Advances in Health Sciences Education **8:** 17–26, 2003.

Practice Makes Perfect: The Critical Role of Mixed Practice in the Acquisition of ECG Interpretation Skills

ROSE M. HATALA, LEE R. BROOKS and GEOFFREY R. NORMAN*
Program for Educational Research and Development, Building T-13, McMaster University, Hamilton, ON, L8S 4K1, Canada
*(*author for correspondence, E-mail: norman@mcmaster.ca)*

Abstract. *Purpose.* To examine the effect of instructional format on medical students' learning of ECG diagnosis. *Method.* Two experiments employed different learning and practice methods. In the first, students were randomly allocated to one of two instructional approaches, one organized around features (e.g., QRS voltage) and the other around diagnostic categories (e.g., bundle branch blocks), followed by a practice phase. In the second experiment, the instruction was standardized, and students were randomly allocated to one of two practice phases, either "contrastive" where examples from various categories are mixed together, or "non-contrastive" where all the examples in a single category are practiced in a single block. *Results.* In the first experiment, there was no significant differences in students' diagnostic accuracy on novel ECG examples. In the second experiment, students exposed to the contrastive approach in the practice phase had superior diagnostic accuracy (46%) compared to 30% accuracy for the non-contrastive session, $p < 0.05$). *Conclusion.* These experiments highlight two important features in the design of instructional materials. First, learning around the features of the problem (analogous to problem – based learning) may have no advantages over learning the category. Second, the design and organization of deliberate practice can result in significant learning gain.

Key words: learning, practice

To many, the design of medical curricula appears the reverse of common sense. Students begin by learning the underlying mechanisms of disease through the basic sciences, proceed to the learning of disease categories, and finally see some examples of the mechanisms in patients. Commonly, the patient examples they do acquire are a demonstration after an instructional session, so are seen as examples of a particular condition. As a consequence, they tend to be oriented to "the signs and symptoms of X" and not to "the differential diagnosis of Y". Once they reach the ward or clinic, they find that the world proceeds in exactly the opposite direction. Instead of starting with a disease, they begin with a patient presentation, and must consider all the multiple and overlapping causes of the condition. They may well discover that their knowledge is organized inadequately, and they have not acquired the knowledge needed to differentiate one cause from another.

It would seem reasonable to go at it the other way around, beginning with patient problems and then learning the concepts necessary to understand the problem. Indeed, this is the basic assumption of problem based learning (PBL), where the curriculum is structured so that all the relevant concepts are acquired in the course of solving a problem. Clearly the idea has considerable appeal, and many authors have elaborated the potential advantages of this approach (Walton and Matthews, 1989; Norman and Schmidt, 1992). In particular, it has been suggested, based on basic research in psychology, that PBL should lead to increased "transfer" – an enhanced ability to apply learned concepts to the solution of new problems (Norman and Schmidt, 1992; Needham and Begg, 1989).

Paradoxically, evidence has shown that differences in performance between students from PBL and conventional curricula are small, absent, or occasionally negative (Vernon and Blake, 1993; Albanese and Mitchell, 1993; Colliver, 2000). A number of explanations have been advanced for these findings, including inadequacy of the theoretical rationale (Norman and Schmidt, 2000), methodological issues (Albanese, 2000), and the possibility that curriculum level experiments contain so many sources of variation that it is impossible to detect the experimental effects (ten Cate, 2001). But it is also possible that a PBL curriculum does not result in increased transfer. Bransford (1999) has argued that learning around a problem may actually impede transfer, since the important concepts are tightly bound to the problem context. He suggests that a prerequisite for successful transfer is a deep understanding of the relevant conceptual knowledge. Resolution of this issue must await evidence derived from more carefully controlled experiments.

It is also possible that any effect of the initial curriculum format may be minimal, and that a far more crucial determinant of problem-solving is practice with additional examples. The literature on expertise emphasizes the importance of what Ericsson has termed deliberate practice: individualized, supervised training on a specific aspect of performance through repetition and refinement (Ericsson and Charness, 1994). Experts in diverse fields, including the arts, sports, and science, have spent many more hours in deliberate practice than comparable amateurs in the field. Through repetition and reflection, the individual gains a deeper understanding of the features of the problem that are key to its solution. The repetition of a general concept in many different problem contexts may result in its availability as a solution for new problems, as Bransford (1999) conjectured.

