Diagnostic reasoning strategies and diagnostic success

S Coderre,[1] *H Mandin,*[1] *P H Harasym*[2] *& G H Fick*[2]

Purpose Cognitive psychology research supports the notion that experts use mental frameworks or 'schemes', both to organize knowledge in memory and to solve clinical problems. The central purpose of this study was to determine the relationship between problem-solving strategies and the likelihood of diagnostic success.

Methods Think-aloud protocols were collected to determine the diagnostic reasoning used by experts and non-experts when attempting to diagnose clinical presentations in gastroenterology.

Results Using logistic regression analysis, the study found that there is a relationship between diagnostic reasoning strategy and the likelihood of diagnostic success. Compared to hypothetico-deductive reasoning, the odds of diagnostic success were significantly greater when subjects used the diagnostic strategies of pattern recognition and scheme-inductive reasoning. Two other factors emerged as independent determinants of diagnostic success: expertise and clinical presentation. Not surprisingly, experts outperformed novices, while the content area of the clinical cases in each of the four clinical presentations demonstrated varying degrees of difficulty and thus diagnostic success.

Conclusions These findings have significant implications for medical educators. It supports the introduction of 'schemes' as a means of enhancing memory organization and improving diagnostic success.

Keywords content-specificity, diagnosis, hypothetico-deductive reasoning, pattern-recognition, problem-solving, scheme-inductive reasoning

Medical Education 2003;37:695–703

Introduction

One of the incentives for medical education research in the 1960s was a desire to inculcate medical students with the ability to solve clinical problems.[1] The results of such studies were consistent: no content-independent process was found that could differentiate experts from novices.[2,3] Successful and unsuccessful diagnosticians, both expert and novice, seemed to use a process termed hypothetico-deductive reasoning, a 'diagnosis-to-data' method of problem solving. Since Elstein's original description, hypothetico-deductive reasoning has been considered the most common form of diagnostic reasoning employed by clinicians. Authors such as Newell[4] and Groen[5] view hypothetico-deductive reasoning as a general reasoning strategy and refer to it as a 'weak method' of problem solving. Patel considers this method inefficient and prone to error.[6]

In contrast, 'pattern recognition' has been identified by other research as a very successful approach used by experts to solve clinical problems.[7,8] Before becoming more expert in problem solving, learners progress through several transitional stages characterized by different knowledge structures: elaborated causal networks, abridged networks, illness scripts and instance scripts.[8] Extensive experience eventually leads to acquisition of a repertoire of problems common to the domain of expertise termed 'illness scripts'. This repertoire permits problem resolution by recognition of new problems as ones that are similar or identical to old ones already solved, and the solutions are recalled. This phenomenon, labelled 'pattern recognition', likely represents a complex mental process requiring rapid retrieval of an appropriate match based on salient cues. Given the amount of expertise required, this second diagnostic reasoning strategy is generally unavailable to novice medical students.

Although research has revealed other strategies experts use in problem-solving,[9–11] none emerged that

Departments of [1]Medicine and [2]Community Health Sciences, University of Calgary, Faculty of Medicine, Calgary, Canada

Correspondence: Dr H Mandin, Division of Nephrology, Foothills Hospital, 1403 29th Street NW, Calgary, Alberta, T2N 2T9, Canada. Tel. 403-944-4300, Fax: 403-944-2876,
E-mail: henry.mandin@calgaryhealthregion.ca

Key learning points

There are three diagnostic reasoning strategies available: hypothetico-deductive reasoning, scheme-inductive reasoning and pattern recognition.

Experts and novices using scheme-inductive reasoning have fivefold greater odds of diagnostic success than using hypothetico-deductive reasoning.

Pattern recognition is associated with 10-fold greater odds of diagnostic success than using hypothetico-deductive reasoning. The aptitude to exploit this strategy occurs with expertise and its use by medical students is not advocated because of potentially dire consequences.

Diagnostic problem-solving strategy, level of expertise and clinical presentations are associated significantly to diagnostic success.

was considered appropriate for introduction into medical school curricula until 1994[12,13] when a predominantly inductive problem-solving strategy was introduced. Employing information from the literature on medical expertise[8,14–17] and drawing on observations made in smaller pilot studies[18,19] expert schemes[a] were recognized and defined.

