DOCUMENT RESUME

ED 367 710 TM 021 241

AUTHOR van de Wiel, Margaretha W. J.; And Others
TITLE The Influence of Task Instruction and Lab Data Format on Clinical Case Recall.
PUB DATE Apr 94
NOTE 14p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Clinical Diagnosis; Cognitive Processes; Data Analysis; Foreign Countries; Higher Education; Internal Medicine; Laboratories; Medical Case Histories; *Medical Students; Motivation; *Physicians; *Recall (Psyc'.ology); *Testing
IDENTIFIERS Accuracy; Data Organization; *Expertise; Knowledge Utilization; Netherlands; *Task Oriented Instructions

ABSTRACT

In the present study, task instruction and lab data format were manipulated to explain the discrepancy between the positive linear recall function with expertise (reported by van de Wiel and others, 1993), and the generally found intermediate effect in clinical case recall. Sixteen second-year medical students, 16 fourth-year students, and 16 internists studied, diagnosed, and recalled four clinical cases. No differences were found between intentional and incidental recall instructions or between cases with numerical and interpreted lab data. Diagnostic accuracy increased with the level of expertise. The overall recall data showed again the intermediate effect. Reanalysis of the 1993 data suggests that the linear recall function was caused by the experts' motivation. Four figures, two tables, and one appendix describing a case of heart failure with two kinds of lab data are included. (Contains 10 references.) (Author/SLD)

The influence of task instruction and lab data format on clinical case recall

Margaretha W.J. van de Wiel

Henny P.A. Boshuizen

Henk G. Schmidt
Department of Educational Development and Research
University of Limburg
Maastricht, The Netherlands

University of Limburg
PO Box 616
6200 MD Maastricht
The Netherlands
E-mail: M.vdWiel@Educ.RULimburg.NL

Paper presented at the Annual Meeting of the American Educational Research Association in New Orleans, Louisiana, April 4-8, 1994

Abstract

In the present study, task instruction and lab data format were manipulated to explain the discrepancy between the positive linear recall function with expertise (Van de Wiel et al., 1993), and the generally found intermediate effect in clinical case recall. Subjects of three levels of expertise studied, diagnosed and recalled four clinical cases. No differences were found between intentional and incidental recall instructions, and between cases with numerical and interpreted lab data. Diagnostic accuracy increased with level of expertise. The overall recall data showed again the intermediate effect. Reanalysis of the 1993 data suggests that the linear recall function was caused by the experts' motivation.

In research on expertise in the domain of medicine, free recall of clinical cases is used to probe the case representations of subjects of different levels of expertise. It has repeatedly been demonstrated that subjects of an intermediate level of expertise remembered more case information than medical experts (e.g. Patel & Groen, 1991; Patel & Medley-Mark, 1986; Schmidt & Boshuizen, 1993; Schmidt, Boshuizen, & Hobus, 1988). The inverted U-shaped relationship between recall and level of expertise is known as the intermediate effect, and has been explained by knowledge encapsulation in expert physicians (Schmidt & Boshuizen, 1992; Schmidt & Boshuizen, 1993). Encapsulation is a form of knowlegde restructuring in which clusters of detailed concepts become encapsulated into a few higher level concepts or diagnostic labels, based on abundant practice in a certain domain. In contrast to students who have to reason through their knowledge bases in order to build a coherent case representation, experienced physicians automatically activate encapsulating concepts in diagnosing a clinical case, and hence their case recall is less extensive than students'.