Neither PBL nor traditional curricula explicitly address the role of practice in learning diagnosis. Typically, instructors show students a series of examples of each diagnostic category, beginning with prototypical and leading to less typical presentations. This may not be an optimal method for learning to distinguish among alternatives. Avrahami et al. (1997) demonstrated that when students were asked to teach other students how to correctly identify examples of a new category, the student-"teachers" chose to show their student-"learners" a few typical examples of the category, then contrasted these with examples not of the category, and then had the students practice on borderline examples. In other words, the students intuitively taught each other by contrasting category members with non-members, in order to highlight the relevant features of category membership.

This basic strategy of contrasting examples of different categories has been formalized as a "contrastive approach" (McKenzie, 1998). Through the juxtaposition of examples from different categories, students learn to extract the critical features which discriminate between categories. Klayman and Brown (1993) compared the effectiveness of a contrastive to a non-contrastive approach in teaching students about 2 fictitious diagnoses (zymosis and proxitis). The non-contrastive group learned the typical features of one diagnosis, then practiced on a series of examples, then learned the second diagnosis followed by a practice phase. The contrastive group learned both diagnoses at once and then practiced on the entire set of examples mixed together. In a subsequent test phase, the contrastive learners were better able to distinguish between new cases of the two "diseases". However, these experiments had numerous constraints as the students only learned two diseases, learning took place by experience with hundreds of examples without formal instruction about the relation between features and diagnoses, and the features were not coherent and were chosen to give a clear advantage to contrastive strategies.

To explore these issues of instruction and practice in a controlled experiment, we focus on the instruction of ECG diagnosis to medical students. Successful ECG diagnosis involves knowing the rules pertaining to each diagnosis (i.e., criteria for Left Ventricular Hypertrophy), correctly identifying features on the ECG (i.e., measuring the voltage of R and S waves) and sorting the relevant from the irrelevant features.

A review of the literature pertaining to ECG instruction revealed that many instructional materials do not address the cognitive aspects involved in ECG diagnosis. Most of the studies taught each ECG diagnostic category in isolation, varying the superficial formats of instruction (lectures vs. computer-based instruction (Fincher et al., 1988) or vs. programmed texts (Stretton et al., 1967)) and did not show significant benefits for a particular instructional format. The single study which employed contrastive methods demonstrated significant improvement in student knowledge, from a baseline accuracy of 23.9% correct ECG diagnoses to a post-intervention accuracy of 74.4%. Unfortunately, there was no comparison to standard teaching techniques and thus the relative benefit of this instructional technique remains unclear (Kingston, 1979).

We will describe two experiments which emphasize cognitive aspects of learning and the role of deliberate practice. In the first experiment, we contrast two instructional approaches. The first strategy, analogous to a PBL approach, organized the materials around features of the ECG – QRS morphology, the ST segment, etc. W will call this "feature-based" instruction (FB). The alternative instruction was organized in the traditional manner around the diagnostic categories (ventricular hypertrophy, MI), which we will call "category based" (CB) instruction.

In the second study, following Bransford's (1999) recommendations, we used a category – based approach to ECG instruction, and manipulated the practice phase. The practice phase was oriented towards giving students ample opportunity to apply the learned concepts to new problems. The experimental manipulation explored the effects of contrastive practice, where students had the opportunity to practice diagnosing the ECG's in a mixed series, versus non-contrastive practice, where students were provided with a series of ECG's of each condition of graded difficulty, then go on to the next diagnosis.

Experiment #1: Feature-Based (FB) versus Category – Based (CB) Instruction

METHODS

Overview

In this study, contrasts were incorporated into both the instruction and the practice phase of the teaching sessions. First year medical students were assigned to one of two ECG instructional sessions. The sessions covered the same diagnoses and ECG examples but differed in the organization of the material. In the FB session, instruction was organized by ECG feature, to encourage students to compare and contrast competing diagnoses. In the CB session, instruction was organized in the traditional format, by ECG diagnosis. Examples were consistent with the instruction, so, for example, in the FB instruction, after the section on QRS voltage, students would be shown a series of ventricular hypertrophy ECGs; in the CB instruction, the same examples would follow the section on ventricular hypertrophy. Effectiveness of the instructional material was measured in post-test containing new ECG examples.

Population

The entire first-year cohort of McMaster medical students from the graduating class of 1999, (n = 100) was recruited to participate in the study at the end of their cardiac sub-unit in the first year. The cardiac sub-unit was the students' first exposure to cardiac physiology and pathophysiology, with instruction occurring over a one-month period based on tutorials, self-directed learning and formal teaching sessions. Student participation was anonymous and voluntary; participating students were paid a small honorarium for completing the post-test.