'Schemes' were considered to reflect an organized knowledge structure for learning as well as a structure for diagnostic reasoning. They were drawn on paper like 'inductive trees' or 'road maps' to recreate the major divisions (or chunks) used by expert clinicians for both storage of knowledge in memory and its retrieval for solving problems[13,14] (see Appendix 1 for an example). This scheme-inductive process differs from the usual inductive process (reasoning from the clinical data to a diagnosis) in one important manner. It is not simply forward reasoning, 'as reasoning with a single diagnosis in mind'.[20] Decisions are explicitly at the forks in the road or branching of the tree. The organizational structure, or 'scheme', proceeds from alternative causal groups, through crucial 'tests', to exclusion of some alternative groupings and adoption of what is left. These tests may be based on an evaluation of symptoms, signs, or results of investigations, singly or in any combination. Consequently, Papa's prototype-based probabilistic model,[9] Bordage's semantic axes,[10] or the forceful features model[11] may serve as such 'tests', and consequently be utilized in the inductive process. After several branching points, when the number of diagnostic options has been considerably reduced, deductive reasoning or pattern recognition may be exploited. Finally, the scheme-inductive process is not content-independent; each of the organizational 'schemes' is specific to the clinical presentation.

Accordingly, there are three different diagnostic reasoning strategies available to learners. Deductive reasoning[b] (hypothetico-deductive) is used by experienced diagnosticians to include or exclude one or two diagnoses or as a fallback strategy when faced with unbounded clinical problems that are outside their domains of expertise. Inductive reasoning[c] ('scheme-inductive' problem solving) is a strategy used when faced with unbounded clinical problems and pattern recognition is not possible.[21] This type of problem solving represents the 'climbing of a conditional inductive tree'.[17] 'Pattern-recognition' is a successful diagnostic reasoning strategy available primarily to experts. These three strategies have not been previously compared to each other in a systematic way. The primary purpose of this study is to determine whether there is a relationship between diagnostic success and the diagnostic reasoning strategy utilized by both experts and novices when solving different types of clinical presentations. Specifically, this study will examine the effect of the independent variable of diagnostic reasoning strategy (hypothetico-deductive versus scheme-inductive

[a]The term 'scheme' was introduced at U of C for the organisational structure that spontaneously emanated from the mind of experts when organising knowledge domains. 'Schemata' on the other hand are mental structures used for both data storage and retrieval into 'bundles' of information (i.e. 'data schemata') and active processing or organisation of information (i.e. 'process schemata'). They facilitate learning and comprehension. 'Schemes' may not be identical to 'schemata' (it is difficult to know with certainty how memory is organized) but likely are close to the actual internal representations of the domain. The term 'scheme' is used here to reflect the fact 'schemata' remain unknown. Subsequent to the introduction of schemes for the promotion of learning, these same schemes proved to be invaluable for teaching, learning and problem-solving.

[b]Deductive reasoning is guided by generated hypotheses. Characteristically, during problem solving, the physician relates the general knowledge of the disease hypothesised to the specific signs and symptoms of the patient.

[c]Scheme-inductive reasoning is guided by a scheme. Characteristically, during problem-solving, the physician seeks specific information from the patient (symptoms, signs, laboratory data) that will distinguish between the categories of conditions at the branching points of the scheme. The specific information obtained from the patient is related to the general categories of conditions at the branch points of the scheme, and the presence or absence of these clinical findings leads to the adoption of one category and exclusion of the rest.

versus pattern recognition) on the dependent variable of diagnostic success (making a correct diagnosis). In addition, the effect of two other independent variables, expertise level (novice versus expert), and type of clinical situation (clinical presentations) on diagnostic success will be examined.

Methods

Examination construction

An examination for four clinical presentations (dysphagia, chronic diarrhoea, nausea with vomiting and elevated liver enzymes) representing different domains in gastrointestinal medicine was constructed. The examination consisted of 12 pencil-and-paper questions, four of which were standard, five-option multiple choice questions (MCQs) and eight of which were extended-matching questions, with a range of 10–16 possible choices. Three questions were created for each of the four clinical presentations. All of the questions asked for the most likely diagnosis, to be chosen from the clinical information in the stem.

