PEDAGOGICAL VALUE OF ADOPTING MULTFACETED APPROACHES

THE PEDAGOGICAL VALUE OF ADOPTING MULTFACETED APPROACHES TO DIAGNOSTIC REASONING

By

TAVINDER K ARK, B.Sc

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

MASTER OF SCIENCE (2005)
(Psychology)

McMaster University
Hamilton, Ontario

TITLE: The pedagogical value of adopting multifaceted approaches to diagnostic reasoning

AUTHOR: Tavinder K Ark, B.Sc. (McMaster University)

SUPERVISOR: Professor Kevin W Eva

NUMBER OF PAGES: vii, 78

Abstract

There has been much debate in the medical education literature regarding the role of analytic and non-analytic reasoning strategies, but the impact of teaching one or the other strategy has yet to be directly tested. Analytical processes are those that entail the systematic and conscious consideration of features and their relationship to potential diagnoses – this is the form of clinical reasoning that has traditionally been advocated by educators. Alternatively, non-analytical processes are automatic, often unconscious means by which to arrive at a clinical diagnosis. One example of such a process is reliance on similarity to previously seen cases. The availability of these strategies are believed to increase with expertise, but debate exists regarding whether or not they are useful pedagogically. The purpose of these experiments was to examine the effectiveness of emphasizing particular decision-making processes during learning using an experimental methodology. Undergraduate Psychology students were trained to identify features on electrocardiograms (ECGs) and assign diagnoses. Experiment 1 focused mainly on understanding the role feature identification plays in decision-making, and whether instructions regarding how to organize the diagnostic features are beneficial. The purpose of experiment 2 was to determine the relative benefit of adopting a combination of feature driven and non-analytical strategies and contrastive instructions (i.e. being explicitly told to compare similarities and differences between diagnostic categories) at the time of training.

ACKNOWLEDGEMENTS

None of this could be possible without mentioning a few people. First and foremost, I'd like to thank my supervisor, Kevin Eva. Kevin has taught me many things among which is to speak and write about my research ideas although I think that may still be a work in progress (is there light at the end of that tunnel?). I'd also like to thank him for his support and patience. Next, I'd like to thank the members on my committee: Geoff, Lee and Bruce. Geoff Norman must be thanked for the honesty he has brought to my research, his sharp mind, intriguing sense of humour and the many occasions we talked about photography. I can only hope to be able to snap an amazing shot of a horizon with a "five dollar" camera from Shoppers drug mart as he did. I'd like to thank Lee Brooks for his input and creative thinking. Even within the short span that I got to work with Lee, I felt the presence of a great mind at work. He taught me that an idea is only as good as the story I tell. Special thanks must go to Bruce Milliken who taught me to take a chance with my research ideas and when I was wrong guided me back on the right path.

I'd like to thank my lab mates: Sam, Aimee, Nikki and Meredith. It was always an adventure in the lab because of the fun and energy the dynamic trio – Sam Hannah, Aimee Skye and Nikki Woods – brought into the lab. Each day was truly unique and it was fun to have the opportunity to listen to the banter, which included trying to improve Sam's wardrobe (although I did like the knee and elbow patches) to talking about the latest reality show like Paradise Hotel. If they had capes and superpowers I'm not sure what would have ensued. Meredith Young must also be thanked for bringing in fresh energy.

I'd also like to thank my mom, dad and brother. It is always nice to come home and have hot food on the table and clothes nicely ironed to wear the next day! I'd like to thank my friends with a special thanks to my silly friend Marie who kept my spirits high, listened to my outrageous stories, was the voice of reason (apparently I have none because the number of times she has stopped me from fleeing in terror thinking that was the wisest course of action when dealing with my thesis has been astronomical) and made the road a little less bumpy through my thesis. She also made me appreciate how dangerous dingoes really are, taught me when birds should fly, and if anything, putt up with my absurd need to make up words and crazy sayings. Jee Murhpy, tickle me silly! I'd like thank all my 'coffee making buddies' at Starbucks and Second Cup, especially James. Without you all I'd have been missing the well needed caffeine to get me through most mornings. How can I not thank my computer, Pugsley, who more than once, wanted to jump off the desk and run away. Every so often he even tried to commit suicide by crashing on me, but in the end we made it! Thank god for the both of us. Lastly, I acknowledge this may be the best part of my thesis.

Table of Contents

List of Tables and Figures

Tables:

Figures:

INTRODUCTION

To determine if a patient has suffered an acute anterior myocardial infarction requires the identification of certain diagnostic features, such as chest pain radiating down the left arm 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 whereby this task can be performed - analytic and non-analytic. Both of these models have been operationalized in many ways and should not be considered mutually exclusive.

Traditionally, medical educators endorse analytical processes of clinical reasoning when teaching medical students. These processes entail the careful analysis of features and their relationship to possible diagnoses. Instructions to carefully identify and consider all the features before generating a diagnostic hypothesis are believed to reduce premature closure during a diagnostic search (i.e. failure to consider all the diagnoses possible; Patel, Groen & Arocha, 1990) and biases that can result when just one diagnosis is kept in mind. As a result, it is understandable why medical educators advocate such a strategy.

Alternatively, in the past fifteen years research has empirically demonstrated a non-analytical component of clinical reasoning. Central to this view is the belief that relying on similarity to previously seen cases when assigning a diagnosis is a valid alternative strategy. That is, one might automatically recognize the correct diagnosis by rapid processes like pattern recognition, which typically occur unconsciously. The availability of this strategy is believed to increase with expertise as a direct result of acquiring a vast repertoire of prior cases to which future cases may be compared. Despite the evidence, some medical educators emphasize avoidance of this strategy because they believe it has detrimental effects if used by novices (Coderre, Mandin & Harasym, 2003).

There are potential benefits and drawbacks to using either strategy and therefore, there has been much debate in the medical education literature regarding the relative value of teaching one or the other strategy. Studies by Regehr, Cline, Norman and Brooks (1994) and Kulatunga-Moruzi, Brooks, and Norman (2001) measured the relative contribution of instructions to be analytic and non-analytic during dermatological diagnosis; the former study involved residents, the latter involved medical students.

In both studies, participants were instructed to either use a "similarity-based" approach – participants were told to 'quickly look at the slide, then choose the most likely diagnosis based on whatever came to mind' – to invoke non-analytical processes, or to take a "feature first" approach by carefully identifying the features present before assigning a diagnosis to foster a type of analytical processing. The dermatological slides were orthogonally organized so that comparisons could be made between similarity and typicality (a case is similar if the current case closely resembles previously seen cases, whereas a case is typical if it contains more of the features characteristic of that disorder). The extent to which non-analytic and analytic processes were operational was measured by the difference in diagnostic accuracy between similar/dissimilar and typical/atypical cases. Likewise, the interaction between case type (similar vs. dissimilar, typical vs. atypical) and instructions given (feature first vs. similarity based) provided an approximation of the extent to which instructions altered participants' diagnostic strategy.

For example, if the feature first instructions resulted in a greater difference in diagnostic accuracy between typical and atypical cases than a similarity based instruction, it suggests that a feature first instruction increased the extent to which analytic processes were operational.

Both of these studies found a main effect of similarity and typicality – similar and typical cases were diagnosed more accurately than dissimilar and atypical ones respectively. This suggests that both analytic and non-analytic processes were operational in the diagnostic decision making process for both residents and medical students. Furthermore, Regehr et al. (1996) found residents were only amenable to feature-oriented instructions, whereas Kulatunga-Moruzi et al. (2001) found medical students were only amenable to similarity based-instructions. The discrepancy found between these two studies suggests that residents do not spontaneously use feature-driven processes as extensively as novices. Yet, they appear able to resort to causal rules and consider features carefully when prompted to do so rather than relying primarily on similarity to previously seen cases when making a diagnosis. Likewise, novices can take advantage of similarity when prompted to do so by instructions, as opposed to solely relying on features, although it is not their typical mode of processing. Lastly, there was no main effect of decision-making instructions in either one of these studies suggesting that it is unnecessary to caution novices against using similarity.

EXPERIMENT 1

Participants in Regehr et al. (1996) and Kulatunga-Moruzi et al. (2001) were instructed to use either "feature first" or "similarity-based" models of clinical reasoning in isolation. However, neither of these studies addressed whether adopting a multi-faceted instructional approach might be more beneficial when teaching novices. In teaching undergraduate psychology students to diagnose electrocardiograms (ECGs), Norman, Brooks, Colle & Hatala (2000) manipulated decision-making instructions in order to determine the effect of "forward" and "backward" reasoning strategies upon diagnostic accuracy. A forward strategy is defined as reasoning from the data to a solution, that is, from the features to a diagnosis. Conversely, backward reasoning is defined as working from the solution back to the data, or in medical terms, working from the diagnosis to the features one would expect to find if it was that diagnosis. However, the instructions in their study that were designed to invoke a backward reasoning strategy could also be described as a combination of feature first and first impression instructions. Participants in this condition were told to use 'the list of possible diagnoses, but not to "jump the gun" (i.e., to trust non-analytic processes, but also to perform a careful feature search). Alternatively, the forward condition participants were instructed to strictly identify features first and then assign a diagnosis (i.e., to be highly analytic).

The Norman et al. (2000) study found that participants in the backward condition outperformed those in the forward condition – 62% to 49% when both conditions had an ECG in front of them at the time of assigning a diagnosis. Furthermore, the participants in the forward group identified on average 1.70 irrelevant features per case, while those in the backward condition only identified 0.98. This may be one explanation for why the forward group participants showed lower diagnostic accuracy relative to those in the backward group – the participants in the forward group had to account for irrelevant features in their diagnosis or dismiss them. The lower diagnostic accuracy observed in the forward group suggests that the instructions to use the disease list may increase diagnostic accuracy by reducing the number of irrelevant features identified.

That being said, perhaps the reduced diagnostic accuracy on the part of participants given a "feature first" instruction (forward condition) was not simply a consequence of the over-identification of irrelevant features and the need to account for them in their diagnosis. Participants may have had a difficult time assigning a diagnosis because they were unaware of how to organize the features they identified. That is, perhaps participants can identify as many features as possible without harm, as long as they have a strategy to organize the features to match up correctly with a diagnostic category.

Both of these possibilities were examined in the current study in order to determine if giving participants an orienting framework in addition to their feature first instructions may help to increase their diagnostic accuracy, even when a large number of irrelevant features are labeled. The additional steps were either to have participants rank the features in accordance with 'obviousness and vividness' or list all the possible diagnoses based on the features labeled using the features to argue for and against each diagnosis.

The purpose of this experiment was to determine if providing instructions regarding how to organize features is beneficial to help students overcome instructions to carefully consider features before generating diagnoses. Moreover, do the instructions to organize features as a) independent sources of information or b) with potential diagnoses lead to the same likelihood of formulating the correct diagnosis? It was hypothesized that participants given the additional instructions would diagnose ECGs more accurately than those given just the feature first instructions because they would retain qualitative information and process the features to a deeper level. The results were anticipated to provide insight regarding whether or not simple over identification of features or the lack of organization of the identified features better explains the reduced diagnostic accuracy observed in the forward condition relative to the backward condition used by Norman et al. (2000).

Methods

Participants

Forty-eight undergraduate students enrolled at McMaster University participated in this experiment for experimental credit. None of the participants had previous experience with ECGs.

Design and Procedure

I – Training Phase

Learning took place in groups of one to three with the experimenter. All participants were presented with general information regarding the 12 leads in an ECG reading and were taught to recognize the general structures of a normal ECG waveform (P wave, QRS complex, ST segment and T wave; See Appendix 1). The experimenter taught each of the ten diagnostic categories, starting with a normal condition, by using a "Features of Various Disorders List" (See Appendix 2). This list contained all the key features for each diagnostic category. For example, Left Bundle Branch Block was presented as typically having (a) RSR' (rabbit ears) present in V5 and V6 and (b) widened QRS complex (Schamroth, 1982).

After participants learned the features of a particular diagnostic category, they were presented with four example ECGs. For the first two ECGs, the experimenter identified the key features for a diagnostic category by using the feature list. For the latter two, participants were asked to identify the relevant features. This process was repeated for each diagnosis (See Appendix 3).

II – Practice Phase

At the end of the training phase, participants were given a practice booklet that consisted of ten ECGs previously seen during training. During this phase, participants were randomly assigned to one of three experimental conditions.

To emphasize an analytic strategy, participants assigned to the "feature first" condition were given the following instructions (See Appendix 4):

> For each ECG, work down the Response Sheet List and indicate all the features that can be seen.

Only after this task was completed were participants asked to assign a diagnosis. Particiants in the other two instructional conditions were given these exact same instructions, but with additional direction.

