Feature identification

For each ECG, the features identified by participants were categorized as: hit indicative, a feature that was present and indicative of the correct diagnosis for that particular ECG; hit not indicative, a

feature that was present in the ECG but not indicative of the correct diagnosis; or false alarm, a feature that was not present in the ECG. An ANOVA analogous to the one performed on diagnostic accuracy was performed on the number of features identified within each feature classification. The mean number of features identified by classification and training condition is shown in Table 2.

There was a significant main effect of the total number of features identified by training condition ($F_{3,792} = 69.1$, $p < .001$). Post hoc analyses revealed that participants in the feature first condition identified more features, on average, than did participants in any of the other three training conditions ($F_{1,396} > 16.0$, $p < .01$ for each comparison). The same pattern of results occurred for both hits not indicative ($F_{1,396} > 11.0$, $p < .001$ in all cases) and false alarms ($F_{1,396} > 65.0$, $p < .001$ in all cases). In contrast, no difference was observed in the mean number of hits indicative identified as a function of training condition. The main effect of old/new and the interaction between new/old cases and training condition were also nonsignificant in all feature analyses ($F_{3,792} < 1$, $p > .3$ in all cases).

Discussion

Like in many areas of science, there has been a tendency for theorists in medical education to strive to identify *the* truth: *the* hallmark of expertise and *the* ideal training technique. In reality, however, medical diagnosis is complicated enough that it is unlikely to ever yield one strategy that will provide a solution for all the problems clinicians will face, even for those who have specialized.^{1,2} The results of our study support an additive model of clinical reasoning in which instructions to be both featured-oriented and to use similarity-based reasoning strategies improved diagnostic performance relative to instructions to use either strategy in isolation. We observed the advantage of this combined approach when instruction to trust feelings of familiarity was given implicitly or explicitly and regardless of whether instruction to systematically consider the features presented in the case was given simultaneously with, or subsequent to, the instruction to generate a diagnostic decision.

Table 1

Diagnostic Accuracy of 48 Undergraduate Psychology Students Who Were Trained to Diagnose Electrocardiograms (ECGs), by Instruction Provided and Novelty of Stimulus, McMaster University, Ontario, Canada, 2003–04*

Training condition	Mean proportion of cases accurately diagnosed		
	Old ECGs [†]	New ECGs [‡]	Total
Similarity-based (Pass 1)	48.3	34.2	41.3
Feature first	48.3	35.8	42.1
Combined explicit	67.5	45.0	56.3
Combined implicit	64.2	42.5	53.3
Total	57.1	39.4	48.3

* Statistical significance ($p < .05$): (a) Combined explicit = Combined implicit > Feature first = Similarity-based, (b) Old ECGs > New ECGs, (c) No interaction between training condition and novelty.

[†] Old ECGs were test ECGs that were presented during the training phase.

[‡] New ECGs were test ECGs that had never been seen before by participants.

Table 2

Mean Number of Features Identified by 48 Undergraduate Psychology Students Who Were Trained to Diagnose Electrocardiograms (ECGs), by Feature Classification and Training Condition, McMaster University, Ontario, Canada, 2003–04*

Training condition	Mean no. of features identified			Total features
	Hits indicative [†]	Hits not indicative [‡]	False alarms [§]	
Similarity-based (Pass 2)	1.22	0.31	1.03	2.56
Feature first	1.39	0.65	2.21	3.99
Explicit combined	1.37	0.31	0.87	2.55
Implicit combined	1.44	0.36	1.27	2.94

* Mean numbers in bold are statistically different ($p < .05$) from the other means in the same column. There was no main effect of novelty and the interaction between training condition and novelty was nonsignificant.

[†] Hits indicative were features that were present on the ECG and indicative of the correct diagnosis.

[‡] Hits not indicative were features that were present on the ECG, but not indicative of the correct diagnosis.

[§] False alarms were features that were not present on the ECG.