Recently we tried to replicate the intermediate effect in clinical case recall, but surprisingly we found a linear recall function with level of expertise (Van de Wiel, Boshuizen, Schmidt & de Leeuw, 1993). A possible explanation for this conflicting finding concerned the presence of approximately 25% of numerical lab data in the four cases of the replication experiment in contrast to virtually no lab data in the endocarditis case of the original experiment (Schmidt & Boshuizen, 1993). In recall studies of numerical lab data Norman, Brooks and Allen (1989) found that medical specialists remembered more lab data than second- and third-year students. Especially under strict diagnose instructions this effect was strong. Therefore, it was concluded that the interpretation of numerical lab data for diagnostic practice requires an effortful analysis even for expert physicians. If subjects in the replication experiment had to process the values of the lab data and the complex relations among them in an analytical fashion, they would have formed elaborate case representations and this could explain their high recall performance. On the other hand, if we would present lab data in an already interpreted format[1], we would expect that these data can be processed immediately and automatically, resulting in less elaborate case representations and recall. This distinction between analytical processing of numerical lab data versus automatic processing of interpreted lab data was investigated in the present study.

Another issue addressed to explain the failure to reproduce the intermediate effect is the task perception of subjects. In typical clinical case representation

[1] For example Sodium 140 mmol/l and Potassium 4.3 mmol/l can be interpreted as respectively Sodium normal and Potassium normal. In addition both could be interpreted as electrolytes normal. Another example of an interpretation of more numerical lab data is respiratory alkalosis.

studies subjects are instructed to study a case in order to formulate a diagnosis. Some researchers (Hobus, Schmidt, Boshuizen & Patel, 1987; Norman et al., 1989) did not tell their subjects that they would be asked for recall after diagnosing the case, which is referred to as incidental recall instructions. Other researchers (e.g. Patel & Medley-Mark, 1986; Schmidt & Boshuizen, 1993) did tell their subjects that they would be asked to recall the case after the diagnostic task, and this will be referred to as intentional recall instructions. The assumption with this type of intentional recall is that the recall task is considered as secondary to the diagnostic task, and therefore probes the genuine problem representations of subjects diagnosing a case without interference of memorization behavior. In fact, it is expected that subjects behave in the same way as if they were not aware to be asked for recall. This assumption does certainly not hold for rather extreme intentional recall instructions, in which subjects were required to study case information with the intention to recall this information as accurate as possible (e.g. Coughlin & Patel, 1987; Norman et al., 1989). Although a diagnosis was sometimes requested after recall, these studies have little in common with diagnostic processing in medical practice and, therefore, are better called case memorization studies. Norman and colleagues (1989) compared a memorization task with an incidental recall task in their studies on numerical lab data revealing a significant interaction effect between expertise level and recall instructions: Experts recalled less information under memorize instructions than under incidental instructions, whereas third-year medical students recalled more information under memorize instructions. This experiment, thus, clearly demonstrated the importance of explicit recall instructions for meaningfully interpreting clinical case recall data. However, the difference between a diagnostic task with intentional or incidental recall instructions has not been tested yet. And since experts in the Van de Wiel et al. study (1993) recalled more when more study time was available, it is not inconceivable that they used that time to enhance their recall after knowing the diagnosis of a case. Therefore we questioned in the present experiment whether our assumption was correct that in a primarily diagnostic task intentional recall may be regarderd as similar to icidental recall.

A third explanation for the discrepancy in recall data between the study of Schmidt and Boshuizen (1993) and the study of Van de Wiel et al. (1993) could be an effect of the different cases used in the two experiments. The intermediate effect has been repeatedly demonstrated with the case of bacterial endocarditis (e.g. Patel & Medley-Mark, 1986; Schmidt & Boshuizen, 1993), whereas four other cases from internal medicine were used in the study of Van de Wiel et al.. In order to investigate a possible case effect we included the endocarditis case and three cases from the Van de Wiel et al. study in our present experiment.

To summarize, in the current case representation study we first investigated the influence of intentional versus incidental recall instructions on recall performance in a diagnostic task; secondly we examined if the recall of two cases containing numerical lab data was higher than the recall of the same cases with the lab data in interpreted format; finally the endocarditis case used in the study of Schmidt and Boshuizen (1993) was presented in order to verify a possible case effect. The study time available was set to three minutes, since the intermediate effect most clearly occurred in this long reading time condition (Schmidt & Boshuizen).