Materials

Course material was modified from a self-study ECG text (Underwood, 1993). The criteria for selecting diagnoses included the incidence of the diagnosis, the seriousness of the illness, diagnoses that were often confused in clinical practice, and diagnoses that could be taught during a single 2 hour teaching session. Arrhythmias were excluded, and the ECG diagnoses covered during instruction included hypertrophy, bundle branch blocks and myocardial infarction among others. The selected ECG examples were collected from numerous sources including patient charts, cardiologists' files and a textbook (Underwood, 1993). All ECGs were then redrawn in a standardized format by the audio-visual department, ensuring high-quality images.

Procedure

Three separate sessions were held, each for 1/3 of the class at the end of their one month cardiac sub-unit. The session was 2 hours in length, followed by a post-test. The first and third sessions were allocated to feature-based instruction, and the second session to category-based instruction. One author gave all of the sessions, in order to standardize the presentation style.

Category – Based (CB) Instruction. The 2 hour session began with a review of basic ECG interpretation, and then taught the following diagnoses: hypertrophy, bundle branch blocks, myocardial infarction, ischemia, pericarditis. After each diagnosis was reviewed, a series of examples of that diagnosis were presented to the students in a practice phase.

One example of each diagnosis was included in the instruction phase (i.e., for hypertrophy and bundle branch blocks, one example each of right and of left; for myocardial infarction and ischemia, one example of anterior changes and one of inferior). During the practice phase, 2 to 3 additional examples of each diagnosis were reviewed.

Feature – Based (FB) Instruction. The instructional examples were organized around the ECG features. Each instructional session began with a review of basic ECG interpretation, identical to that received by the learners in the CB group. Subsequent materials covered: QRS voltage (hypertrophy), QRS duration (bundle branch blocks), QRS morphology (bundle branch blocks, hypertrophy, myocardial infarction), ST segment (myocardial infarction, ischemia and pericarditis). During instruction, students were encouraged to compare and contrast the difference in the features between the competing diagnoses. The number of instructional examples varied from 2 to 5 depending on the topic. Before proceeding to the next topic, between 3 and 6 ECGs reviewing the relevant diagnoses within the current topic were presented for students in a practice phase (for example, in the section on QRS voltage, students were shown 5 ventricular hypertrophy ECG's and 1 normal, in the section on QRS morphology, students saw one MI and 2 bundle branch block ECG's).

All ECGs were identical to those used in conventional instruction. However, to contrast the competing diagnoses, 4 ECGs were shown twice during instruction, and 2 ECGs were shown twice in the practice phase. As a result, students were exposed to a total of 6 additional ECGs in the FB session compared to the CB session.

Outcome

The effectiveness of the educational interventions was assessed by a written test containing 9 novel ECGs. Students recorded the most likely diagnosis and the features in each ECG that led to the diagnosis. Responses were scored blindly for accuracy of ECG diagnosis, from incorrect(0) to partially correct (0.5) to completely correct(1); each participant had a maximal possible score of 9 for diagnostic accuracy. Data were analyzed using a repeated measures ANOVA with one between-subject factor (CB vs. FB) and one within-subject factor (Case).

RESULTS

Seventy-one of 100 potential students participated. Forty-eight students were allocated to the first and third(FB) sessions. Twenty-three students attended the second (CB) session.

There were no significant differences between the groups in diagnostic accuracy. The CB group's mean score was 5.68 (63%) on a 9-item post-test compared to 6.14 (68%) for the FB groups ($F(1,71) = 2.54$, $p = 0.12$).

DISCUSSION

The results of this study are disappointing. When feature – based instruction was employed, students' performance did not improve over a category-based approach. One possibility is that the initial instruction was confusing rather than illuminating, as it highlighted the discriminating features of each disorder (i.e., QRS voltage, or ST segment changes) but did not present an orderly list of typical features in each diagnostic category. Bransford (1999) has indicated that many studies of transfer are unsound because investigators pay insufficient attention to the initial learning. Avrahami et al. (1997) showed that students instructing others begin with typical cases. In Klayman and Brown's (1993) study with artificial diseases, students first learned the typical features of the disorder during an instruction phase that preceded the contrastive practice phase. Applying this to ECG instruction, students likely need a framework of typical ECG features in order to understand the disorder, before appreciating which features among the list are most discriminating. Simply giving students the list of discriminating features, as was attempted in this experiment, appears to be a less successful instructional strategy.