While the reductionist approach and our small sample size could be perceived as limitations, the advantages of such a strategy have been well described.¹² Simple verbal reports of reasoning strategy are insufficient for this purpose because by definition, nonanalytic processes are often unavailable to conscious introspection,¹³ and the requirement to verbalize can alter the reasoning processes themselves,¹⁴ thereby calling into question the validity of the inferences that can be drawn.¹⁵ As such, it is necessary to carefully design experimental manipulations to determine the relative benefit of instruction to be analytic or nonanalytic during diagnostic decision making. Relative to curricular-level interventions, carefully controlled, time-limited trials, such as this one, aimed at specific educational issues, provide the opportunity to identify the active ingredient in learning activities independent of numerous confounding variables that have a tendency to occlude curriculum level effects. While ecological validity is sacrificed to some extent, this study demonstrates that it need not be forfeited entirely as the instructional devices (i.e., the training materials and teaching methods) used are perfectly compatible with those used during actual medical training. In fact, the basic instructional approach we used in this study was modeled closely on one used by clinical teachers in McMaster University's Undergraduate MD program. The invaluable by-product of this design strategy is that the experimental effects tend to be large enough, when present, that studies are often sufficiently powered to reveal

statistically significant differences with relatively small sample sizes. The participants in our study, absolute novices in ECG diagnosis, achieved performance levels on this limited task equivalent to that of second-year medical students (who have been shown the same materials in previous research) regardless of condition; those who received combined instructions revealed diagnostic accuracy equivalent to that of second year residents.¹⁶

An obvious limitation of our study is that it may not apply to all aspects of medical expertise. Extracting information from an ECG is only one small component of competence. Nevertheless, other studies, using participants of all levels of expertise, have revealed that ambiguity is an inherent part of many diagnostic tasks and may confer an advantage to diagnostic-directed search in both visual and verbal materials, at least when done in conjunction with careful consideration of features.^{3,17} Two previously published studies of dermatological diagnosis, one with residents,¹⁸ and one with medical students,¹⁹ are particularly relevant. In both cases, participants were instructed either to use a "feature first" (i.e., analytic) approach, carefully considering the features presented on dermatological slides before assigning a diagnostic label, or to take a "similarity-based" (i.e., nonanalytic) approach by providing diagnostic labels based on one's first impression of the slide. Stimuli were carefully selected such that a comparison could be drawn between slides for which dimensions of similarity (relative to slides seen during training) and typicality (as

defined by an expert dermatologist) were made orthogonal. Differences in diagnostic accuracy between similar/dissimilar and between typical/atypical provided an approximation of the extent to which nonanalytic and analytic processes, respectively, influenced the diagnosis.

Relevant to our study, both of these studies revealed main effects of similarity and typicality (i.e., diagnostic accuracy was better for similar and typical cases relative to dissimilar and atypical cases, respectively), indicating that analytic and nonanalytic processes influenced the diagnostic decisions of both residents and undergraduates. Furthermore, there was no main effect of instruction in either study, suggesting that it is inappropriate and unnecessary to caution students to avoid using pattern recognition. Our study builds on this work by addressing the residual issue of whether there might be benefits to adopting a multifaceted instructional approach when teaching novices.

Further research is required to formally identify the reason for the poor performance resulting from use of either of the isolated reasoning instructions, but a pair of interesting hypotheses is supported by the data we have collected. First, the analysis of feature calls across condition suggests, as do the data of Norman et al.⁹ that learners who are taught to carefully identify features before generating diagnostic hypotheses are able to do so too well. Participants in the feature first condition identified more hits indicative of incorrect diagnoses and more false alarms than did participants in either the similarity-based or the combined conditions. Taken in combination with the finding that participants in all four training conditions were equally likely to identify features consistent with the correct diagnosis, this finding suggests that diagnosticians who try to objectively list features without the guidance of diagnostic hypotheses can be led astray by finding themselves awash in a list of features that cannot be reconciled into a coherent diagnostic entity.

Second, and perhaps more surprising, is the finding that participants in the similarity-based condition were able to overcome their initially incorrect diagnostic decisions upon being asked to

consider the features more systematically. A number of studies have shown tentative diagnoses bias one's consideration of clinical cases,⁸ even when the tentative diagnoses are generated by the diagnostician herself.²⁰ This biasing can influence the identification of features, making people less likely to see or interpret features as indicative of alternative diagnoses, thereby creating the potential for self-fulfilling prophecies and diagnostic error.^{4,21} In prior research, we have shown, however, that one way to overcome the bias created by diagnostic alternatives is to induce a more careful, analytic consideration of the features present in a case.²² Finding that the similarity-based group was able to overcome a decision in favor of incorrect diagnoses after performing a more careful feature analysis provides a unique confirmation of those earlier results while also supporting the hypothesis that one should not rely exclusively on any one form of processing.