Subjects

The examination was administered to 20 non-experts, final-year clinical clerks at the University of Calgary, and 20 gastroenterology experts (defined as qualified specialists who had been in practice for more than 5 years and were devoting more than 80% of their clinical time to gastroenterology).

Data collection

First, with no time restraint, the subjects were asked to answer the 12 questions. Two types of scores were generated:

Assessment of cognitive process After the completion of the 12 questions, the subjects, with the examination paper in hand and any written notes made during the examination, were asked to explain how they arrived at each diagnosis. A panel of two judges (experts in the gastro-enterologic presentations being tested and in the recognition of the diagnostic reasoning process) interviewed the examinees. Without prompting, and without revealing examination results, the examinees were asked to think aloud[22] and describe how each diagnosis was derived. Based on the examinees' verbal discourse for that question, the two judges determined the predominant diagnostic process used. Once the diagnostic process was assigned, the examinee was encouraged to proceed to the next question, until a diagnostic process had been assigned for all 12 questions.

It was determined that hypothetico-deductive reasoning was the diagnostic strategy utilized when prior to selecting the most likely diagnosis, the subjects analysed one by one each alternative diagnosis presented with the clinical vignettes. Determination that a scheme-inductive diagnostic reasoning strategy was used occurred by analysis of the verbal discourse using modified propositional analysis.[23] A proposition is defined as 'the smallest unit of meaning that underlies the surface structure of a text'. This analysis consisted of searching the examinees' discourse for key predetermined propositions that linked categories and thus provided evidence for chunking (i.e. scheme use). These key chunking propositions were determined by the authors based on information from texts, databases and consultation with experts not participating in the study. The propositions are shown in Table 1.

Determination that 'pattern recognition' was used occurred when the subject directly reached a single diagnosis with only perfunctory attention to the alternatives.

The interviews were audiotaped or videotaped for later review. Such reviews were required when the two judges identified different reasoning strategies. The most frequent cause for differences in identification of diagnostic reasoning strategy was examinees' use of more than one strategy in a given problem. Concurrence about the diagnostic reasoning strategy was reached in all subjects after tape review and discussion.

Assessment of diagnostic answers A dichotomous score was assigned to the questions: a mark of 1 for the correct diagnosis and a mark of 0 for an incorrect answer.

Data analyses

Reliability of the examination diagnostic scores was estimated using Cronbach's alpha coefficient. Item statistics were generated for each item including a discrimination index. Inter-rater reliability of diagnostic reasoning scores was estimated by a Pearson correlation coefficient.

Construct validity was determined by a significant difference between experts and novices in the use of different diagnostic strategies and ability to arrive at a correct diagnosis. It was expected that experts would utilize pattern recognition as a diagnostic strategy more frequently than hypothetico-deductive reasoning, with utilization frequency of scheme-inductive reasoning being located in between the other two strategies. A chi-squared test was used to determine if there were

Table 1 Propositions demonstrating evidence of chunking

Clinical presentation	Key chunking propositions
Dysphagia	Oropharyngeal versus oesophageal
	Mechanical versus motility
Elevated liver enzymes	Hepatocellular versus cholestatic
	Intra- versus extrahepatic cholestasis
Nausea and vomiting	Gastrointestinal versus non-GI causes
	Gastrointestinal versus metabolic versus CNS versus drugs
Diarrhoea	Small bowel versus large bowel
	Steatorrhoea (malabsorption) versus non-steatorrhoea
	Osmotic versus secretory versus inflammatory versus motility

significant differences in frequencies between experts and novices in use of diagnostic strategies. The chi-squared test was repeated for each clinical presentation (i.e. elevated liver enzymes, nausea and vomiting, diarrhoea and dysphagia). To reduce the possibilities of a type 1 error, the level of significance was set at 0·01 for the chi-squared test and 0·05 for all other statistical tests. In a similar fashion, it was expected that experts would be more successful than novices at arriving at the correct diagnosis in all four clinical presentations. A one-way MANOVA was used to determine if there were significant differences between experts and novices (independent variable) mean diagnostic success scores (dependent variable) on four clinical presentations. If a significant difference was found, an ANOVA was used to determine which clinical presentations exhibited significant differences between experts and novices.