Method

Subjects. In total 48 subjects participated in this experiment. These were 16 second-year and 16 fourth-year medical students of the University of Limburg. The experts were 16 internists from five different hospitals in Limburg with at least four years of experience and an average of 15 years of experience in internal medicine. Subjects received a small compensation for their participation.

Material. The materials consisted of four clinical case descriptions and two blank response sheets after each case. Each clinical case description reported some contextual information, the complaint, findings from history taking and physical examination, the relevant laboratory data and some additional findings. Three cases, stomach carcinoma, pheohromocytoma and heart failure, were earlier used by Van de Wiel et al. (1993) and these cases contained numerical lab data. In addition, the discriptions of pheochromocytoma and heart failure were rewritten into descriptions containing lab data in an interpreted form. A fourth case description was the endocarditis case used by Patel and Groen (1986) and Schmidt and Boshuizen (1993). The case descriptions were about half a page in length and consisted of 42, 33, 43, 32, 40 and 50 propositions respectively. In appendix A case descriptions of heart failure with numerical lab data and interpreted lab data are provided.

Procedure. Subjects were instructed to study a case for maximal 3 minutes in order to formulate a diagnosis. Before the presentation of the first clinical case (stomach carcinoma) only half of the subjects was told that they would subsequently be asked to write down whatever they remembered from the case (intentional condition); the other half was not aware that they would be asked to recall the case (incidental condition). Instructions for the next three cases were for all subjects as in the intentional condition. The subjects of each level of expertise were divided into two groups, each consisting of 4 subjects from the intentional conditon and 4 subjects from the incidental. One group was subsequently presented the pheochromocytoma case with numerical lab data, the endocarditis case, and the heart failure case with interpreted lab data. The other group was subsequently presented the pheochromocytoma case with interpreted lab data, the endocarditis case, and the heart failure case with numerical lab data. An overview of the experimental design is provided in table 1. Subjects were free to use as much time as they needed to write down their diagnosis and recall.

Table 1 Experimental design

Case	Subjects*	Task instructions	Lab data format
1. Stomach carcinoma	1 - 8	Intentional	Numerical
	9 - 16	Incidental	Numerical
2. Pheochromacytoma	1 - 4 en 9 - 12	Intentional	Numerical
	5 - 8 en 13 - 16	Intentional	Interpreted
3. Bacterial endocarditis	1 - 16	Intentional	Interpreted
4. Heart failure	1 - 4 en 9 - 12	Intentional	Interpreted
	5 - 8 en 13 - 16	Intentional	Numerical

* Assignment of subjects to experimental conditions for each expertise group.

Analysis. Diagnoses were scored on a scale ranging from 0 (completely inaccurate diagnosis) to 6 (completely accurate diagnosis) for each case: points were attributed to accurate diagnostic elements which, for each case, summed up to 6. Based on a technique of propositional analysis for medical protocols (Patel & Groen, 1986) recall protocols were segmented into small meaningful information units referred to as propostions. For each proposition in the free recall, it was decided whether it matched any propositions in the original case. The number of correctly recalled propostions was counted. Reliabilities of these procedures exceeded .95. Anova and repeated measures Manova were used to analyze the data.

Results and discussion

Diagnosis. None of the manipulated variables had an effect on diagnostic accuracy, therefore we plotted diagnostic accuracy on each case presented against level of expertise (figure 1). Repeated measures Manova revealed a significant main effect for both level of expertise (F(2, 45) = 64,37, p = .0001), and casuistry (F(3, 135) = 6.79, p = .0003). The diagnostic task differentiated between levels of expertise as might be expected: subjects with more experience performed better.