A second possibility is that, to the extent that practice is a critical determinant of mastery, the two instructional methods were very similar in the practice examples.. Examples following every chapter in the feature-based instruction were deliberately linked to the chapter topics, and this had the effect of creating relatively homogeneous subsets. For example, the section on QRS voltage contained 5 ventricular hypertrophy and one normal, the section on QRS duration contained 6 bundle branch blocks and one normal.

In the second experiment, in order to examine the effect of contrastive practice, we held instruction constant, and experimentally varying the nature of the practice.

Experiment #2: Contrastive vs. Non-Contrastive Practice

METHODS

Overview

First year medical students were assigned to one of two ECG instructional sessions. Both sessions covered the same diagnoses and ECG examples and differed only in the organization of the practice phase. In both sessions the instruction was organized in the category-based format, by ECG diagnosis. In the practice phase of the contrastive session, new ECGs were presented to the students for review, covering all of the diagnoses included in instruction. Emphasis was placed on comparing and contrasting sequential examples. In the practice phase of the non-contrastive session, the identical ECGs were re-organized and presented to the students at the end of each topic, thus students saw multiple examples of a single diagnosis at a time. Effectiveness of the instructional material was measured in a post-test containing new ECG examples.

Population

Two-thirds of McMaster's first-year medical students from the graduating class of 2002 (n = 100) were recruited to participate in the study at the beginning of their cardiac sub-unit. Sixty-six students participated (100% participation). Thirty-three students were allocated to each instructional session.

Student participation was anonymous and voluntary; participating students were paid a small honorarium for completing the post-test.

Materials

As in experiment #1, the ECG diagnoses covered during instruction were limited to myocardial infarction, ventricular hypertrophy, and bundle branch blocks, pericardit's and ischaemia. Identical ECGs were used for both sessions, but the organization of the ECGs differed.

Procedure

Two separate instructional sessions were held, each for 1/3 of the class at the end of their one month cardiac sub-unit. The instructional session was 2 hours in length, followed by a post-test. The first session was allocated to contrastive practice and the second session to non-contrastive practice. One author gave all of the sessions, in order to standardize the presentation style.

Non-contrastive practice. The 2 hour instructional session began with a review of basic ECG interpretation, and then taught the diagnoses sequentially, with 2 examples of each diagnosis highlighted during instruction. After each diagnostic category was taught, four new examples of that diagnosis (i.e., for hypertrophy, 2 examples of left and 2 examples of right ventricular hypertrophy) were presented to the students as a practice phase.

Contrastive practice. The 2 hour instructional session was organized as in Non-Contrastive Practice. After covering all of the diagnoses, students completed a practice phase where they reviewed the series of 12 novel ECGs, mixed across the diagnoses. These practice ECGs were identical to those used at the end of each diagnostic category in the Non-Contrastive Instruction. Students were encouraged to compare and contrast the difference in the features between the competing diagnoses.

Outcome. The effectiveness of the educational intervention was assessed by a written test containing 6 novel ECGs. Students recorded the most likely diagnosis and the features in each ECG that led to the diagnosis. Responses were scored blindly for accuracy of ECG diagnosis, from incorrect(0) to partially correct(0.5) to completely correct(1); each participant had a maximal possible score of 6 for diagnostic accuracy. As before, data were analyzed using a repeated measures ANOVA with one between-subject factor (Contrastive vs. Non-contrastive) and one within-subject factor (Case).

RESULTS

There was a significant benefit to the contrastive session. Overall diagnostic accuracy on the 6 item post-test was 2.8 (47%) (sd = 0.94) for the contrastive session compared to 1.8 (30%) (sd = 1.05) for the non-contrastive session ($F(1,65) = 14.58$, $p < 0.05$).

DISCUSSION

The current studies were designed to evaluate the effectiveness of different ECG teaching and practice sessions for novice learners. When feature-based versus category-based instruction was employed, in Experiment #1, no impact on student learning was demonstrated. In Experiment #2, the use of contrastive practice, where students saw the ECG's in a mixed sequence, led to higher diagnostic accuracy than seeing the same ECG examples during a practice phase that did not contrast between competing diagnoses. Contrastive practice, where students are encouraged to identify similarity and difference among examples of different categories, may be more effective in helping students notice the types of features that discriminate between competing diagnostic alternatives.