Effects of expertise, diagnostic reasoning, and clinical presentations on diagnostic success A logistic regression analysis was used to determine which of the three independent variables being studied (diagnostic reasoning strategy, expertise and clinical presentation) had an impact on diagnostic success (the dependent variable). This analysis enables the modelling of the odds of making the correct diagnosis in terms of the independent variables. The regression was carried out using the generalized estimating equation approach. This approach enables a modelling of the association between responses from the same examinee (e.g. an exchangeable correlation that assumed a fixed but possibly non-zero correlation between responses from any given examinee). It is analogous to but more general than an analysis of variance of a split unit study. Analysis was carried out using the Stata software system (see www.stata.com or write Stata Corporation, 4905 Lakeway Drive, College Station TX 77845, USA).

Results

Reliability and validity of diagnostic reasoning scores

The two judges were able to agree by and large on the strategy used (hypothetico-deductive, scheme-inductive and pattern recognition). Initial diagnostic reasoning scores correlation between the two judges was 0·84.

Table 2 provides evidence for construct validity of the diagnostic reasoning classification. The chi-squared test indicated a difference in frequency of use of diagnostic strategies between experts and novices on the four clinical presentations (i.e. $P < 0{\cdot}01$). The total frequency of observed strategies was 60 as there were 20 experts and 20 novices and each responded to three clinical cases within each clinical presentation.

Reliability and validity of diagnostic success scores

The reliability coefficient for the diagnostic examination scores was 0·87. Discrimination indices of the item scores were above 0·5, indicating excellent discrimination. Construct validity of the items was demonstrated by the overall superiority of the experts over the non-experts across clinical problems, except with the clinical presentation 'nausea and vomiting' (appropriate statistical testing in Table 3). The mean diagnosis scores in Table 3 range from 0 to 3, as a score of 0 or 1 was assigned to each case (outlined in Methods), and there were three different cases for each clinical presentation.

Relationship of independent to dependent variables

The results of a logistic regression analysis are summarized in Table 4. Table 4 indicates the influence that the independent variables of diagnostic reasoning, expertise and clinical presentation had on the odds of diagnostic success (dependent variable) in the examination. The results are presented as the odds ratio (OR) of diagnostic success, relative to a chosen baseline level of each variable. Table 4 provides the independent variable, baseline level, OR of success, confidence limits and *P*-values.

Table 4 indicates that the odds of diagnostic success when examinees use pattern recognition are over 10 times (and over five times if scheme-inductive reasoning is utilized) relative to the odds of success with hypothetico-deductive reasoning. Table 4 also indicates there was no significant difference in odds of diagnostic success between scheme-inductive reasoning and pattern recognition.

Expertise is also correlated significantly with diagnostic success (the odds of success for experts are

Table 2 Frequency of diagnostic reasoning process for experts ($n = 20$) and novices ($n = 20$) on four clinical presentations and three cases/presentation

	Elevated liver enzymes	Nausea and vomiting	Diarrhoea	Dysphagia	Frequency
Experts					
Hypothetico-deductive	2	7	4	3	16
Scheme-inductive	29	11	26	40	106
Pattern recognition	29	42	30	17	118
Frequency	60	60	60	60	240
Novices					
Hypothetico-deductive	26	22	33	19	100
Scheme-inductive	26	4	17	36	83
Pattern recognition	8	34	10	5	57
Frequency	60	60	60	60	240
Chi-squared (d.f. = 2)	77·6	24·4	70·2	42·7	

$P < 0{\cdot}01$.

almost eight times the odds of success for the non-experts). In addition, there is a significant effect due to clinical presentations. Thus, diagnostic success is correlated in part with three independent variables: diagnostic reasoning, expertise and clinical presentation. In the logistic regression analysis, the possibility of two- or three-variable interaction was considered. There was no evidence of interaction between these variables.

Discussion

The present study was designed to investigate the relationship between diagnostic reasoning and diagnostic success. Irrespective of the potential impact of the other independent variables studied (expertise or clinical presentation), the reasoning strategy utilized for a particular medical problem was significantly related to the odds of making a correct diagnosis. The two other variables studied, expertise and clinical presentation, also were linked with the odds of diagnostic success. Not surprisingly, as shown in Table 3, experts achieved higher scores than novices. This is consistent with many previous studies.[21] The effect of the clinical presentations likely corresponds to a manifestation of content specificity.[2]

Table 3 Mean diagnosis scores of experts and novices on four clinical presentations (Wilks lambda = 0·29, $F_{4,35} = 21{\cdot}79$, $P = 0{\cdot}00$)

Clinical presentation	Expert	Novice	*P*-value
Elevated liver enzymes	2·85	1·75	0·00*
Nausea and vomiting	2·95	2·73	0·09
Diarrhoea	2·95	1·58	0·00*
Dysphagia	2·70	1·56	0·00*

*Indicates *P*-values below 0·05 level of significance.