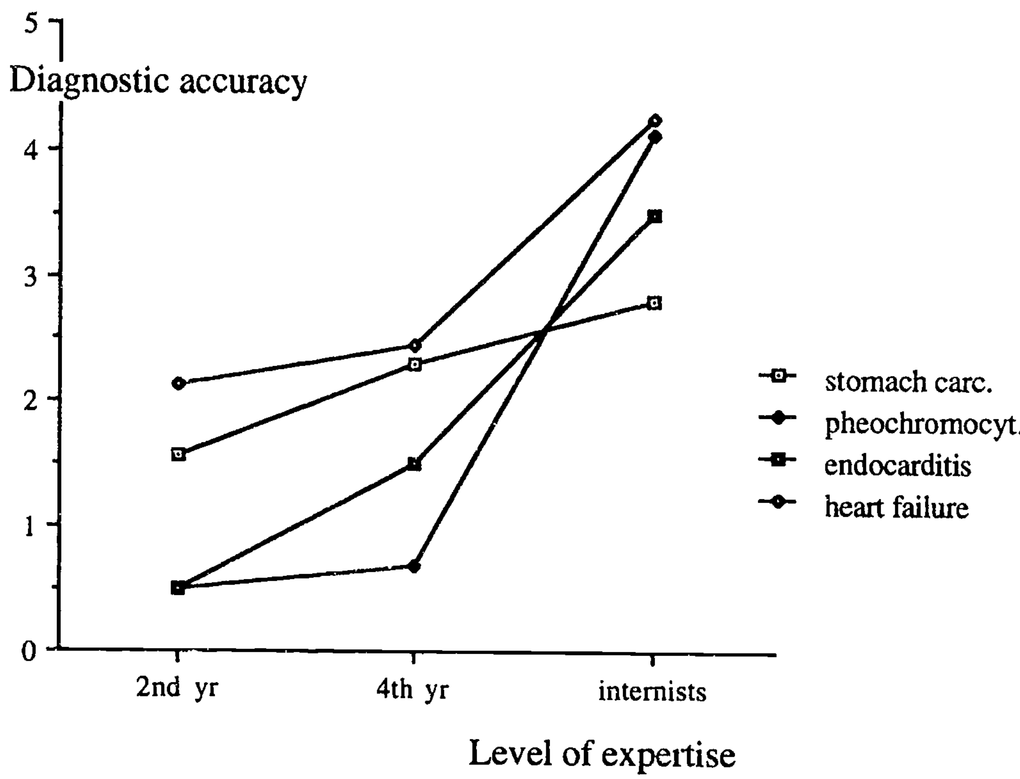

Figure 1. Diagnostic accuracy as a function of expertise level and casuistry.

Incidental versus intentional recall task. Overall Anova revealed a significant effect of expertise level on percentage of propositions recalled (F(2, 42) = 16,30, p = .0001). Figure 2 shows that the relationship between recall and level of expertise for both intentional, and incidental recall instructions is an inverted U-shaped function. However, no significant effect of task instruction was found (F(1, 42) = 2.66, p = .11). Although in figure 2 a trend may be observed that recall performance is higher under intentional recall instructions, this improvement was only 7% for the internists, whereas the difference between internists' recall performance in the study of Schmidt and Boshuizen (1993) and in the study of Van de Wiel et al. (1993) was 33% (see table 2). Therefore, we conclude that if the task of subjects is presented as a primarily diagnostic task, the intentional recall instruction does not significantly enhance recall performance. The assumption seems justified, then, that recall prompted by incidental and intentional task instructions both reflect case representation as a result of diagnostic processing.