There are limitations to our experiments. Strictly speaking, students were not randomized to instruction, although we have no reason to suspect any systematic bias emerged from this allocation. In addition, in Experiment #2, the contrastive method was employed with the earliest group of students who would have had the lowest baseline knowledge. Despite this disadvantage, the contrastive learners demonstrated higher diagnostic accuracy after the intervention. A second limitation was the single teaching session employed in both experiments. A more sustained teaching effort, perhaps repeated over a series of seminars with increased opportunity for practice on new ECG examples, should lead to higher diagnostic accuracy. Perhaps for this reason, diagnostic accuracy was limited to 47% for the better performing contrastive learners in Experiment #2.

This series of experiments highlights two important features in the design of instructional materials. First, deliberate practice of the concepts highlighted in instruction is an important component of an instructional session. Deliberate practice allows the student to problem-solve in the domain, and helps to highlight the salient features of the diagnostic task. Moreover, deliberate practice with feedback is a critical component of maintaining competence (Guest, Regehr and Tiberius, 2001). Second, the order of examples in a practice session can have a large impact on performance. Contrastive methods used in a practice phase, after teaching the students the diagnostic categories and typical features within those categories, is an relatively more effective instructional strategy. The contrastive approach should be applicable to other perceptual and/or diagnostic domains where noticing features is an important component of the task, and the list of competing alternatives for a given diagnosis is limited.

These studies may also have implications for curriculum design in two ways. First, to the extent that Experiment #1 is analogous to a problem-based approach, where the student may use the problem to explore and contrast various diagnostic categories, this may be a no more effective strategy for initial learning than one organized around prototypical cases. More important, the results of Experiment #2 suggest that the structure of the practice phase is of at least equal importance to the design of the instruction phase. Medical educators may need to focus greater attention on the practice phase of the curricula, namely the clerkship, to explore the use of deliberate practice and the benefits of contrastive methods for diagnostic tasks. Introducing a practice component with specific manipulation of the sequence of examples and feedback can lead to more effective learning on the part of the student.

Acknowledgements

This research was supported by a grant from the ACMC/MRC Research in Medical Education Fund.

References

Albanese, M.A. & Mitchell, S. (1993). Problem based learning: A review of literature on its outcome and implementation issues. *Academic Medicine* **68**: 52–81.

Avrahami, J., Kareev, Y. et al. (1997). Teaching by examples: implications for the process of category acquisition. *Quarterly Journal of Experimental Psychology* **50A**: 586–606.

Bransford, J.D., Brown, A.L. & Cocking, R.R. (1999). *How People Learn: Brain, Mind, Experience and School*. Washington, DC: National Academy Press.

Colliver, J. (2000). Effectiveness of problem based learning curricula: Theory and practice. *Academic Medicine* **75**: 259–266.

Ericsson, K.A. & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist* **49**: 725–747.

Fincher, R., Abdulla, A. et al. (1988). Computer-assisted learning compared with weekly seminars for teaching fundamental electrocardiography to junior medical students. *Southern Medical Journal* **81**: 1291–1294.

Gick, M.L. & Holyoak, K.J. (1980). Analogical problem solving. *Cognitive Psychology* **12**: 306–355.

Guest, C.B, Regehr, G. & Tiberius, R.G. (2001). The life long challenge of expertise. *Medical Education* **35**: 78–81.

Kingston, M. (1979). Electrocardiograph course. *Journal of Medical Education* **54**: 107–110.

Klayman, J. & Brown, K. (1993). Debias the environment instead of the judge: An alternative approach to reducing error in diagnostic (and other) judgment tasks. *Cognition* **49**: 97–122.

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

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

Norman, G.R. & Schmidt, H.G. (1992). The psychological basis of problem based learning. *Academic Medicine* **67**: 557–565.

Norman G.R. & Schmidt, H.G. (2000). Effectiveness of problem based learning curricula: Theory, practice and paper darts. *Medical Education* **34**: 721–728.

Stretton, T., Hall, R. & Owen, S.G. (1967). Programmed instruction in medical education. *British Journal of Medical Education* **1**: 165–168.

Ten Cate, O. (2001). Whatever happened to the student? *Advances in Health Sciences Education* **6**: 81–88.

Underwood, D.A. (1993). *Clinical Electrocardiography: Self-assessment and Review.* Pennsylvania, PA: W.B. Saunders Company.

Vernon, D.T.A. & Blake, R,L. (1993). Does problem based learning work? A meta-analysis of evaluative research. *Academic Medicine* **68**: 550–563.

Walton, H.J. & Matthews, M.B. (1989). Essentials of problem based learning. *Medical Education* **23**: 542–558.