Participants in the "Argue for alternatives" condition were first given the feature first instructions. Only once they identified the features for all ten ECGs were they instructed to (See Appendix 4):

> Now that you have gone through the feature list, list all the possible diagnoses relevant to the features you have identified. Then argue for and against each possible diagnosis. Once you have done this, make a final diagnosis. Use the disease list to help you.

Participants in the "Rank Features" condition were given the following additional instructions in addition to the "feature first" instructions (See Appendix 4):

> Now that you have gone through the feature list, rank the features in accordance to obviousness and vividness. Once you have done this, make a diagnosis.

All participants were allowed to view the ECGs when assigning their diagnosis. They were also told the correct answer after diagnosing each ECG; whenever an incorrect diagnosis was assigned, immediate feedback was given that reiterated their respective instructions.

III – Test Phase

During the test phase, participants were asked to diagnose twenty ECGs, ten of which were novel and ten of which were seen during training but distinct from those presented during the practice phase. Participants were not given any feedback regarding the correct answer during the test phase. Otherwise, the procedure and instructions were identical to that of the practice phase for each participant.

Results

Diagnostic Accuracy

The mean diagnostic accuracy is reported in Table 1 as a function of instructional condition and novelty of the stimuli. A repeated measures analysis of variance was performed on the test phase data with instructional condition included as a between subject grouping factor. Novelty (old/new) and slide (nested within novelty) were included as within subject repeated measures.

There was a significant main effect of instruction, $F(2,810) = 5.45$, $p<0.05$. Post hoc analyses revealed that participants in the argue for alternatives (49%) group significantly outperformed both the rank features (42%) and feature first (38%) groups ($F(1,396) > 4.59$, $p < 0.05$ in both cases). There was no difference between the feature first and rank features conditions, $F(1,396) = 1.34$, $p > 0.05$.

TABLE 1: Mean Diagnostic Accuracy (%) during test as a function of both instruction and novelty of the stimulus

Condition	Old ECGs	New ECGs	Total
Feature first	38.7	36.2	37.5
Argue for alternatives	50.6	47.5	49.1
Rank features	42.5	40.6	41.6
Total	44.0	41.5	42.7

No difference was observed between old ECGs (those that had been seen during training) and new ECGs (those that had never been seen before by participants; F(1,810)

$= 0.74$, $p > 0.05$). Nor was the interaction between instruction and novelty significant, $F(2,810) = 0.02$, $p > 0.05$.

Feature Identification

For each ECG, features were categorized as (a) hit indicative – a feature that was present and indicative of the correct diagnosis, (b) hit not indicative – a feature that was present but indicative of another diagnosis, or (c) 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 as a function of classification and instruction is presented in Table 2. There was a significant main effect of hits indicative, $F(1,810) = 7.94$, $p < 0.05$. Post hoc analyses revealed that participants in the argue for alternatives condition identified more hits indicative, on average, than participants in the feature first and rank features conditions ($F(1,396) > 6.79$, $p < 0.05$ in both cases). There was also a significant main effect of the total number of features identified by instruction, $F(2,810) = 18.8$, $p < 0.05$. Post hoc analyses revealed that participants in the argue for alternatives and rank features conditions identified more features than participants in the feature first condition, $F(1,396) > 11.0$, $p < 0.05$ for both comparisons. A similar pattern of results was found for false alarms ($F(1,396) > 11.96$, $p < 0.05$ in all cases). In contrast, no difference was found between the mean number of hits not indicative identified as a function of instruction, $F(1,396) = 1.5$, $p > 0.05$. Lastly, the main effect of old/new cases and the interaction between condition and new/old cases were also non-significant in all feature analyses, $F(2, 810) < 0.46$, $p > 0.05$ for both comparisons.

TABLE 2: Mean Number of features identified during the test phase as a function of both feature classification and instruction

Condition	Hit Indicative	Hit not Indicative	False Alarm	Total Features
Feature First	1.34	0.50	1.12	2.96
Argue for Alternatives	1.57	0.56	1.38	3.51
Rank Features	1.41	0.56	1.40	3.37
Total	1.44	0.54	1.30	3.28

Additional analyses were conducted on the feature data of the rank features and argue for alternatives condition. A chi square test was conducted to examine the relationship between the feature classifications ranked as most obvious and vivid (that is, feature categories ranked as "1" by participants) by the rank features participants and the percent correct/incorrect diagnoses made (see Table 3). The analysis revealed that participants who ranked a hit indicative as most obvious and vivid would assign a correct diagnosis 63.4% of the time. If a hit not indicative or false alarm was ranked as 1, then only 23.5% or 17.6% of the time they assigned a correct diagnosis. This difference in proportions across the three conditions was significant, $\_^2(2) = 58.02$, $p < 0.05$.

TABLE 3: Mean Diagnostic Accuracy (%) during the test phase as a function of ranking a feature classification as most obvious and vivid (1) and correct/incorrect diagnoses

Rank Feature as 1:	Mean Diagnostic Accuracy (%)	
	Correct	Incorrect
Hit Indicative	63.4	36.6
Hit Not Indicative	23.5	76.5
False Alarm	17.6	82.4

For those in the argue for alternatives condition, a Mantel-Haenszel chi square test was performed on all the feature classifications (hit indicative, hit not indicative and the false alarm), true and untrue statements used to argue for/against the correct/incorrect diagnoses. A feature used to argue against a diagnosis could be categorized as one of the feature classifications or as (a) true statement – a feature that was not present in an ECG but required for it to be that diagnosis (i.e. a participant argued against a diagnosis by stating that a certain feature was missing and as a result the diagnosis cannot be this disorder. If this feature was correctly used to argue against a diagnosis it was classed as a true statement – meaning it is a valid argument against a diagnosis) or (b) untrue statement – a feature that could have been either present or not in an ECG, but was incorrectly used to argue against a diagnosis (i.e. a participant argued against a diagnosis by stating that this feature was not present when in actuality it was, they just failed to identify it, or they could have used a feature wrongly to argue against a diagnosis. In any of these instances, the feature was classed as an untrue statement). The difference in proportion across all features classifications was significant, $\_^{2}(2) = 9.038$, $p < 0.05$.

The mean diagnostic accuracy is reported in Table 4 as a function of the classification of features to argue for/against the correct and incorrect diagnoses. When participants argued for the correct diagnosis, 78.4% of the time they used a hit indicative and very rarely used a hit not indicative (12.1%) or false alarm (1.6%). When arguing for the incorrect diagnosis, participants used false alarms 98.4% of the time, followed by hits not indicative at 87.9% – they rarely used hit indicatives (21.6%). In contrast, participants argued against an incorrect diagnosis using all three features classifications (hit indicative – 83.3%; hit not indicative - 81.3%, false alarm - 65.2%) and true statements (correctly identifying features they were missing but required for it to be that diagnosis; 80%). When participants argued against the correct diagnosis, false alarms were mostly used (34%).

TABLE 4: Mean Diagnostic Accuracy (%) during the test phase as a function of the feature classifications used to argue for/against the correct/incorrect diagnoses

Argument type	Feature Classification	Mean Diagnostic Accuracy (%)	
		Correct	Incorrect
Argue For	Hit Indicative	78.4	21.6
	Hit Not Indicative	12.1	87.9
	False Alarm	1.6	98.4
Argue Against	Hit Indicative	16.7	83.3
	Hit Not Indicative	18.8	81.3
	False Alarm	34.8	65.2
	True Statement	20.0	80.0
	Untrue Statement	13.8	86.2

Discussion

Participants in the argue for alternatives group had a higher diagnostic accuracy than those in the rank or feature first conditions. Participants in the rank condition did not diagnose ECGs any more accurately than those in the feature first condition. That is, having participants argue for and against possible diagnoses was a beneficial strategy, whereas having participants rank features based on obviousness and vividness before assigning a diagnosis was not. There are several possible explanations for this finding.

We anticipated that having participants argue for alternatives would yield higher diagnostic accuracy than the features first condition because participants would be provided with an organizational framework to aid in the grouping of features. However, the same was predicted for the rank condition and this group did not improve relative to the feature first condition. One possibility is that the argue for alternatives instructions involved ranking the features to argue for and against the possible diagnoses. That is, for a participant to argue for a diagnosis, he/she must decide if a feature is really vivid and obvious in order to do so. As a result, participants in the argue for alternatives condition may have undertaken a more elaborate organizational process than those in the rank condition.

Another possibility is that the instruction to explicitly consider diagnoses is required to enable optimal organization of the features. In fact, past research has found that giving prior information, such as a tentative diagnosis, alters diagnostic accuracy by drawing attention to those features that would be missed otherwise (Berbaum, Franken, Dorfman, Barloon, Ell, Lu, Smith & Abu-Yousef, 1986). Arguing for diagnostic alternatives appears to have made participants more likely to identify features that were present and indicative of the correct diagnosis (hit indicative) than the other two conditions. Both the rank and argue for alternatives condition elicited more features that were not present in the ECG (false alarm) than the feature first condition, but only participants in the argue for alternatives condition were able to overcome this increase in incorrect feature calls. Norman et al. (2001) suggested the irrelevant features identified by participants given a feature first instruction may have been what hurt their diagnostic accuracy in the end – they had to either dismiss or account for the irrelevant features.

The results from the argue for alternatives conditions suggest that the explicit instructions to consider diagnoses provides a structure from which this can be done more effectively. This would suggest that the over identification of the irrelevant features does not itself lead to poor diagnostic performance; instead, the inability to organize the features provides the main problem. In contrast, the rank instruction appears to be an ineffective additional step to help organize the features because the instructions may have encouraged participants to focus only on a few select features and their diagnoses rather than considering them all at the time of assigning a diagnosis.

When participants ranked a false alarm as most obvious (ranked as "1"), 82.4% of the time the participants would assign an incorrect diagnosis. Likewise, if they ranked a hit indicative as most obvious, then 36.6% of the time it would lead to the incorrect diagnosis. It is difficult to determine if having the incorrect diagnosis in mind leads to a false alarm being ranked as 1 or if the false alarm being ranked as 1 led to the incorrect diagnosis; co-selection is a real possibility.

On the other hand, the argue for alternative participants seem to have considered all the features. If we examine the features they used, they would usually argue for the correct diagnosis and against the incorrect using hit indicatives. Very rarely did they argue for an incorrect diagnosis with hit indicatives. When they had to argue against the diagnoses, they either used (a) all three feature categories (hit indicative, hit not indicative and false alarm) or (b) would state the features that were missing but required for it to be that diagnosis.

In summary, the rank instructions appear to focus participants attention only on the features they rank as 1, whereas the argue for alternatives instructions requires participants to focus on all the features. Based on these results, it suggests that diagnostic accuracy is not hurt simply by the presence of irrelevant features identified, but also by not considering all the features present. Just ranking all the features will not overcome the distracting impact of false alarms identification; the features must be organized with potential diagnoses in mind to increase the likelihood of concluding for the correct diagnosis.

The result of this study focused primarily on understanding the role feature identification plays in decision-making based on the findings of the Norman et al. (2000) study. However, Norman et al. (2000) also provided the first opportunity to study the advantages of adopting a multi-faceted (a combined reasoning) strategy. This study was not designed to compare instructions in which novices were told to use "similarity-based" and "feature first" instructions in isolation. Ark, Brooks, Eva (submitted) designed a study to do exactly that. That study, combined with experiment 2 provides further insight into the effectiveness of emphasizing particular diagnostic processes as pedagogic strategies.

EXPERIMENT 2

In Ark et al. (submitted; see Appendix 5), students were trained to complete an ECG diagnosis task similar to that of experiment 1. To emphasize an analytical reasoning strategy, one quarter of participants were instructed to carefully identify all features that could be seen before assigning a diagnosis (feature first). To emphasize a non-analytic reasoning strategy, one quarter of participants were told to trust similarity (i.e., that new ECGs often look like ECGs that they have seen before) and to diagnose based on this impression (first impression pass 1). Afterwards, the same participants were asked to work through the ECGs again, but this time to identify features and re-diagnose each ECG keeping their initial diagnosis in mind (first impression pass 2 or sequentially combined). The remaining half of participants received both feature first and first impression instructions simultaneously. Half of them were given both sets of instructions explicitly (explicit combined), while the remaining participants were given feature first instructions and simply informed that some of the test ECGs had been seen during training (implicit combined).

Therefore, we operationalized the combined instructions in three different ways by having participants: (1) either systematically consider the features presented in a case simultaneously or sequentially, (2) explicitly be told to trust their feelings of familiarity in addition to their instructions to carefully consider features, and (3) implicitly utilize similarity by informing them some of the ECG cases were seen during training.