Diagnostic reasoning process has been previously investigated, but in a fashion not comparable to the present study. Previous studies tended to compare the diagnostic performance of experts to that of novices or groups less expert.[21] In contrast, the present study compares problem-solving processes utilized by both novices and experts. The logistic regression analysis shown in Table 4 demonstrates that examinees using 'strong'[11] methods of problem solving such as pattern recognition or scheme-inductive reasoning had greater odds (approximately five- to 10-fold) of diagnostic success than examinees using hypothetico-deductive reasoning.

Pattern recognition was associated with the greatest likelihood of diagnostic success in the present study. The aptitude to exploit this strategy occurs with the availability of multiple examples in memory acquired through experience. When a new case is recognized as being similar to one seen before, pattern recognition leads to diagnosis. Its use by medical students in not usually advocated because their inadequate experience might lead to potentially grim consequences. Instead, students ought to be encouraged to see many examples of clinical problems and opportunity for such encounters should become an integral part of planning for medical school curricula.

Scheme-inductive problem-solving also emerged in the present study as a strategy associated with diagnostic success. Although others have bestowed diverse labels to the process, this diagnostic reasoning strategy has been consistently equated with enhanced diagnostic success.[18,19] In another recent study, difficult diagnostic problems were selected specifically to preclude the use of 'pattern recognition'. Experts, by using 'data

Table 4 Logistic regression of the odds of diagnostic success

Independent variable	Baseline level	OR (95% CI)	*P*-value
Diagnostic reasoning			
Scheme-inductive	Hypothetico-deductive	5·12 (2·65–9·91)	< 0·0001
Pattern recognition	Hypothetico-deductive	10·34 (4·35–24·58)	< 0·0001
Pattern recognition	Scheme-inductive	2·02 (0·84–4·84)	0·1148
Expertise			
Expert group	Non-expert group	7·69 (3·56–16·58)	< 0·0001
Clinical presentation			
Nausea and vomiting	Liver enzymes	5·41 (2·09–14·01)	0·0005
Diarrhoea	Liver enzymes	1·04 (0·50–2·15)	0·9108
Dysphagia	Liver enzymes	0·46 (0·22–0·93)	0·0295

chunking around physiologic principles' and 'judicious and comprehensive choice of alternative diagnoses', had a 91% diagnostic success rate compared to 25% for novices.[21] The diagnostic reasoning strategy described in this study is similar to scheme-inductive problem solving.

The 'small worlds' hypothesis of Kushniruk *et al.*[24] is also remarkably similar to scheme-inductive problem-solving. The hypothesis explains that expert physicians base the organization of their knowledge on similarities between disease categories, forming 'small worlds' consisting of small subsets of diseases and their distinguishing features. Such a knowledge structure is identical to a 'scheme'.[13] Experts then analyse the related diagnostic hypotheses by means of a succession of limited comparisons. They use efficient strategies for discriminating among these alternative hypotheses in a stepwise process. This process is virtually indistinguishable from scheme-inductive problem-solving.

The third factor affecting the odds of diagnostic success in the present study, the clinical presentations, likely represents an expression of content specificity. Although the presence of content specificity has been described as axiomatic, factors that explicate this ubiquitous finding remain uncertain.[1] Eva *et al.*[1] suggest that the low correlation of performance across problems cannot be explained solely by differences in knowledge. They state, 'Low correlation of performances across problems may reflect different strategy choices, not different aspects of the problems per se.' It is possible that these two ideas are not mutually exclusive. In our study, scheme utilization, reflective of advanced knowledge organization, correlated positively with diagnostic success. Advanced knowledge structures likely play a permissive role in the selection of diagnostic strategies.[7,8,25,26] In other words, 'reduced/dispersed knowledge' (small amounts of information/long lists of static diagnoses originating from rote memorization), permits hypothetico-deductive reasoning. 'Elaborated/compiled knowledge' (clinical findings, anatomic locations, pathophysiologic explanations and disease taxonomies/encapsulated, higher-order knowledge structures that link abridged intricate networks into a scheme of relationships and diagnoses) permits scheme-inductive problem-solving. 'Illness/instance scripts' make pattern recognition possible. The problem-solving strategy utilized is dependent on the knowledge structure available to the problem-solver; the knowledge structure available depends on the domain where the problem to be solved resides.