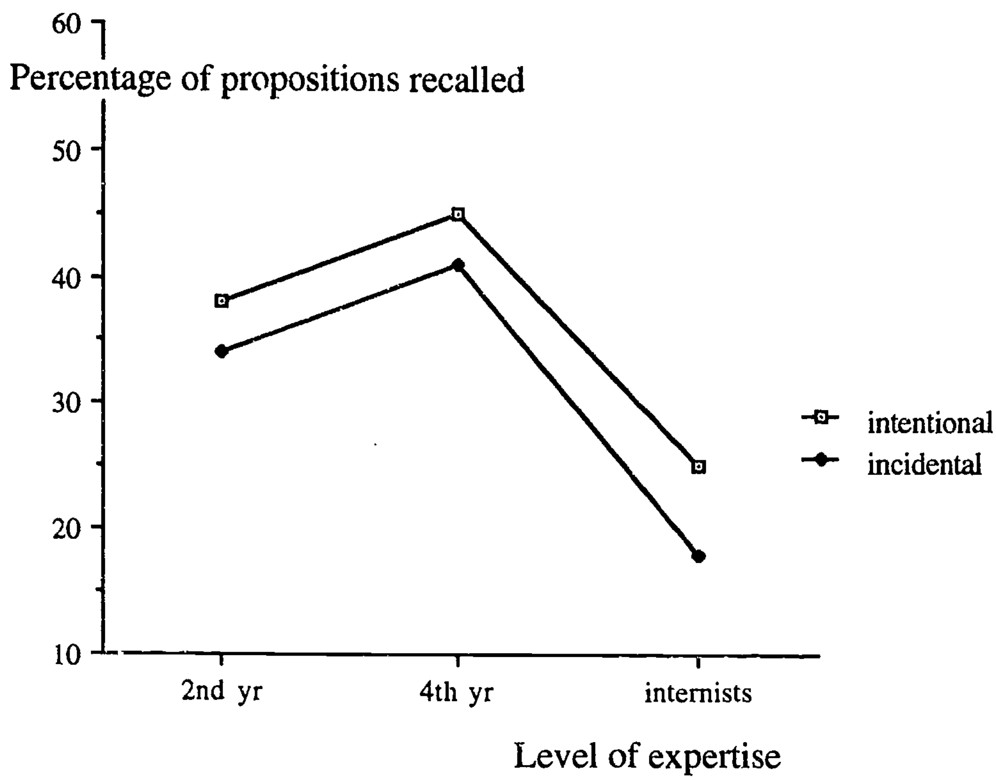

Figure 2. Percentage of propositions recalled as a function of expertise level and task instruction for the stomach carcinoma case.

Recall of cases with numerical versus interpreted labdata. The relation between percentage of propositions recalled, level of expertise and labdata format is depicted in figure 3. The main effect of labdata format on recall performance was not significant for both the pheochromocytoma case ($F(1, 42) = 2.29$, $p = .14$) and the heart failure case ($F(1, 42) = .001$, $p = .97$). This indicates that the processing of numerical lab data in these cases does not require a more effortful analysis than the processing of interpreted labdata. Further, figure 3 schows that internists remember less from the pheochromocytoma case than students, although the main effect of level of expertise on recall was not statistically significant ($F(2, 42) = 2.78$, $p = .074$). In addition, we did not find a significant main effect of level of expertise on recall in the heart failure case ($F(2, 42) = 1.67$, $p = .20$).