No difference was found in diagnostic accuracy between the groups given the feature first (42%) and first impression instructions (41%), but the groups instructed to use both strategies (explicitly or implicitly or sequentially) performed significantly better (56%, 53% and 50%, respectively). The results of this study suggest an additive model of clinical reasoning in which instructions to be feature oriented and to trust similarity improves performance in novice diagnosticians. The advantage of this combined approach was observed whether instructions to trust familiarity were given implicitly or explicitly, and regardless of whether the instruction to systematically consider the features presented in a case was given simultaneously or sequentially with instruction to use familiarity.

ECGs were either old ECGs (those that have been seen before) or new ECGs (those that participants did not see before). We anticipated that the diagnoses of participants in the explicit combined and first impression conditions would be more influenced by the old versus new distinction because they were specifically told that new ECGs often look like old ones and to use that sense of familiarity when making a diagnosis. However, no novelty by condition interaction was observed. The findings were consistent with past research that previously seen cases are diagnosed more accurately than new ones (Allen et al., 1992; Brooks, et al., 1991) and stresses the role that similarity plays in guiding diagnostic decisions regardless of the specific instructions provided (Bargh & Chartrand, 1999; Roediger, 1990).

There are still many questions left unanswered with regards to the coordination of analytic and non-analytic instructions as a pedagogic strategy. For example, did the feature first or first impression instructions result in poor diagnostic accuracy relative to a combined one because the instruction hampered performance in the former groups or because it improved performance in the latter groups? Perhaps participants naturally utilize a multi-faceted approach, and instead of improving their performance by suggesting a combined approach, we hindered participants by forcing them into utilizing either process in isolation. This possibility was examined in Experiment 2. Participants were either given a combined instruction or no instruction at all. In addition, we wished to determine if the benefits of instructing participants to utilize a combined approach could also be observed after a time delay rather than just immediately after training.

Furthermore, we aimed to compare the efficacy of a combined strategy to that of contrastive learning models that have been shown to enable transfer. Many medical schools and textbooks alike teach medical students the features that are associated with a particular diagnosis in isolation (Stuyt, de Vries Robbe & von der Meer, 2003). Studies have found that simply stating the principles and giving examples is not an effective way to enable transfer of the principles to new, yet similar cases (Gentner, 1997, 2003). In contrast, analogical encoding does appear to yield some benefit (Gentner, 2003; Gentner & Markham, 1997; Thompson, Gentner & Loewenstein, 2000; Loewenstein, Thompson & Gentner, 1999).

Our ability to access and transfer knowledge from memory depends crucially on how it was learned. If people can extract principles during learning, then these principles can become the starting point for understanding similarities with new cases involving the same principles. One way to promote such an abstraction is by drawing parallels between two or more instances. That is, the comparison of two examples highlights the similar aspects and promotes the abstraction of common relational structures, increasing the likelihood that the principle will be retrieved and applied to future novel instances. This process described above is called analogical encoding. Moreover, McKenzie (1998) suggested that by comparing and contrasting examples of different categories, students learn to extract the critical features that discriminate between categories. Such contrastive learning will make the features that are similar and different between the two diseases more salient, but will especially make the diagnosticity of each symptom more obvious.

In the contrastive learning condition of this experiment, diagnostic categories that (a) share symptoms/features in common (b) are highly similar, and (c) easily confused with one another, were compared and contrasted at the time of learning. For example, Left Ventricular Hypertrophy with Strain and Ischemia were compared and contrasted to one another because they both share ST depression and T wave inversions in the same leads on an ECG. As a result, participants in earlier studies often diagnosed an ECG with ischemia when they have chosen left ventricular hypertrophy with strain and vice versa.

In summary, there are four manipulations in this experiment. One is the contrast of a combined instruction relative to non-specific instructions. The second is the comparison of a contrastive and noncontrastive mode of learning. The third is time – participants in all groups were tested immediately after learning and then after a week-long delay to determine if benefits of instructing participants to use a combined approach could be observed after a time delay in addition to just immediately after training. The fourth is novelty of the ECGs – a comparison of old and new ECGs. Main effects of contrastive learning and combined instructions were predicted. The results of this study were anticipated to provide insight regarding whether or not the diagnostic accuracy of new ECGs could be improved and whether or not participants naturally utilize a combined approach and therefore do not need explicit instructions to do so. If they naturally utilize a combined approach, then explicit instructions to do so might be expected to have minimal effect.

Methods

Participants

Forty-eight undergraduate Psychology students enrolled at McMaster University participated in this experiment for course credit. None of the participants had previous experience with ECGs.

Design and Procedure

I – Training Phase

Learning took place during one-on-one training sessions with the experimenter. All participants were presented with general information regarding the 12 leads in an ECG and taught to read a normal ECG waveform from an illustration – this part of the experiment mirrored experiment one because the materials were identical. The only difference in materials is that participants learned eight categories in this experiment, whereas they learned ten in experiment one. After this general introduction to ECGs, participants were randomly assigned to one of two learning conditions.

Combining Needham & Begg's (1991), Gentner's (1997, 2003) and McKenzie's (1998) research, participants in this experiment learned the diagnostic categories by either contrastive or noncontrastive learning. Participants in the noncontrastive condition were taught each of the eight diagnostic categories beginning with normal using a "Features of Various Disorders List" – the same list that was used in experiment 1. Like experiment 1, for each diagnostic category participants were given four examples in sequence. The experimenter identified the key features that belonged to that particular disorder for the first two ECGs. Then participants identified the same relevant features for the last two ECGs. This process was repeated until all diagnostic categories and their examples were presented (See Appendix 4).

Participants assigned to the contrastive learning condition were asked to compare and contrast the features that were similar and dissimilar between a diagnostic category and normal ECG before comparing it to a confusable differential. Participants were encouraged to self generate the similarities and differences between categories as much as possible. However, in instances where they failed, the experimenter would provide hints at the differences. This happened occasionally. For example, participants were asked to compare and contrast the features that were similar and dissimilar for two example ECGs of Left Ventricular Hypertrophy with strain to a normal ECG. They then compared and contrasted two examples of Ischemia (this disorder is often confused with Left Ventricular Hypertrophy with strain) to normal. Finally, they compared the two different examples of Left Ventricular Hypertrophy with strain to two different example ECGs of Ischemia. This process was repeated until all diagnostic categories were compared to normal and their differential. On average, training time for participants in the contrastive condition lasted forty to forty-five minutes, while those in the noncontrastive lasted only thirty to thirty-five minutes.

II – Immediate Test Phase

After the training phase, participants were presented with a practice booklet consisting of seven previously seen ECGs from training. During this phase, participants were randomly divided further into one of two decision-making instructional conditions. To emphasize a combined (analytic and non-analytic) strategy, participants were given the following instructions (See Appendix 4):

> 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 told to use similarity while simultaneously carefully considering the features present. Participants in the "no instruction" condition were verbally told to (See Appendix 4):

> Assign a diagnosis to each ECG using whatever strategy comes naturally to you.

That is, no instructions were given to these participants with regards to the strategy they should employ when assigning a diagnosis. Instead they were told to approach the cases in whatever manner came naturally to them.

All participants were allowed to view the ECGs when assigning their diagnosis and were told the correct answer after diagnosing each ECG. Whenever they assigned an incorrect diagnosis, immediate feedback was given that reiterated both their learning and decision-making instructions.

III – Delay Test Phase

The delay test phase occurred a week after training. During this phase, participants were asked to diagnose fifteen ECGs, eight of which were novel and seven of which had been seen during training but were distinct from those seen during the Immediate Test phase. Participants were not given any feedback with respect to the correct answer; otherwise the procedure and decision-making instructions were identical to those given during the immediate test phase for each participant.

Results

Diagnostic Accuracy

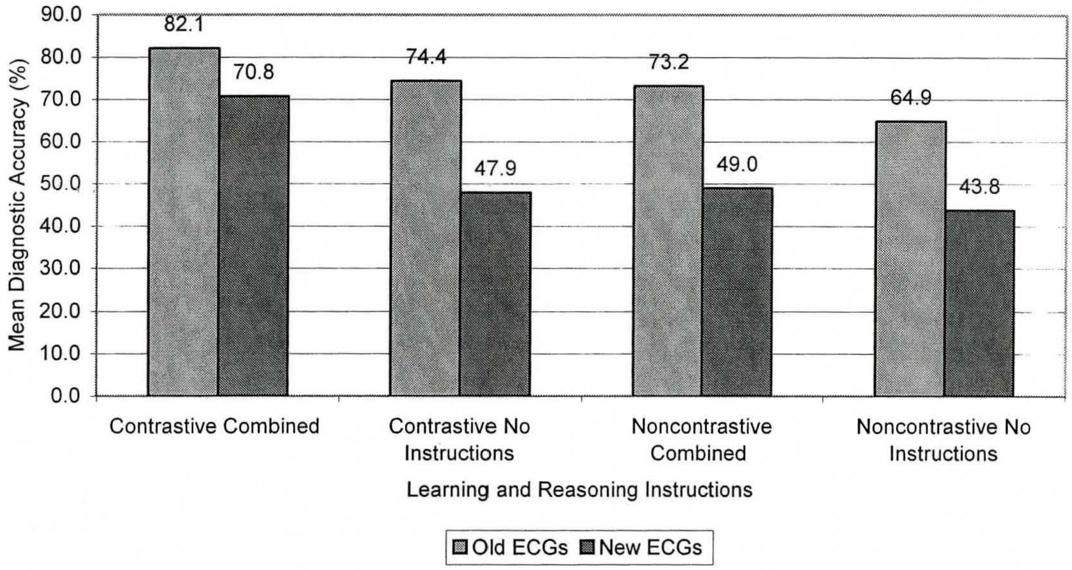
A repeated measures analysis of variance was performed with learning instructions and decision-making instructional conditions as the between subject grouping factors. Time (immediate/delay), novelty (old/new) and slide (nested within novelty) were included as within subject repeated measures. No significant main effect of time was observed between the immediate or delay phase, $F(1,572) = 0.781$, $p > 0.05$); nor were the interactions between time and learning/decision-making instructions significant ($F(1,572) < 0.121$, $p > 0.05$ in both comparisons). Therefore, the data were collapsed across this factor in further analyses. The means are illustrated in Table 5. Participants in the contrastive learning condition performed significantly better (71%) than those in the non-contrastive condition (61%), $F(1,572) = 19.2$, $p < 0.05$. Those in the combined instructional condition (71%) significantly outperformed those in the no instruction group (61%), $F(1.572) = 17.9$, $p < 0.05$.

TABLE 5: Mean diagnostic accuracy (%) as a function of learning and decision-making instructions

	Contrastive Learning	Noncontrastive Learning	Total
Combined Instructions	78.0	64.4	71.2
No Instructions	64.8	57.2	61.0
Total	71.4	60.8	66.1

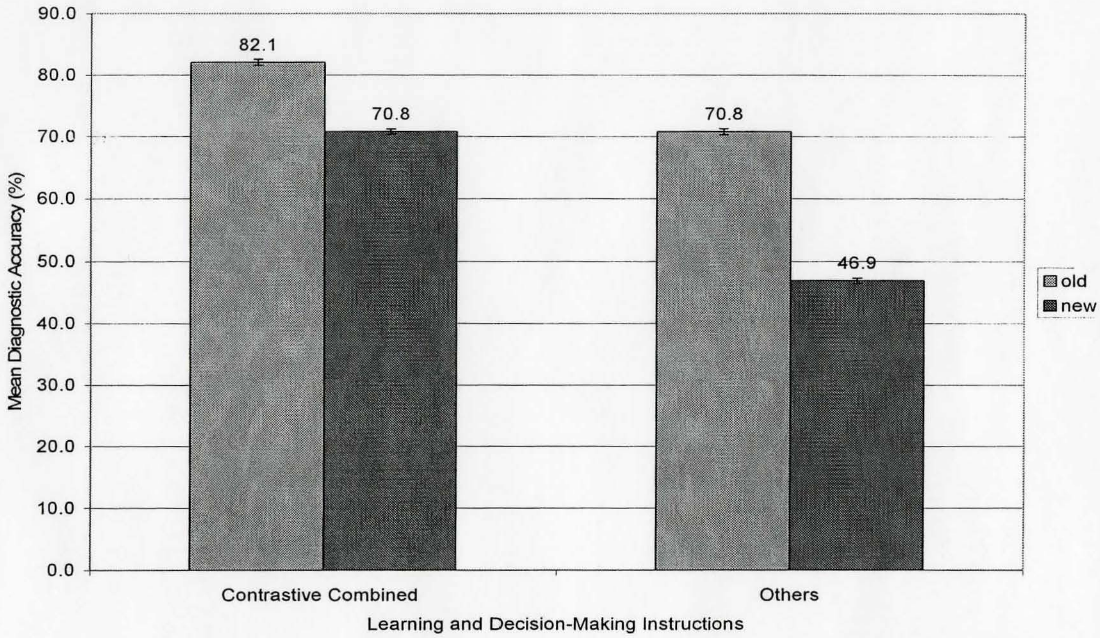
On average, old ECGs (74%) seen during training were diagnosed more accurately than new ones (53%; $F(1,572) = 10.7$, $p < 0.05$, the means are illustrated in Figure 1). All interactions between novelty, learning and decision-making instructions were non-significant, $F(1,572) < 1.56$, $p > 0.05$ in all comparisons.