In their study, Elstein *et al.*[2] explained that expert clinicians used hypothetico-deductive reasoning because all clinical information was not presented initially, and the search for further information is most efficiently guided by working back from possible hypotheses to associated symptoms. In the present study, all the necessary information was made available in the clinical vignettes. A search for additional information was not needed. Consequently, hypothetico-deductive reasoning was selected by some of the subjects for reasons other than data collection. Perhaps it was the absence of an organized knowledge structure in memory that resulted in the selection of this strategy. Although the problem-solvers in Elstein's study were specialists in internal medicine, this specialty encompasses a huge medical knowledge domain. Even experts are ignorant in some domains (reduced knowledge), have limited knowledge in other areas (dispersed knowledge) and have structured knowledge (elaborated networks) as the medical domains come nearer to their personal area of special expertise (abridged networks and illness/instance scripts). Hypothetico-deductive reasoning represents a generic approach that can be utilized in the absence of organized knowledge structures, permitting diagnosis when medical problems reside outside the area of expertise of the problem solver.

There are some limitations to the present study. Eva and colleagues have described think-aloud protocols as methodologically flawed.[20] They suggest that think-aloud protocols are not reliable for identifying forward versus backward reasoning. One study analysed in support of this conclusion revealed that the apparent direction of reasoning (forward or backward) depended on when the participants were prompted to think aloud (online or post-hoc). Doubt was cast on whether the propositions generated (conditional for forward and causal for backward) truly indicate differences in forward and backward reasoning or simply the capacity to cause post-hoc think aloud responses to appear like forward reasoning. Thus, the likelihood exists that diagnostic success affected the way in which participants discussed their reasoning process.

The possibility that diagnostic success affected the manner in which the subjects in the present study discussed their reasoning process is not at all probable. None of them knew whether a correct diagnosis was reached. As for the think-aloud protocol used in this study, it differed in several important ways from the one described by Eva *et al.*[1] Propositional analysis used in the present study was not directed at the number of hypotheses or propositions, or at conditional versus causal links. Whether the reasoning strategy was forward or backward was not under consideration. Mental frameworks or 'schemes' were the focus of the present study. The analysis consisted of searching the examinees' discourse for key predetermined propositions that linked categories and thus provided evidence for chunking. Hypothetico-deductive reasoning and pattern recognition were identified in a manner completely different from the one utilized by Eva and colleagues (see Methods). The difference in the protocol used in the present study compared to the one used by Eva is substantial and therefore it cannot be assumed that the same methodological flaw exists. In addition, parallel investigations of scheme-inductive reasoning employing a completely different method for ascertaining how information is organized (an indirect method termed 'concept sorting')[19] provided results very similar to the ones described in the present study and supportive of the notion that scheme-inductive reasoning is associated with success.

A second limitation is the manner in which the diagnostic problem-solving process selected by the subjects was ascertained. Concurrence and not initial judgements were used (see Methods). 'High fidelity' simulations[2] that can test problem-solvers for key diagnostic tasks were not utilized. Despite such limitations, the conclusion that diagnostic problem-solving strategy, expertise and clinical presentations are significant factors associated with diagnostic success is well supported by the evidence presented. Additional studies are needed to compare the effect of problem-solving strategies in other domains of medicine and with groups having intermediate levels of expertise.

Acknowledgements

This research was supported by a grant from the Medical Council of Canada. The authors wish to acknowledge and thank Glenn Regehr for reviewing the manuscript and providing insightful and helpful suggestions that improved the quality and clarity of this paper.

References

1 Eva KW, Neville AJ, Norman GR. Exploring the etiology of content specificity. factors influencing analogic transfer and problem solving. *Acad Med* 1998;**73**:S1–5.