Our findings have very clear educational implications. Clinical teachers should not guard against the use of nonanalytic reasoning strategies (short of blind guessing, of course) when counseling medical trainees regarding how to proceed in learning the diagnostic categorization schemes they will need to apply over the course of their careers. On the contrary, four experimental studies now have shown that instruction to use similarity (i.e., pattern recognition) during diagnostic decision making will result in diagnostic accuracy at least as good as using more analytic, feature-based strategies. Coderre et al.⁷ were also able to show that pattern-recognition strategies are beneficial and that the increase in diagnostic accuracy that can be gained from their use may be unrelated to level of expertise. Our study adds to these findings, however, by

illustrating that various reasoning/teaching strategies need not be mutually exclusive and, in contrast, can complement one another, leading to greater diagnostic accuracy when used together than when either an analytic or nonanalytic strategy is used in isolation.

Acknowledgments

The authors thank Mary Lou Schmuck for her assistance with the data analyses contained within this report, as well as a pair of anonymous reviewers for their suggestions on how to improve the manuscript. No external funds were received for this project.

References

- Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ.* 2005;39:98–106.
- Custers EJFM, Regehr G, Norman GR. Mental representations of medical diagnostic knowledge: a review. *Acad Med.* 1996;71(10 suppl):S55–S61.
- Norman GR, Brooks LR. The non-analytic basis of clinical reasoning. *Adv Health Sci Educ.* 1997;2:173–84.
- Hatala R, Norman GR, Brooks LR. Impact of a clinical scenario on accuracy of electrocardiogram interpretation. *J Gen Intern Med.* 1999;14:126–29.
- Barrows HS, Norman GR, Neufeld VR, Feightner JW. The clinical reasoning process of randomly selected physicians in general medical practice. *Clin Invest Med.* 1982;5:49–56.
- Elstein AS, Shulman LS, Sprafka SA. *Medical Problem Solving: An Analysis of Clinical Reasoning.* Cambridge, MA: Harvard University Press, 1978.
- Coderre S, Mandin H, Harasym PH, Fick GH. Diagnostic reasoning strategies and diagnostic success. *Med Educ.* 2003;37:695–703.
- LeBlanc VR, Brooks LR, Norman GR. Believing is seeing: the influence of a diagnostic hypothesis on the interpretation of clinical features. *Acad Med.* 2002;77(10 suppl):S67–S69.
- Norman GR, Brooks LR, Colle CL, Hatala RM. The benefit of diagnostic hypotheses in clinical reasoning: experimental study of an instructional intervention for forward and backward reasoning. *Cogn Instruct.* 2000;17:433–48.
- Eva KW, Norman GR. Heuristics and biases—a biased perspective on clinical reasoning. *Acad Med.* 2005;39:870–72.
- Schamroth L. *An Introduction to Electrocardiography.* 6th ed. Oxford: Blackwell Scientific Publications, 1982.
- Norman GR, Schmidt HG. Effectiveness of problem-based learning curricula: theory, practice, and paper darts. *Med Educ.* 2000;34:721–28.
- Bargh JA, Chartrand TL. The unbearable automaticity of being. *Am Psychol.* 1999;54:462–79.
- Roediger HL. Implicit memory: retention without remembering. *Am Psychol.* 1990;45:1043–56.
- Eva KW, Brooks LR, Norman GR. Forward reasoning as a hallmark of expertise in medicine: logical, psychological, and phenomenological inconsistencies. In: Shohov SP (ed). *Advances in Psychological Research.* Vol 8. New York: Nova Science Publishers, 2002:41–69.
- Hatala R, Norman GR, Brooks LR. Impact of a clinical scenario on accuracy of electrocardiogram interpretation. *J Gen Intern Med.* 1999;14:126–29.
- Brooks LR, Coblenz CL, Norman GR, Babcock CJ. Expertise in visual diagnosis: a review of the literature. *Acad Med.* 1992;67(10 suppl):S78–S83.
- Regehr G, Cline J, Norman GR, Brooks LR. Effects of processing strategy on diagnostic skill in dermatology. *Acad Med.* 1994;69(10 suppl):S34–S36.
- Kulatunga-Moruzi C, Brooks LR, Norman GR. Coordination of analytic and similarity-based processing strategies and expertise in dermatological diagnosis. *Teach Learn Med.* 2001;13:110–16.
- Eva KW, Brooks LR. The under-weighting of implicitly generated diagnoses. *Acad Med.* 2000;75(10 suppl):S81–S83.
- Brooks LR, LeBlanc VR, Norman GR. On the difficulty of noticing obvious features in patient appearance. *Psychol Sci.* 2001;11:112–17.
- Eva KW. The influence of differentially processing evidence on diagnostic decision-making [dissertation]. McMaster University, Hamilton, Ontario, Canada 2001.