FIGURE 1: Mean diagnostic accuracy (%) as a function of learning and decision-making instructional conditions together and novelty of the stimulus



A post hoc analysis was conducted on participants in the combined contrastive condition versus the other three learning and decision-making instructional conditions (contrastive no instructions, noncontrastive combined and noncontrastive no instructions) collapsed together and an analogous ANOVA to the one performed on the diagnostic accuracy in this experiment was conducted. Participants in the contrastive combined (78%) condition diagnosed ECGs more accurately than participants in any of the other three (62%) learning/decision-making conditions, $F(1,920) = 32.5$, $p < 0.05$). Old ECGs were diagnosed more accurately than new ones, $F(1,920) = 68.6$, $p < 0.05$, and the were diagnosed more accurately than new ones, $F(1,920) = 68.6$, $p < 0.05$, and the interaction between novelty, learning and decision-making conditions was significant, $F(1,920) = 4.75$, $p < 0.05$ The contrastive combined (71%) participants diagnosed new ECGs more accurately than those in the other three (47%) conditions.

FIGURE 2: Mean diagnostic accuracy (%) as a function of Contrastive Combined versus the other learning and decision-making instructional conditions together and novelty of the stimulus



An additional analysis was performed on the combinations of learning and decision-making instructional conditions. A Mantel-Haenszel chi square was conducted to examine the relationship between the number of times participants diagnosed an ECG incorrectly because they confused it with its differential (see Table 6). The differences in proportion across the four conditions was insignificant, but in the right direction ($\chi^2(3) = 3.22$, $p > 0.05$). The analysis suggests participants in the contrastive combined condition (41%) were less likely to confuse an incorrect diagnosis with its differential. Those in the contrastive no instructions, noncontrastive combined and noncontrastive no instructions confused an incorrect diagnosis with the differential 64%, 56% and 59% of the time.

TABLE 6: Mean percentage of times participants diagnosed an ECG incorrectly because they choose the differential diagnosis as a function of learning and decision-making instructional conditions together

Conditions	Mean Percentage (%) of times participants incorrectly concluded in favour of the confusable diagnostic alternative
Contrastive Combined	40.9
Contrastive No Instructions	64.4
Noncontrastive Combined	56.0
Noncontrastive No Instructions	58.5

Feature Identification

The feature data were not analyzed by decision-making instructional conditions because the instructions between the two groups varied in terms of the emphasis placed on the identification of features. That is, the instructions given to participants in the no instructions condition did not require the identification of features, whereas participants in the combined condition were encouraged to label the features observed. As a result, obvious differences would arise between these two conditions.

On the other hand, the feature data from the two learning conditions could be analyzed. Therefore, the features were classified into the same categories (hits indicative, hits not indicative, false alarms and total features) and an ANOVA analogous to the one performed on the number of features identified within each classification in experiment 1 was conducted. No significant difference was found between the learning conditions (contrastive and noncontrastive) in any of the features classification, $F(1, 572) > 0.22$, $p > 0.05$. See Table 7 for the mean number of features as a function of classification and the learning instructions.

TABLE 7: Mean number of features identified as a function of both feature classification and learning instructions

Learning Instruction	Hits Indicative	Hit Not Indicative	False Alarm	Total Features
Contrastive	0.877	0.136	0.326	1.33
Noncontrastive	0.762	0.129	0.409	1.30
Total	0.820	0.133	0.368	1.32

Discussion

Instructions to be both feature-oriented and to use similarity-based reasoning strategies (combined condition) led to significantly better performance than instructions that gave no guidance regarding the strategy that should be employed. The results suggest that participants in the "no instructions" condition either (a) do not naturally utilize a combined strategy, or (b) utilize this strategy but not as effectively as when given explicit instructions to do so. Research by Paivio (1969, 1975, 1991), Reynolds & Paivio (1968) and Sadoski, Kealy, Goetz and Paivio (1997) found that the use of concreteness in composing definitions led to improved performance in language tasks relative to abstract or no definitions. According to these studies, phrases or sentences – like our combined instructions – are more imaginable, comprehensible and memorable when made concrete. Similarly, while the no instruction participants may have employed some form of a combined strategy, it may have been ineffective because they did not have an explicit and concrete "instruction" to which to refer.

Participants who compared and contrasted diagnostic categories (contrastive condition) had a higher diagnostic accuracy than those who were simply told the features that are produced by a specific disorder (noncontrastive condition). This finding is also consistent with a past study that found medical students who were taught the disorders and their differential diagnosis when interpreting the various abnormal ECG patterns, diagnosed ECGs more accurately than those taught the features produced by a particular disease (Kingston, 1979). There are a few possible explanations for this finding. Participants in the contrastive condition may have diagnosed ECGs more accurately because they learned the diagnosticity of each symptom and therefore could extract the critical features that discriminate among competing diagnoses. This interpretation is supported by the finding that contrastive learners, at least when given combined instructions, were less likely to conclude in favor of the incorrect confusable alternatives. Alternatively, the noncontrastive participants may have been unable to filter out the inappropriate or unnecessary information because a previous study has found that contrastive learners are less influenced by nondiagnostic features than noncontrastive learners (Goldstone, 1996).

There were a few unexpected findings in this study. One was the lack of main effect of time and a lack of interaction between learning direction and decision-making instructions. In general, old ECGs were diagnosed more accurately than new ones, which is consistent with previous findings (Experiment 1, Ark et al., Submitted, Brooks et al., 1991, Allen et al., 1992) and emphasizes that humans have a spontaneous tendency to allow similarity to guide their diagnostic decisions regardless of the specific learning guidance or instructions provided at the time of diagnosis.

Participants did show signs of applying their learning to the solution of new problems – new ECGs were diagnosed accurately 40% to 71% of the time. In particular, the contrastive combined participants not only had the highest diagnostic accuracy overall, but specifically diagnosed new ECGs (71%) more accurately than the other three learning and decision-making conditions (contrastive no instructions, 48%; noncontrastive combined, 49%; noncontrastive no instructions, 44%). That is, although all the learning and decision-making groups exhibited signs of transfer, those in the contrastive combined condition were able to transfer concepts learned from prior examples to solve new problems more readily than the other groups. This demonstrates that combined instructions and contrastive learning together promotes the use of examples in solution of transfer of information to novel cases to a greater extent than contrastive learning on its own.

Why is the combined instruction so important in order to prevent confusion between diagnostic categories that are similar and to increase the transfer of information to novel cases? Why is it that the contrastive no instructions participants did not show similar benefits? Contrastive learning makes participants consider the differential diagnosis more often and the diagnosticity of the features becomes more salient than if they learn in a noncontrastive manner (McKenzie, 1998). That is, by teaching participants the differential diagnosis, that diagnosis becomes more salient than less relevant categories. Those in the contrastive combined condition incorrectly diagnosed an ECG, 41% of the time because they confused it with its differential. Those in the contrastive no instruction condition revealed such errors occur 64% of the time. It appears the decision-making instruction may be an important step in applying the information learned during contrastive teaching instructions and overcoming the saliency of incorrect confusable categories.

Perhaps, the reason why participants given a combined instruction in addition to the contrastive learning were not hindered by the saliency of features or the incorrect confusable alternative is because the combined instruction encourages the development of a framework/strategy around which key features and diagnostic hypotheses can be organized and interpreted (Kushniruk et al., 1998; Brewer & Nakamura, 1984). Conversely, perhaps participants given no instructions on top of contrastive learning could not filter out what is inappropriate or unnecessary information at the time of assigning a diagnosis because they did not develop a framework/strategy to utilize during the decision-making process. Further research needs to be conducted to truly understand the relationship between contrastive learning and combined instructions and to determine why the two instructions together reduced diagnostic confusion between similar categories and increased transfer of previously learned information to the solution of new problems.

In conclusion, it appears both contrastive learning and combined instructions lead to a significantly higher diagnostic accuracy than noncontrastive learning and no instructions. Participants might naturally utilize a combined strategy. However, the explicit instructions to use such a strategy appear to be needed in order for the benefits to be observed. Although Gentner suggests that analogical encoding and thus transfer only succeeds when comparisons are made on a structurally deeper level (i.e. pathophysiology of particular cardiovascular disease; Gentner & Toupin, 1986), our study suggests that the transfer of information can also occur based on increased appreciation of the resemblance of surface elements such as features (i.e. chest pain radiating down the left arm or a rabbit ear in leads V1 and V2). Participants given contrastive learning together with combined instructions were able to a) transfer their previous knowledge from prior example ECGs to solve new cases, and b) reduce confusion between similar diagnostic categories, to a greater extent than using either strategy in isolation.

GENERAL DISCUSSION

Both of these studies examined the effectiveness of emphasizing particular diagnostic decision making processes as pedagogic strategies. The purpose of experiment one was to understand the role feature identification plays in decision-making and determine if instructions to organize the diagnostic features are beneficial especially when more irrelevant features are identified. In experiment one the explicit instructions to consider and argue for diagnostic alternatives appears to provide the organization needed to overcome irrelevant feature calls. The instructions to rank the features in accordance to 'obviousness and vividness' led to as many irrelevant feature as those in the argue for alternatives group, but in this case diagnostic accuracy was hurt. This suggests that the overidentification of features may not have lead to the poor diagnostic accuracy observed in the Norman et al. (2000) study; instead it is a problem with organization and interpretation of the features. That is, if participants organize features by treating them as independent sources of information, then they are unlikely to formulate the correct diagnosis. However, participants that organized features using diagnoses led to improved performance in diagnostic accuracy.

Moreover, arguing for diagnostic alternatives also made participants more likely to identify features of the correct diagnosis. Berbaum et al. (1986) showed that a tentative diagnosis can alter diagnostic accuracy by drawing attention to those features that would otherwise be missed. In fact, the consideration of a diagnosis can activate the representation of that disease presentation and therefore bring to mind the features – ones that were already identified and those missing but required – for it to be this diagnosis.

The current study further confirms that the identification of features can be strongly influenced by category-level information such as a diagnosis. That is, the diagnosis can serve to focus the attention of participants to determine which features to look for and where to find them.

Many have expressed concerns that considering a diagnostic hypothesis may induce biases in the report of features by causing them to be reinterpreted in the light of the diagnostic hypothesis (Hatala, Norman & Brooks, 1996, 1999; Norman, Brooks, Coblentz & Babcook, 1992; Norman, Brooks, Regehr, Marriott & Shali, 1996; LeBlanc, Dore, Norman & Brooks, 2004). This in turn might lead participants to report features that are not present or to miss contradictory features that are present. Although participants in the argue for alternatives condition did identify a greater number of irrelevant features, they were not hindered by it as their diagnostic accuracy shows. Also, these participants were led to consider more than one diagnosis and their features; this may have helped prevent biases that would otherwise be present if they just considered one diagnostic hypothesis.

In experiment two, the application of prior information to the solution of new problems was greatly increased when participants were given contrastive learning directions (i.e. they were asked to compare and contrast diagnostic categories) and combined instructions (i.e. instructed to use both similarity and features) together. That is, the contrastive learning was effective for the transfer of information when participants were given a combined instruction at the time of assigning a diagnosis; otherwise, the increased rates of transfer did not occur as seen in the contrastive no instruction condition.

Generally, there are three hypothesized dimensions to a problem in which transfer can occur: Context, content and deep structure (Gentner & Toupin, 1986). The context includes the physical context such as the room in which a test is taken. The content consists of two broad areas – the semantic domain (i.e. the etiology of a disease) and surface elements (i.e. features of a disease like rabbit ears). Finally the deep structure involves the underlying principle (i.e. the actually disease process). Although true analogical transfer is considered to occur between those things that share a similar deeper level, similarity between problems can exist at any of these levels. Gentner and colleagues suggest that analogical transfer can only succeed on a deeper structural level (Gentner & Toupin, 1986), but we were able to show that transfer can also occur by making comparisons at the level of surface elements (i.e. surface resemblance of features). Additional studies need to be conducted to understanding why contrastive learning and combined decision-making instructions not only leads to an overall higher diagnostic accuracy, but a greater extent of transfer to new ECGs.