2 Elstein AS, Shulman LS, Sprafka SA. *Medical Problem-Solving: an Analysis of Clinical Reasoning*. Cambridge, MA: Harvard University Press 1978.

3 Neufeld VR, Norman GR, Barrows HS, Feigner JW. Clinical problem solving by medical students. a longitudinal and cross-sectional study. *Med Educ* 1981;**15**:315–22.

4 Newell A. *Artificial Intelligence and the Concept of Mind. Computer Models of Language and Thought*. Oxford: Freeman 1973, 52.

5 Groen G, Patel V. Medical problem-solving: some questionable assumptions. *Med Educ* 1985;**19**:95–100.

6 Patel V, Groen G, Norman GR. Effects of conventional and problem-based medical curricula on problem solving. *Acad Med* 1991;**66**:380–9.

7 Gilhooly KJ. Cognitive psychology and medical diagnosis. *Appl Cognitive Psychol* 1990;**4**:261–72.

8 Schmidt HG, Norman GR, Boshuizen HPA. A cognitive perspective on medical expertise: theory and implications. *Acad Med* 1990;**65**:611–21.

9 Papa FJ, Stone RC, Aldrich DG. Further evidence of the relationship between case typicality and diagnostic performance: implications for medical education. *Acad Med* 1996;**71**:S10–2.

10 Bordage G, Lemieux M. Semantic structures and diagnostic thinking of experts and novices. In Proceedings of the 30th Annual Conference on Research in Medical Education. *Acad Med* 1991;**66**:S70–2.

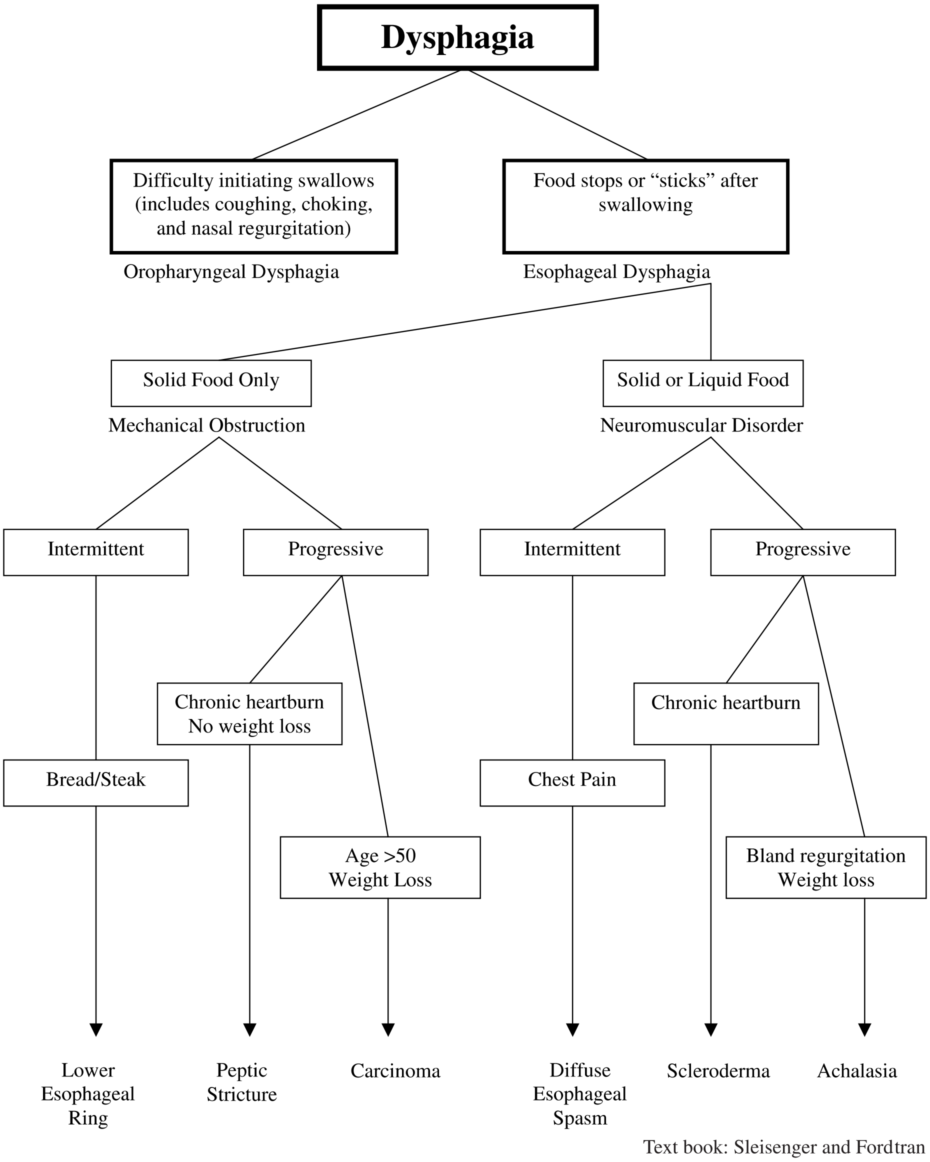
11 Grant J, Marsden P. The structure of memorized knowledge in students and clinicians: an explanation for diagnostic expertise. *Med Educ* 1987;**21**:92–7.