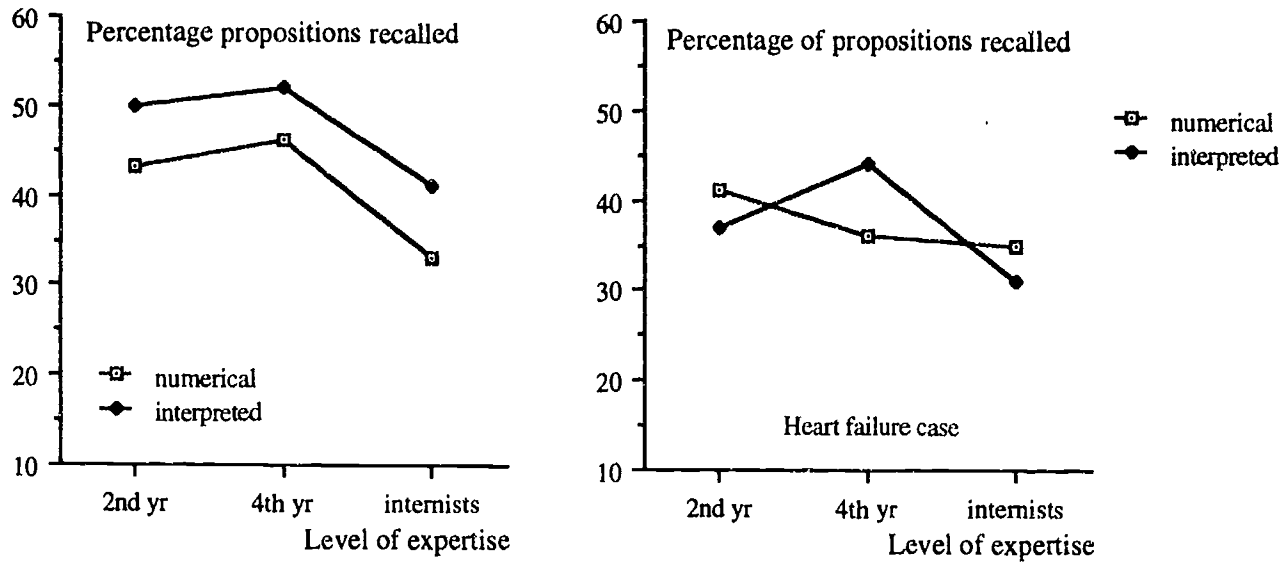

Figure 3. Percentage of propositions recalled as a function of expertise level and lab data format for the pheochromocytoma case and the heart failure case.

Recall of endocarditis case. Analysis of the endocarditis case revealed a significant effect of level of expertise on number of propositions recalled ($F(2, 45) = 9.56$ $p = .0003$)). Figure 4 shows that the expert physicians remembered less from the endocarditis case than subjects of an intermediate level of expertise. Pairwise comparisons between the three expertise groups by the Student-Newman-Keuls test (significance level of .05) showed that the internists recalled significantly less than the second- and fourth-year students. This result is in line with the intermediate effect found by Schmidt and Boshuizen (1993) for the endocarditis case.

To summarize the recall data, we found neither an effect of the different task instructions, nor of the manipulations of lab data. This allowed us to perform a repeated measures analysis over the recall data of all four cases, and to depict the average percentage of propositions recalled for each case as a function of level of expertise (figure 4). Manova revealed significant effects of level of expertise on recall ($F(2, 45) = 7.93$, $p = .0011$), of casuistry on recall ($F(3,135) = 16.86$, $p = .0001$), and a significant interaction effect ($F(6,135) = 2.85$, $p = .012$). Pairwise comparisons between the three expertise groups by the Student-Newman-Keuls test (significance level of .05) showed that medical experts recalled significantly less than students, confirming the intermediate effect in clinical case recall. Thus, these findings consistenly replicate the intermediate effect reported by Schmidt and Boshuizen (1993) under different task instructions and with different material.