In sum, the way novices are taught the features of diagnostic categories and the decision-making strategy they are instructed to utilize at the time of assigning a diagnosis are both important processes in the transfer of knowledge from previously encountered examples to the solution of new problems. Specifically, contrastive learning and combined instructions led to better learning; both of these strategies together are more likely to retrieve the correct knowledge from memory than if either strategy is used in isolation. Therefore, teaching novices the similarities and difference between diagnostic categories (especially the differential diagnosis) and providing them with explicit instructions to utilize a combined strategy appears to provide the most effective pedagogic strategy and should aid in the education of novices and facilitate their progress to a more expert level of problem solving.

REFERENCES

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.K., Brooks, L.R., & Eva, K.W. (Submitted). The best of both worlds: Clinical teachers need not guard against teaching pattern recognition to novices. Journal of American Medical Association.

Bargh, J.A., & Chartrand, T.L. (1999). The unbearable automaticity of being. American Journal of Psychology, 54, 462-479.

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

Brewer, W., & Nakamura, G. (1984). The nature and functions of schemas, in: R.Wyer, T. Srull (Eds.), Handbook of Social Cognition, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 119-160.

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.

Custers EJFM., Regehr, G., & Norman, G.R. (1996). Mental Representations of medical diagnostic knowledge: A review. Academic Medicine, 71, S55-61.

Eva, K.W. (2005). What every teacher needs to know about clinical reasoning. Medical Education, 39, 98-106.

Eva, K.W. The influence of differentially processing information on diagnostic decision-making [dissertation]. McMaster University, Hamilton, Ontario, Canadai, 2001.

Eva, K.W., & Brooks, L.R. (2000). The Under-weighting of Implicity Generated Diagnoses. Academic Medicine, 75, S81-83.

Eva, E.W., Neville, A.J., & Norman, G.R. (1998). Exploring the etiology of content specificity: Factors influencing analogic transfer and problem solving. Academic Medicine, 73, S1-S5.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. Journal of Educational Psychology, 95(2), 393-408.

Gentner, D. (2003). Why we're so smart. In D. Gentner and S. Goldwin-Meadow (Eds.). Language in mind: Advances in the study of language and thought. Cambridge, MA: MIT Press, 195-235.

Gentner, D., Loewenstein, J., Thompson, L. (2003). Learning and Transfer: A general role for analogical encoding. Journal of Educational Psychology, 95, 393-403.

Gentner, D., & Markman, A.B. (1997). Structure mapping in analogy and similarity. American Psychologist, 52, 45-56.

Goldstone, R.L. (1996). Isolated and interrelated concepts. Memory & Cognition, 24, 608-628.

Hatala, R., Norman, G.R., & Brooks, L.R. (1996). The effect of clinical history on physicians' ECG interpretation skills. Academic Medicine, 71(10), S68-70.

Hatala, R., Norman, G.R., & Brooks, L.R. (1999). Impact of a clinical scenario on accuracy of electrocardiogram interpretation. Journal of General Internal Medicine, 14, 126-129.

Kulatunga-Moruzi, C., Brooks, L.R., & Norman, G.R. (2001). Coordination of Analytic and Similarity-Based Processing Strategies and Expertise in Dermatological Diagnosis. Teaching and Learning in Medicine, 13, 110-116.

Kingston, M.E. (1979). Electrocardiograph course. Journal of Medical Education, 54, 107-110.

Kushniruk, A.W., Patel, V.L., & Marley, A.A.J. (1998). Small worlds and medical expertise: implications for medical cognition and knowledge engineering. International Journal of Medical Informatics, 49, 255-271.

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 (Suppl.), S67-S69.

LeBlanc, V.R., Dore, K., Norman, G.R., & Brooks, L.R. (2004). Limiting the playing field: does restricting the number of possible diagnoses reduce errors due to diagnosis specific feature identification? Medical Education, 38, 17-24.

Loewenstein, J., Thompson, L.L., Gentner, D. (1999). Analogical encoding facilitates knowledge transfer in negotiation. Psychonomic Bulletin and Review, 6(4), 586-597.

McKenzie, C.R.M. (1998). Taking into account the strength of an alternative hypothesis. Journal of Experimental Psychology: Learning, Memory, and Cognition, 24, 771-792.

Needham, D.R. & Begg, I.M. (1991). Problem-orientated training promotes spontaneous analogical transfer: Memory-orientated training promotes memory for training. Memory & Cognition, 19, 543-557.

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.

Norman, G.R., Brooks, L.R., Regehr, G., Marriott, M., & Shali, V. (1996). Impact of feature interpretation on medical students diagnostic performance. Academic Medicine, 71(1), S108-109.

Norman, G.R., Brooks, L.R., Coblentz, C.L., & Babcook, C.J. (1992). The correlation of feature identification and category judgments in diagnostic radiology. Memory and Cognition, 20, 344-355.

Paivio, A. (1965). Abstractness, imagery, and meaningfulness in paired-associated learning. Journal of Verbal Learning and Verbal Behavior, 4, 32-38.

Paivio, A. (1975). Imagery and synchronic thinking. Canadian Psychological Review, 16, 147-163.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. Canadian Journal of Psychology, 45, 255-287.

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 as a function of task difficulty. Memory and Cognition, 18, 394-406.

Regehr, G., Cline, J., Norman, G.J., & Brooks, L.R. (1994). Effects of Processing Strategy on Diagnostic Skill in Dermatology. Academic Medicine, 69 (Suppl.), S34-S36.

Reynolds, A., & Paivio, A. (1968). Cognitive and emotional determinants of speech. Canadian Journal of Psychology, 22, 164-175.

Roediger, H.L. (1990). Implicit memory: Retention without remembering. American Journal of Psychology, 45, 1043-1056.

Sadoski, M., Kealy, W.A., Goetz, E.T., & Paivio, A. (1997). Concreteness and Imagery Effects in the written composition of definitions. Journal of Educational Psychology, 89, 518-526.

Schamroth, L. An introduction to Electrocardiography (6th Edition). Oxford: Blackwell Scientific Publications, 1982.

Stuyt, P.M.J., de Vries Robbe, P.F., & van der Meer, J.W.M. (2003). Why don't medical textbooks teach? The lack of logic in the differential diagnosis. The Netherlands Journal of Medicine, 61, 383-387.

Thompson, L., Gentner, D., & Loewenstein, J. (2000). Avoid missed opportunities in managerial Life: Analogical training more powerful than individual case training. Organizational Behavior and Human Decision Processes, 82, 60-75.

Appendix 1 ECG Training Manual

All participants were presented with the "ECG Training Manual," which consisted of two pages of basic electrocardiogram information. The first page listed the names, the axis of the heart they collected electrical impulses from and the location of the where they are placed on the chest and body via an illustration. The second page was an illustration of a normal ECG waveform. This waveform had the major components of an ECG waveform labeled, which participants would need to understand in order to do the experiment.

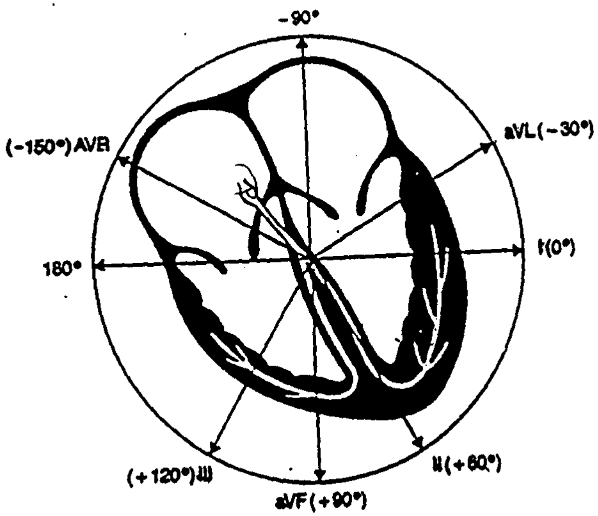
CHAPTER 1:

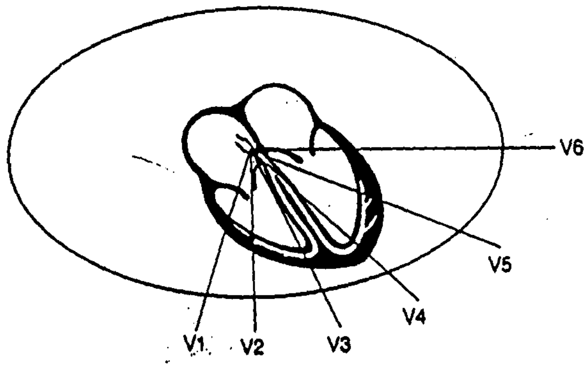
THE ECG TRAINING MANUAL

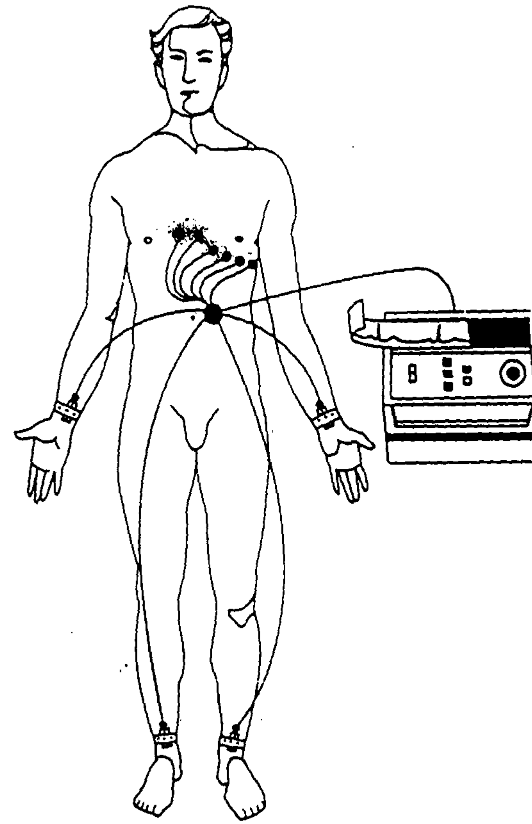
THE BASICS

THE 12 LEADS OF THE ECG

The ECG (electrocardiogram) is a clinical test that records the electrical activity of the heart in a graphical format.





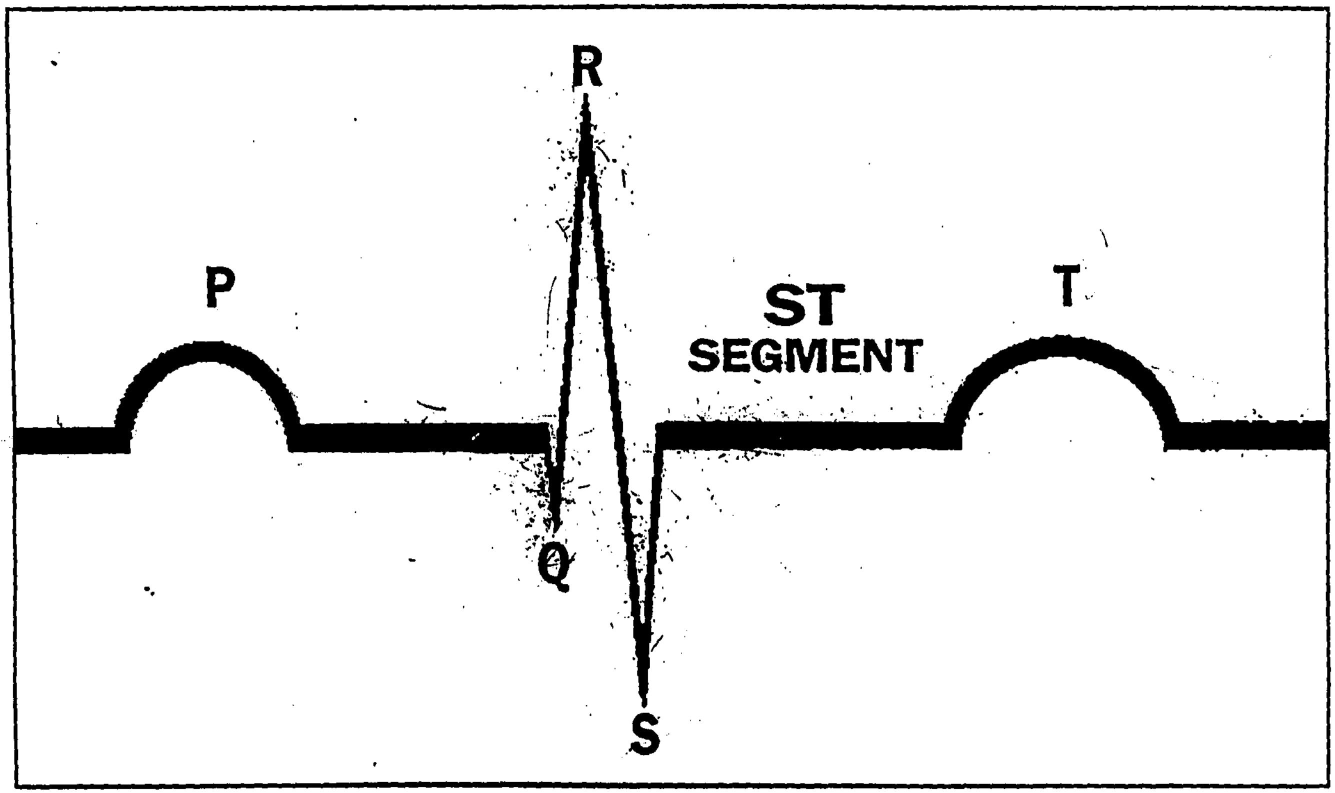


LEADS	LOCATION
I, AVL V5 V6	left Lateral
II, III, AVF	Inferior
V1, V2, V3 V4,	Anterior

AVR

THE ECG WAVE FORM

The electrical impulse seen in any lead is recorded on the ECG graph paper as a basic wave form.