12 Mandin H, Harasym P, Eagle C, Watanabe M. Developing a 'clinical presentation' curriculum at the University of Calgary. *Acad Med* 1995;**70**:186–93.

13 Mandin H, Jones A, Woloschuk W, Harasym P. Helping students learn to think like experts when solving clinical problems. *Acad Med* 1997;**72**:173–9.
14 Regehr G, Norman GR. Issues in cognitive psychology: implications for professional education. *Acad Med* 1996;**71**:988–1001.
15 Glaser R. The role of knowledge. *Am Psychologist* 1984;**30**:93–104.
16 West C, Farmer J, Wolff P. *Instructional Design: Implications for Cognitive Science.* Englewood Cliffs. NJ: Prentice Hall 1991, 1–35.
17 Platt JR. Strong inference. *Science* 1964;**146**:347–53.
18 Beck AL, Bergman DA. Using structured medical information to improve students' problem-solving performance. *J Med Educ* 1986;**61**:749–56.
19 McLaughlin K, Mandin H. Using 'concept sorting' to study learning processes and outcomes. *Acad Med* 2002;**77**:831–6.
20 Eva KW, Brooks LR, Norman GR. *Forward Reasoning as a Hallmark of Expertise in Medicine: Logical, Psychological, and Phenomenological Inconsistencies*. In: Shohov, SP, eds. *Advances in Psychological Research*, **Vol. 8**. New York: Nova Scotia 2002; p. 42.
21 Norman GR, Trott A, Brooks L, Kinsey-Smith E. Cognitive differences in clinical reasoning related to postgraduate training. *Teaching Learning Med* 1994;**6**:114–20.
22 Davison G, Vogel R, Coffman S. Think-aloud approaches to cognitive assessment and the articulated thoughts in simulated situations paradigm. *J Consulting Clin Psychol* 1997;**65**:950–8.
23 Frederiksen CH. Representing logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychol* 1975;**7**:371–458.
24 Kushniruk A, Patel V, Marley A. Small worlds and medical expertise: implications for medical cognition and knowledge engineering. *Int J Med Informatics* 1998;**49**:255–71.
25 Bordage G. Elaborated knowledge: a key to successful diagnostic thinking. *Acad Med* 1994;**69**:883–5.
26 Mandin H. Evaluation: the engine that drives us forwards or back. *J Clin Invest Med* 2000;**23**:70–6.

Received 24 *July* 2002; *editorial comments to authors* 23 *September* 2002; *accepted for publication* 2 *December* 2002

Appendix 1


Dysphagia
Difficulty initiating swallows (includes coughing, choking, and nasal regurgitation)
Food stops or "sticks" after swallowing
Oropharyngeal Dysphagia
Esophageal Dysphagia
Solid Food Only
Solid or Liquid Food
Mechanical Obstruction
Neuromuscular Disorder
Intermittent
Progressive
Intermittent
Progressive
Chronic heartburn
No weight loss
Chronic heartburn
Bread/Steak
Chest Pain
Age >50
Weight Loss
Bland regurgitation
Weight loss
Lower Esophageal Ring
Peptic Stricture
Carcinoma
Diffuse Esophageal Spasm
Scleroderma
Achalasia
Text book: Sleisenger and Fordtran