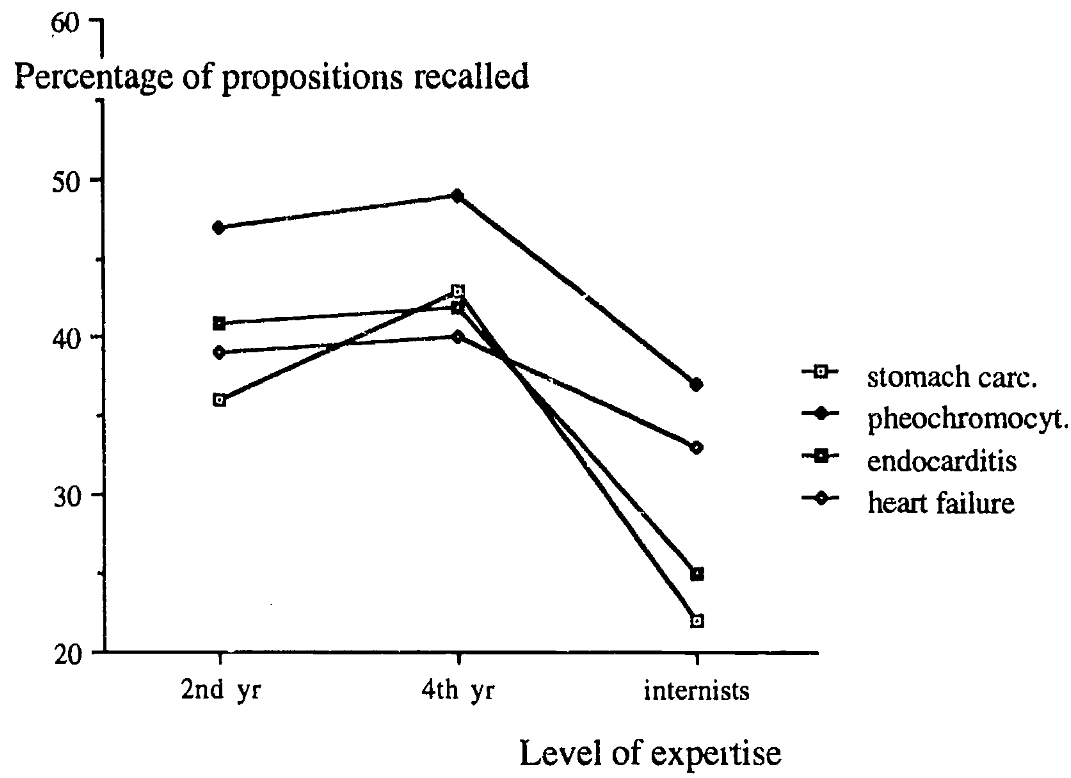

Figure 4. Average percentage of propositions recalled as a function of expertise level and casuistry.

The problem is, however, that we conducted the present experiment in order to explain the discrepancy between the linear recall function in the study of Van de Wiel et al. (1993) and the inverted U-shaped recall function in the study of Schmidt and Boshuizen (1993). Evidently, none of the suggested explanations has been approved, and the intermediate effect has been demonstrated once more. The only possible explanation which remains is a different motivation of the experts in the study of Van de Wiel and colleagues. The hypothesis that the experts in that study were highly motivated to perform as good as possible on task requirements was suggested by comparing the data on reading times, diagnostic accuracy and recall performance between the Schmidt and Boshuizen study, the Van de Wiel et al. study and the present study (see table 2). Experts in the Van de Wiel et al. study not only had a significantly higher recall performance than the experts in the two other studies, but also a better diagnostic accuracy and longer actual reading times. Together with the fact that the internists in the Van de Wiel et al. study recalled more case information under longer processing time conditions, this strongly suggests that the internists in that study used the extra processing time in order to enhance their recall. Thus, once the subjects fulfilled the primary task of formulating a diagnosis, cognitive capacity could be devoted to be able to recall as much as possible from the case. To answer the question why the experts in the Van de Wiel et al. study were more motivated to score on experimental tasks we can only speculate. One possible explanation could be that all internists in that study were working at the same department of internal medicine of the academic hospital in Maastricht and were asked to participate in our experiment by one of the professors in a staff meeting. The subjects in the other two studies, however, were directly approached by the experimenters. In addition, the internists in the present study were working at five different hospitals in the environment of Maastricht.

Table 2. Average actual reading times, diagnostic accuracy and percentage of propositions recalled for fourth-year students and internists under long reading time conditions in three clinical case recall studies.

	Subjects	Schmidt & Boshuizen (1993)	Van de Wiel, et al. (1993)	Present study
Actual reading time*	4th-year stud	?	175 (14)**	166 (33)
	internists	?	152 (33)	91 (34)
Diagnostic accuracy	4th-year stud	1.0 (.8)	2.5 (2.1)	2.5 (2.0)
	internists	4.3 (1.3)	5.3 (1.4)	4.1 (1.7)
Perc. of propositions recalled	4th-year stud	56 (13)	55 (12)	43 (14)
	internists	31 (14)	64 (18)	31 (16)

* Reading time in seconds. ** Standard deviations are provided between brackets.