Appendix 2 The Features of Various Disorders

The "Features of Various Disorders" list outlined the key features of each diagnostic category (including normal) that participants learned. This list was available through the duration of the entire experiment. Note: For experiment 1, all 9 categories including normal was used. For experiment 2, Right Bundle Branch Block and Hyperkalemia were excluded.

Features of Various Disorders

Normal

ÿ Predominant S wave in V1 and V2
ÿ Predominant R waves in V5 and V6
ÿ T wave is positive in leads with an R wave, but negative in AVR

Right Ventricular Hypertrophy

ÿ Predominant R wave in V1 (R > S wave or only R wave)
ÿ R wave in V1 is greater than 7 mm
ÿ R wave may get progressively smaller from V2 to V4

Left Ventricular Hypertrophy with Strain

ÿ Large S waves in V1 and V2
ÿ Large R waves in V5 and V6
ÿ ST depression in V4 to V6 and/or
ÿ T waves inverted in V4 to V6

Left Bundle Branch Block

ÿ Widening of QRS complex (> 3 small blocks wide)
ÿ RSR' (rabbit ears) in V5 and V6

Right Bundle Branch Block

ÿ Widening of QRS complex (> 3 small blocks wide)
ÿ RSR' (rabbit ears) in V1 and V2

Acute Anterior Myocardial Infarction

ÿ ST elevation in V1 to V4
ÿ Q waves present in V1 to V4 (not necessary for diagnosis)

Acute Inferior Myocardial Infarction

ÿ ST elevation in II, III, AVF
ÿ Q waves present in II, III, AVF (not necessary for diagnosis)

Ischemia

- ST depression and/or T wave inversion in any of the leads

Pericarditis
- ST elevation across many leads

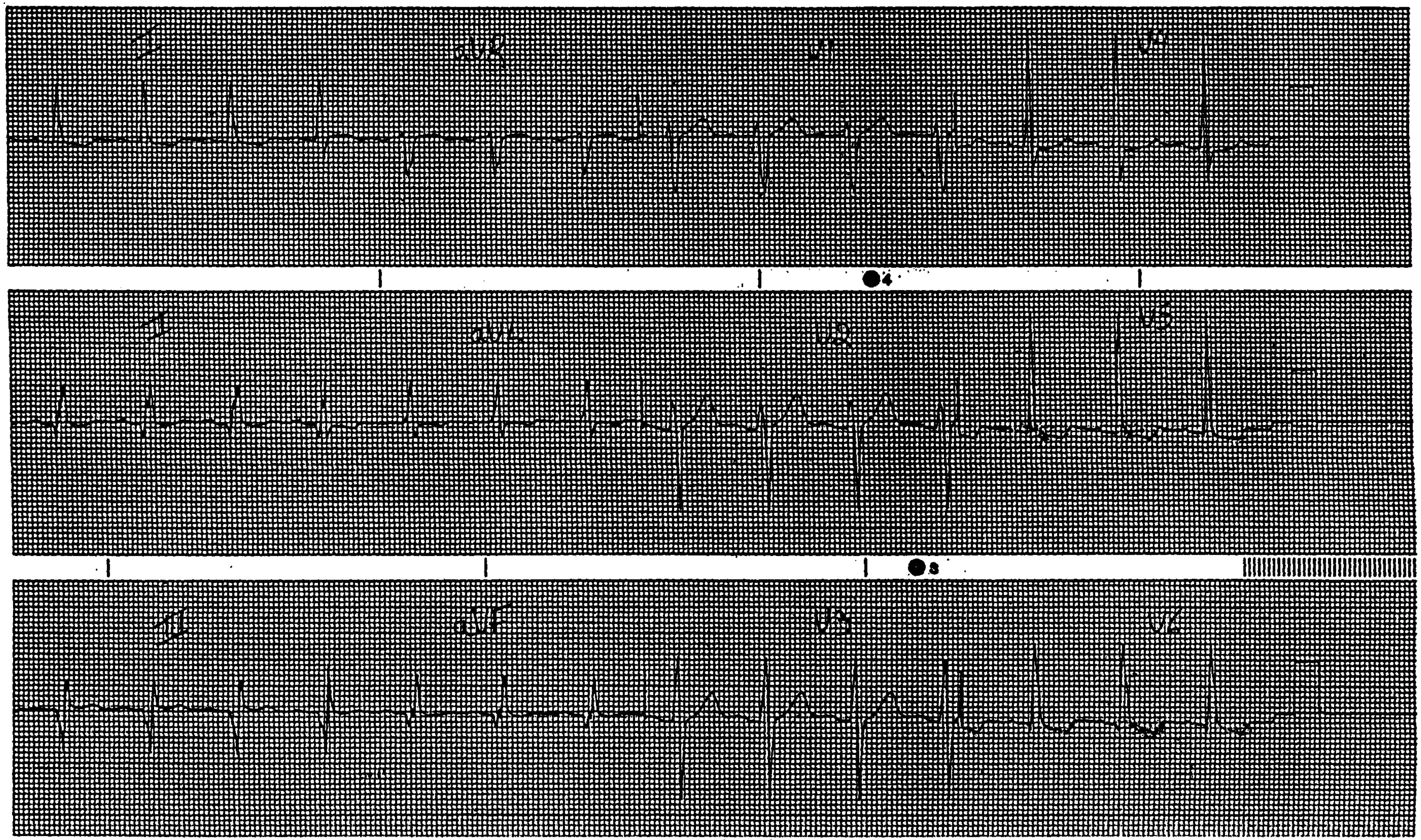
Hyperkalemia
- Peaked T waves across many leads, particularly from V2 to V6
- As hyperkalemia progresses, widening of QRS complex can occur

Appendix 3 Left Ventricular Hypertrophy with Strain

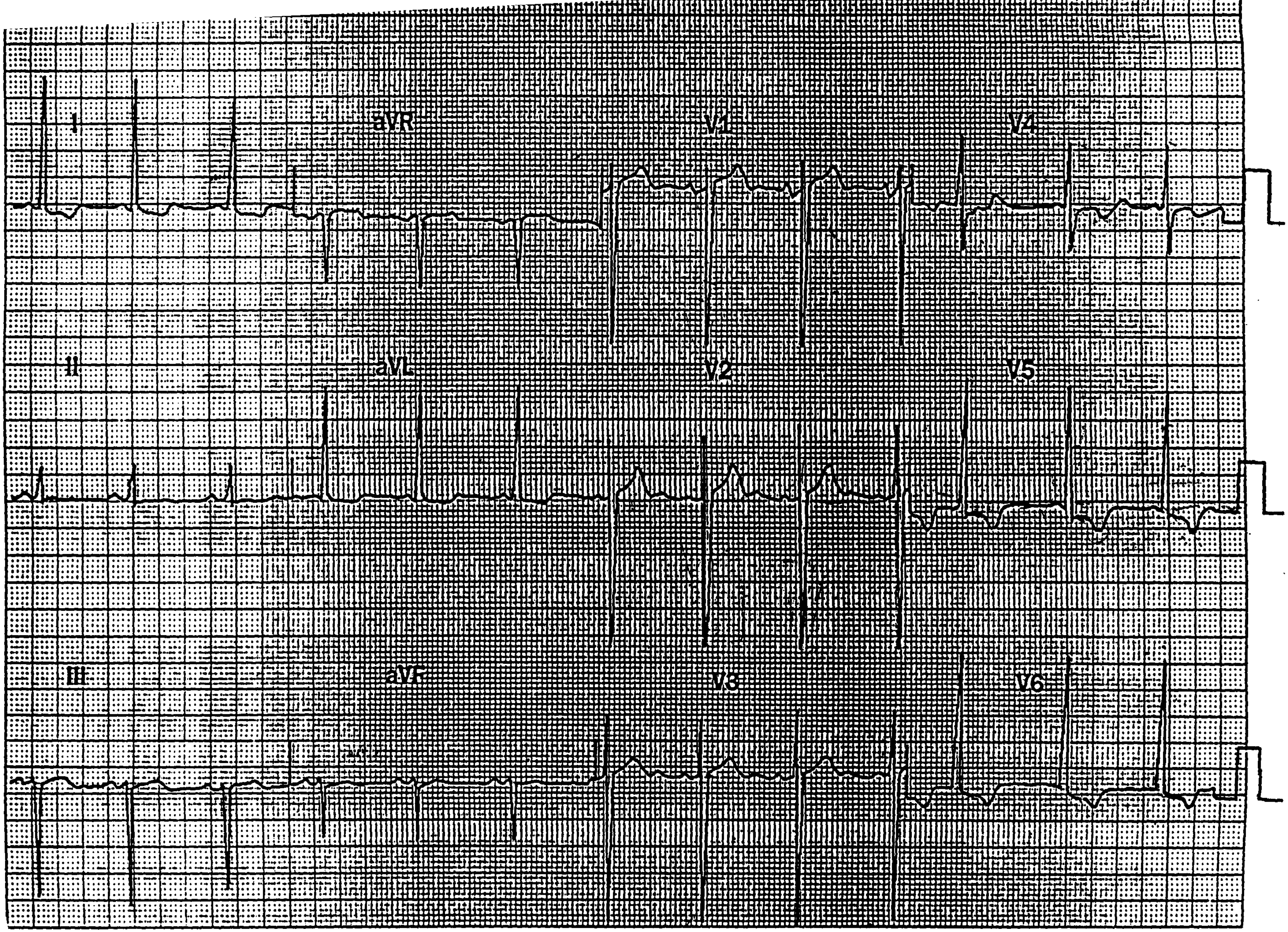
The diagnostic categories were presented in an "ECG Training Booklet" to participants. Attached is an example of how one diagnostic category (Left Ventricular Hypertrophy with Strain) was presented with four example ECGs in this booklet to participants.

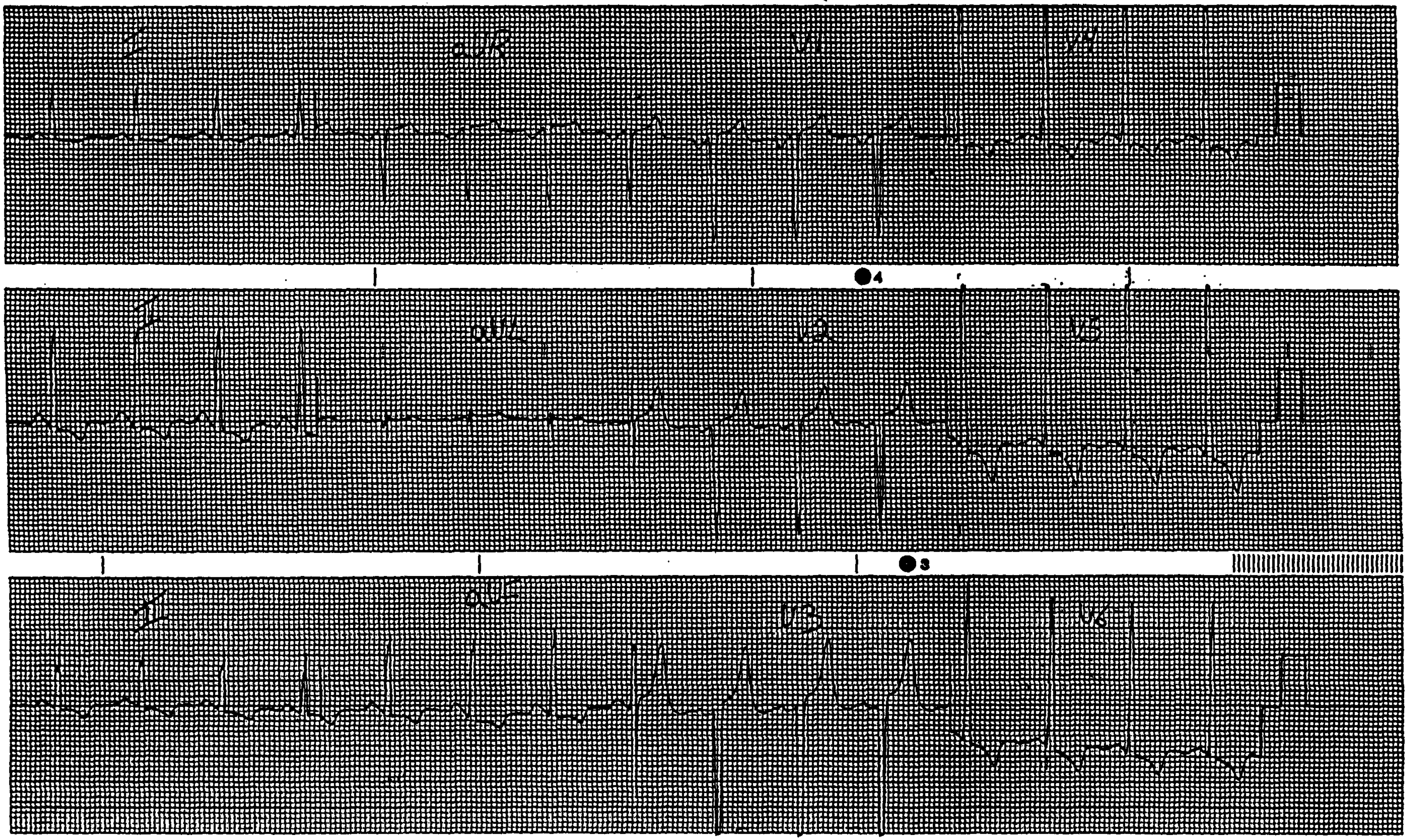
Left Ventricular Hypertrophy with Strain


I
aVR
V1
V4
II
aVL
V2
V5
III
aVF
V3
V6


I
aVR
V1
V4
II
aVL
V2
V5
III
aVF
V3
V6

On the answer sheet provided, write down the features that are present in the following ECGs


I
aVR
V1
V4
II
aVL
V2
V5
III
aVF
V3
V6


I
aVR
V1
V4
II
aVL
V2
V5
III
aVF
V3
V6

Appendix 4 Decision-Making Instruction Answer Booklets

Answer booklets were designed to specifically reflect the decision-making instructions. That is, participants received different answer booklets according to the decision-making instructions they were randomly assigned to. The booklets are presented below by experiment and the label assigned to that decision-making instructions.

a. Experiment 1

Feature First

RESPONSE SHEET

For each ECG, work down the Response Sheet List and indicate all features that can be seen.

- ☐ Q waves
  - ☐ present in V1 to V4
  - ☐ present in II, III, AVF
- ☐ R waves present in V1 and V2
- ☐ R wave in V1 is > 7 mm
- ☐ Predominant R wave in V1
- ☐ R wave getting progressively smaller in V2 to V4
- ☐ Large S waves in V1 and V2
- ☐ Large R waves in V5 and V6
- ☐ Sum of largest S wave in V1 or V2 and R wave in V5 or V6 is > 35 mm
- ☐ RSR' complex (rabbit ears)
  - ☐ V1 and V2
  - ☐ V5 and V6
- ☐ Widened QRS complexes
- ☐ Peaked T waves in more than 1 lead (particularly V2 to V6)
- ☐ T wave inversion
  - ☐ V4 to V6
  - ☐ II, III, AVF
  - ☐ I, AVL
  - ☐ V1, V2
- ☐ ST depression
  - ☐ V1 and V2
  - ☐ V4 to V6
  - ☐ II, III, AVF
  - ☐ I, AVL
- ☐ ST elevation
  - ☐ across multiple leads
  - ☐ V1 to V4
  - ☐ II, III, AVF

Now that you have gone through the feature list, make a diagnosis. Use the disease list to help you.

Please write your diagnosis in the space below.

Diagnosis:____

Argue for Alternatives

RESPONSE SHEET

For each ECG, work down the Response Sheet List and indicate all features that can be seen.

- ☐ Q waves
  - ☐ present in V1 to V4
  - ☐ present in II, III, AVF
- ☐ R waves present in V1 and V2
- ☐ R wave in V1 is > 7 mm
- ☐ Predominant R wave in V1
- ☐ R wave getting progressively smaller in V2 to V4
- ☐ Large S waves in V1 and V2
- ☐ Large R waves in V5 and V6
- ☐ Sum of largest S wave in V1 or V2 and R wave in V5 or V6 is > 35 mm
- ☐ RSR' complex (rabbit ears)
  - ☐ V1 and V2
  - ☐ V5 and V6
- ☐ Widened QRS complexes
- ☐ Peaked T waves in more than 1 lead (particularly V2 to V6)
- ☐ T wave inversion
  - ☐ V4 to V6
  - ☐ II, III, AVF
  - ☐ I, AVL
  - ☐ V1, V2
- ☐ ST depression
  - ☐ V1 and V2
  - ☐ V4 to V6
  - ☐ II, III, AVF
  - ☐ I, AVL
- ☐ ST elevation
  - ☐ across multiple leads
  - ☐ V1 to V4
  - ☐ II, III, AVF

Now that you have gone through the feature list, list all the possible diagnoses relevant to the features you have identified. Then argue for and against each possible diagnosis. Once you have done this, make a final diagnosis. Use the disease list to help you.

Please write your possible diagnoses in the space below.

Possible Diagnosis:______________________________
Arguments for: **Arguments against:**

Possible Diagnosis:______________________________
Arguments for: **Arguments against:**

Possible Diagnosis:______________________________
Arguments for: **Arguments against:**

Possible Diagnosis:______________________________
Arguments for: **Arguments against:**

Possible Diagnosis:______________________________
Arguments for: **Arguments against:**

Please write your final diagnosis in the space below.

Diagnosis:______________________________

Rank Features

RESPONSE SHEET

For each ECG, work down the Response Sheet List and indicate all features that can be seen.

☐ Q waves
- ☐ present in V1 to V4
- ☐ present in II, III, AVF

☐ R waves present in V1 and V2

☐ R wave in V1 is > 7 mm

☐ Predominant R wave in V1

☐ R wave getting progressively smaller in V2 to V4

☐ Large S waves in V1 and V2

☐ Large R waves in V5 and V6

☐ Sum of largest S wave in V1 or V2 and R wave in V5 or V6 is > 35 mm

☐ RSR' complex (rabbit ears)
- ☐ V1 and V2
- ☐ V5 and V6

☐ Widened QRS complexes

☐ Peaked T waves in more than 1 lead (particularly V2 to V6)

☐ T wave inversion
- ☐ V4 to V6
- ☐ II, III, AVF
- ☐ I, AVL
- ☐ V1, V2

☐ ST depression
- ☐ V1 and V2
- ☐ V4 to V6
- ☐ II, III, AVF
- ☐ I, AVL

☐ ST elevation
- ☐ across multiple leads
- ☐ V1 to V4
- ☐ II, III, AVF

Now that you have gone through the feature list, rank the features according to obviousness and vividness.

Once you have done this, make a diagnosis. Use the disease list to help you.

Please write your diagnosis in the space below.

Diagnosis:____

b. Experiment 2

Combined Instructions

RESPONSE SHEET

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.

- ☐ Q waves
  - ☐ present in V1 to V4
  - ☐ present in II, III, AVF
- ☐ R waves present in V1 and V2
- ☐ R wave in V1 is > 7 mm
- ☐ Predominant R wave in V1
- ☐ R wave getting progressively smaller in V2 to V4
- ☐ Large S waves in V1 and V2
- ☐ Large R waves in V5 and V6
- ☐ Sum of largest S wave in V1 or V2 and R wave in V5 or V6 is > 35 mm
- ☐ RSR' complex (rabbit ears)
  - ☐ V1 and V2
  - ☐ V5 and V6
- ☐ Widened QRS complexes
- ☐ Peaked T waves in more than 1 lead (particularly V2 to V6)
- ☐ T wave inversion
  - ☐ V4 to V6
  - ☐ II, III, AVF
  - ☐ I, AVL
  - ☐ V1, V2
- ☐ ST depression
  - ☐ V1 and V2
  - ☐ V4 to V6
  - ☐ II, III, AVF
  - ☐ I, AVL
- ☐ ST elevation
  - ☐ across multiple leads
  - ☐ V1 to V4
  - ☐ II, III, AVF

Once you have done this, make a diagnosis. Use the disease list to help you.

Please write your diagnosis in the space below.

Diagnosis:____

No Instructions

RESPONSE SHEET

☐ Q waves
- ☐ present in V1 to V4
- ☐ present in II, III, AVF

☐ R waves present in V1 and V2

☐ R wave in V1 is > 7 mm

☐ Predominant R wave in V1

☐ R wave getting progressively smaller in V2 to V4

☐ Large S waves in V1 and V2

☐ Large R waves in V5 and V6

☐ Sum of largest S wave in V1 or V2 and R wave in V5 or V6 is > 35 mm

☐ RSR' complex (rabbit ears)
- ☐ V1 and V2
- ☐ V5 and V6

☐ Widened QRS complexes

☐ Peaked T waves in more than 1 lead (particularly V2 to V6)

☐ T wave inversion
- ☐ V4 to V6
- ☐ II, III, AVF
- ☐ I, AVL
- ☐ V1, V2

☐ ST depression
- ☐ V1 and V2
- ☐ V4 to V6
- ☐ II, III, AVF
- ☐ I, AVL

☐ ST elevation
- ☐ across multiple leads
- ☐ V1 to V4
- ☐ II, III, AVF

Please write your diagnosis in the space below.

Diagnosis:____

Appendix 4 Ark, Brooks, Eva (submitted)

March 16, 2005

The best of both worlds:
Clinical teachers need not guard against teaching pattern recognition to novices

Tavinder K Ark,[1] Lee R Brooks,[1] and Kevin W Eva[2]

Department of Psychology[1] and Department of Clinical Epidemiology and Biostatistics[2]
McMaster University, Hamilton, Ontario, Canada

Word count: 3,486 (excluding title page, abstract, tables and references)

Direct correspondence to:

Kevin W. Eva
Department of Clinical Epidemiology and Biostatistics
Program for Educational Research and Development
MDCL 3522
McMaster University
1200 Main Street West
Hamilton, Ontario, Canada
L8N 3Z5

Phone: (905) 525-9140 x27241
Fax: (905) 572-7099
E-mail: evakw@mcmaster.ca

Abstract

Purpose: There has been much debate in the medical education literature regarding the extent to which feature driven and non-analytic reasoning strategies define expertise, but the relative value of teaching one or the other strategy remains uncertain. The purpose of this study was to use an experimental methodology to compare the diagnostic accuracy achieved upon receiving instruction to use each strategy in isolation and relative to that of a combined approach.

Method: Students were trained to identify features on electrocardiograms (ECGs) and assign diagnoses. One quarter of participants were instructed to carefully identify all features present before assigning a diagnosis (feature first). One quarter of participants were taught to trust impressions of similarity and diagnose based on this impression (first impression). The remaining half of participants received both sets of instructions. Half were given both sets of instructions explicitly (explicit combined), while the remaining participants were simply informed that some of the test ECGs had been seen during training (implicit combined).

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

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

Introduction

To determine if a patient has suffered an acute anterior myocardial infarction requires the identification of certain diagnostic features, such as chest pain radiating down the left arm and ST elevations in leads V1 to V4 in an electrocardiogram (ECG). Research in both cognitive psychology and medical education propose two classes of mechanisms whereby this task can be performed; analytic and non-analytic. Both of these classes have been operationalized in many ways and should not be considered mutually exclusive.[i,ii]

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. Among other ways, analytic approaches tend to be invoked with admonitions to (a) carefully identify and consider all clinical features before generating a diagnostic hypothesis, or (b) 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 non-analytic basis of clinical reasoning.[iii] Central to this view is the argument that rapid processes such as pattern recognition provide a valid alternative mechanism whereby diagnostic decision-making might be informed. 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.[iv]

It is now broadly recognized that this form of reasoning nicely describes much of the activity in which experts (having acquired a vast repertoire of cases) engage as many researchers have identified the inadequacy of constraining models of expertise in clinical reasoning to single classes of mechanisms, often pointing to the finding that process differences have failed to capture the essence of expertise.[v,vi] Still, uncertainty remains regarding how, and if, non-analytic processes can be operationalized and invoked to the benefit of learners.[vii] The study reported in this paper was designed to assess the potential strengths/weaknesses of adopting a multi-faceted approach to clinical instruction when training absolute novices.

Simple verbal reports of reasoning strategy are insufficient for this purpose because (a) by definition, non-analytic processes are often unavailable to conscious introspection,[viii] and (b) the requirement to verbalize can alter the reasoning processes themselves,[ix] thereby calling into question the validity of the inferences that can be drawn.[x] As such, it is necessary to carefully design experimental manipulations to determine the relative benefit of instruction to be analytic or non-analytic during diagnostic decision-making. Two previously published studies of dermatological diagnosis, one with residents,[xi] and one with medical students,[xii] have done just that.

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., non-analytic) approach by providing diagnostic labels based on one's first impression of the slide. In addition, 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 non-analytic and analytic processes, respectively, impacted upon the diagnosis. The interaction between these variables and type of instruction (feature first vs. similarity-based) provided an approximation of the extent to which instruction altered participants' typical diagnostic strategy. For example, if similarity-based instruction leads to a greater difference between similar/dissimilar cases than feature-based instruction, it suggests that the similarity-based teaching increased the extent to which non-analytic processes were operational.

Of direct relevance to the current study is that both of these experiments 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 non-analytic 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. Distinct between the two studies was that Regehr, et al.[xi] found that residents were only amenable to feature-oriented instruction whereas Kulatunga-Moruzi, et al.[xii] found that medical students were only amenable to similarity-based instruction.

This discrepancy is supportive of Schmidt, Norman, and Boshuizen's contention that novices learn analytic rules that subsequently provide a foundation on which experiential knowledge can be built (i.e., that diagnosticians come to rely more heavily on non-analytic processes as they gain more experience).[xiii] What it does not address, however, is whether or not there might be benefits to adopting a multi-faceted instructional approach when teaching novices as each participant was instructed solely to use the "feature-first" or "similarity-based" models of diagnostic reasoning.

Doing so 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,[xiv] it must also be recognized that there are drawbacks to reviewing cases in a more systematic and non-biased manner. Norman, et al., for example, taught undergraduate psychology students to diagnose electrocardiograms (ECGs) in a feature-driven manner by having them systematically identify features before considering diagnostic possibilities.[xv] Students in this condition identified more features than 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 provides a first opportunity to assess the advantage of instruction to adopt a multi-faceted (i.e., combined reasoning strategies) approach in that it essentially compares instruction in which learners are told to be non-analytic (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 are given in isolation. The current study builds on this result in two ways. First, we have created a more purely non-analytic instruction condition comparable to the similarity-based instructions of Regehr, et al. to determine whether or not combined instruction truly provides diagnostic advantage in this domain relative to instruction to adopt either an analytic or non-analytic strategy in isolation. That is, can absolute novices take advantage of the best that both worlds have to offer? Second, we have operationalized the combined instruction three different ways: (1) By having participants complete a non-analytic and analytic consideration of each ECG sequentially (i.e., providing a diagnostic decision before carrying out the more analytic search), (2) by explicitly telling learners to trust feelings of familiarity in addition to carefully considering features, and (3) by implicitly informing participants about the benefits of similarity by warning them that some test cases were seen during training. Differences across these conditions, it was anticipated, would provide guidance regarding how educational prescriptions should be provided should a multi-faceted instruction strategy prove beneficial.

Methods

Participants

Forty-eight undergraduate Psychology students enrolled at McMaster University participated in this experiment for course credit. None had previous experience with

ECGs. Ethics approval was received from the Hamilton Health Sciences Research Ethics Board.

Design and Procedure

I - Training Phase

Learning took place during one-on-one training sessions with the experimenter. All participants were presented with general information regarding the 12 leads in an ECG and taught to read a normal ECG waveform using materials created for the purpose of teaching medical students.[xv] The experimenter taught participants each of 10 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 (a) RSR' (rabbit ears) present in V1 and V2 and (b) a widened QRS complex.[xvi]

For each diagnostic category, participants were then presented with four examples in sequence. The experimenter 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.

II - Practice Phase

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

To emphasize a non-analytic reasoning strategy, participants in the "First Impression" 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 re-diagnose each ECG, keeping their initial diagnosis in mind, but changing it if necessary.

To emphasize an analytic reasoning strategy, participants assigned to the "Feature First" 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.

Participants in the "Implicit Combined" reasoning 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.

Participants in the "Explicit Combined" reasoning 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 first impression group). In addition, whenever an incorrect diagnosis was assigned, immediate feedback was given that reinforced the instructions provided.

III – Test Phase

During the test phase, participants were asked to diagnose twenty ECGs, ten of which were novel and ten of which 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 that of the practice phase for each participant.

Results

Diagnostic Accuracy

The mean diagnostic accuracy is reported in Table 1 as a function of condition and novelty of the stimuli. A repeated measures analysis of variance was performed on these data with condition included as the between subjects grouping factor. Slide (nested within old/new) and old/new were included as within subjects repeated measures. There was a significant main effect of condition, $F(3,792) = 9.22$, $p<0.001$. 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 first impression (41%) groups ($F(1,396) > 9.0$, $p < 0.01$ in all cases). There was no difference between the feature first and the first impression conditions, nor between the implicit combined and explicit combined conditions, $F(1,396) < 1$, $p > 0.4$ in both cases.

TABLE 1:

Mean diagnostic accuracy (%) during test as a function of both instruction an novelty of stimulus

Condition	Old ECGs	New ECGs	Total
First Impression (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

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), $F(1,792) = 49.03$, $p<0.001$. The interaction between condition and new/old cases was non-significant, $F(3,792) < 1$, $p > 0.3$.

When the first impression group was given the opportunity to revise their diagnosis 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, $F(1,396) < 3$, $p > 0.05$.

Feature Identification

For each stimulus, features were categorized as (a) hit indicative - a feature that was present and indicative of the correct diagnosis for that particular ECG, (b) hit not indicative - a feature that was present in the ECG but not indicative of the correct diagnosis, or (c) 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 as a function of classification and condition is illustrated in Table 2.

TABLE 2:

Mean number of features identified during the test phase as a function of both feature classification and instruction

Condition	Hits Indicative	Hits not Indicative	False Alarms	Total Features
First impression	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

There was a significant main effect of the total number of features identified by condition, $F(3,792) = 69.08$, $p<0.001$. Post hoc analyses revealed that participants in the feature first condition identified more features, on average, than participants in any of the other three conditions ($F(1,396) > 16$, $p < 0.01$ for each comparison). The same pattern of results occurred for both hits not indicative ($F(1,396) < 11$, $p < 0.001$ in all cases) and false alarms ($F(1,396) < 65$, $p < 0.001$ in all cases). In contrast, no difference was observed in the mean number of hits indicative identified as a function of condition ($F(3,792 = 2.74$, $p > 0.05$. The main effect of old/new and the interaction between new/old cases and condition were also non-significant in all feature analyses, $F(3,792) < 0.48$, $p > 0.?$ in all cases.

Discussion

Like many areas of science, there has been a tendency for theorists in medical education to strive to identify <u>the</u> truth: <u>The</u> hallmark of expertise and <u>the</u> 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 those who have specialized.[i,ii] The results of this experiment 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. The advantage of this combined approach was observed 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.

Moreover, the advantage of a combined approach is observed when the similarity-based instruction is given implicitly or explicitly, or if the feature first or first impression instructions are given simultaneously or sequentially

While the reductionist approach and small sample size used in this study could be perceived as limitations, the advantages of such a strategy have been well described.[xvii] 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) utilized are perfectly compatible with those used during actual medical training. 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. Our participants in this study, absolute novices in ECG diagnosis, achieved performance levels on this limited task equivalent to that of senior medical students regardless of condition. Those who received combined instructions revealed diagnostic accuracy equivalent to that of second year residents.

Further research would be 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 are supported by the data we have collected. First, the analysis of feature calls across condition suggests, as do the data of Norman, et al.,[xv] 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 first impression or the combined conditions. Taken in combination with the finding that participants in all three conditions were equally likely to identify features consistent with the correct diagnosis 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 can not be reconciled into a coherent diagnostic entity.

Second, and perhaps more surprising is the finding that participants in the first impression 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, even when the tentative diagnoses are generated by the diagnostician him/herself.[xviii] The impact of this biasing is that it 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.[iv,xix] 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.[xx] Finding that the first impression group was able to overcome a decision in favour 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.

An interesting, but unanticipated finding is the lack of interaction between the new/old variable and instruction – previously seen cases were diagnosed better than new cases is consistent with past research,[xxi,xxii] and highlights the role of similarity in guiding diagnostic decisions regardless of the specific instruction provided. We anticipated, however, that the difference would have been smallest in the feature first group that was told to carefully identify features prior to considering diagnostic possibilities. The lack of an interaction between new/old and instruction is consistent with Regehr, et al.'s findings,[xi] collected from residents, but counter to Kulatunga-Moruzi, et al.'s findings,[xii] collected from medical students. At present we tentatively interpret these differences as indicative of experience. Residents have sufficient experience to make it difficult for them to avoid using similarity-based reasoning when considering new cases. As such, being told to use similarity induces little change. In contrast, as indicated in the introduction, medical trainees are often warned against formulating diagnostic hypotheses before carefully considering the evidence; in this light it is not surprising that being told to formulate hypotheses early in a case encounter had an effect in the study performed by Kulatunga-Moruzi and colleagues. The lack of interaction in the current study, performed with absolute novices, might suggest that humans have a spontaneous tendency to allow similarity to guide their decisions, but that this tendency is dampened during the early years of medical training due to traditional clinical instruction to be objective.

Again, the conjectures in the preceding paragraph are purely speculative at this point, but regardless of the reason for the presence/absence of the new/old by instruction interaction, the main effects of all three studies have very clear educational implications. Clinical teachers should not guard against the use of non-analytic 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. 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.[vii] The current 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 non-analytic strategy is used in isolation.

References

[i] Eva KW. What every teacher needs to know about clinical reasoning. Med Educ 2005; 39: 98-106.

[ii] Custers EJFM, Regehr G, Norman GR. Mental representations of medical diagnostic knowledge: A review. Acad Med 1996; 71: S55-61.

[iii] Norman GR, Brooks LR. The non-analytic basis of clinical reasoning. Adv Health Sci Educ 1997; 2: 173-84.

[iv] Hatala R, Norman GR, Brooks LR. Impact of a clinical scenario on accuracy of electrocardiogram interpretation. J Gen Int Med 1999; 14: 126-9.

[v] Barrows HS, Norman GR, Neufeld VR, Feightner JW. The clinical reasoning process of randomly selected physicians in general medical practice. Clin Inv Med 1982; 5: 49-56.

[vi] Elstein AS, Shulman LS, Sprafka SA. Medical problem solving: An analysis of clinical reasoning. Cambridge, MA: Harvard University Press, 1978.

[vii] Coderre S, Mandin H, Harasym PH, Fick GH. Diagnostic reasoning strategies and diagnostic success. Med Educ 2003; 37: 695-703.

[viii] Bargh JA, Chartrand TL. The unbearable automaticity of being. Amer Psychol 1999; 54: 462-479.

[ix] Roediger HL. Implicit memory: Retention without remembering. Amer Psychol 1990; 45: 1043-56.

[x] 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 (Volume 8, pp.41-69). New York: Nova Science Publishers, Inc, 2002.

[xi] Regehr G, Cline J, Norman GR, Brooks LR. Effects of Processing Strategy on Diagnostic Skill in Dermatology. Acad Med 1994; 69 (Suppl.): S34-6.

[xii] 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: 110-6.

[xiii] Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise: Theory and implications. Acad Med 1990; 65: 611-21.

[xiv] 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 (Suppl.):S67-9.

[xv] 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. Cognit Instruct 2000; 17: 433-48.

[xvi] Schamroth, L. An introduction to Electrocardiography (6th Edition). Oxford: Blackwell Scientific Publications, 1982.

[xvii] Norman GR, Schmidt HG. Effectiveness of problem-based learning curricula: Theory, practice, and paper darts. Med Educ 2000; 34: 721-8.

[xviii] Eva KW, Brooks LR. The Under-weighting of Implicitly Generated Diagnoses. Acad Med 2000; 75: S81-3.

[xix] Brooks LR, LeBlanc VR, Norman GR. On the difficulty of noticing obvious features in patient appearance. Psych Sci 2001; 11: 112-7.

[xx] Eva KW. The influence of differentially processing information on diagnostic decision-making [dissertation]. McMaster University, Hamilton, Ontario, Canada, 2001.

[xxi] Allen SW, Norman GR, Brooks LR. Experimental Studies of Learning Dermatologic Diagnosis: The Impact of Examples. Teach Learn Med 1992; 4: 35-44.

[xxii] Brooks LR, Norman GR, Allen SW. Role of Specific Similarity in a Medical Diagnostic Task. J Exp Psych: Gen 1991; 120: 278-87.