References

Coughlin, L. D., & Patel, V. J. (1987). Processing of critical information by doctors and medical students. *Medical Education, 62,* 818-828.

Hobus, P. P. M., Schmidt, H. G., Boshuizen, H. P. A., & Patel, V. L. (1987). contextual factors in the activation of first hypotheses: Expert-novice defferences. *Medical Education, 21,* 471-476.

Norman, G. R., Brooks, L. R., & Allen, S. W. (1989). Recall by expert medical practitioners and novices as a record of processing attention. *Journal of Experimental Psychology: Learning, Memory, and Cogntion, 15(6),* 1166-1174.

Patel, V. L., & Groen, G. J. (1991). The general and specific nature of medical expertise: A critical look. In A. Ericsson, & J. Smith (Ed.), *Toward a general theory of expertise: Prospects and limits.* (pp. 93-125). Cambridge: Cambridge University Press.

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

Patel, V. M., & Medley-Mark, V. (1986). Relationship between representation of textual information and underlying problem representation in medicine. *Cognitive research series.* Montreal: McGill University. (CME86-CS1)

Schmidt, H. G., & Boshuizen, H. P. A. (1992). Encapsulation of biomedical knowledge. In A. E. Evans, & V. L. Patel (Ed.), *Advanced models of cognition for medical training and practice.* . New York, NY: Springer Verlag.

Schmidt, H. G., & Boshuizen, H. P. A. (1993). On the origin of intermediate effects in clinical case recall. *Memory and Cognition, 21(3),* 338-351.

Schmidt, H. G., Boshuizen, H. P. A., & Hobus, P. P. M. (1988). Transitory stages in the development of medical expertise: the "intermediate effect" in clinical case representation studies. In: *Proceedings of the Tenth Annual Conference of the Cognitive Science Society Meeting.* (pp. 139-145). Hillsdale, NJ: Lawrence Erlbaum.

Van de Wiel, M. W. J., Boshuizen, H. P. A., Schmidt, H. G., & de Leeuw, P. W. (1993). *A failure to reproduce the intermediate effect in clinical case recall.* Paper presented at the Annual meeting of the American Educational Research Association. Atlanta, Georgia, April 12 - 16.

Appendix A

Case of heart failure with numerical lab data

A 70-year-old female is admitted into hospital because of increasing shortness of breath. History taking reveals that the patient has been very tired lately and tolerates her food badly. Sometimes she has chest pain, especially after dinner.

Physical examination shows a pale and tired woman. She has an irregular, unequal pulse of 100/min. The blood pressure is 110/70 mmHg and jugular venous pressure is elevated. The patient has wide-spread peripheral edema, and positive jugular venous pulsations. The heart is enlarged to all sides, and auscultation reveals a holosystolic murmur at the apex radiating towards the axilla. Lungs: at both sides rales at lung bases. Liver and spleen not palpable.

Laboratory results shows a ESR of 2 mm/u (normal: < 12 mm/u), a Hemoglobine-level of 10.8 mmol/l (normal: 7.5-10.0 mmol/l) and a PCV of 0.54 (normal: 0.36-0.47). Electrolyts normal. Creatinine 85 µmol/l (normal: 53-97 µmol/l), CPK 40 U/l (normal: 40-200 U/l). pH is 7.50 (normal: 7.35-7.45), PO_2 11,6 kPa (normal: 8.7-13.1 kPa), pCO_2 3.6 kPa (normal: 4.5-5.9 kPa), HCO_3^--concentration 21 mmol/l (normal: 22-28mmol/l) and O_2-saturation 97% (normal: 93-98%).

The thoracic X-ray shows congestion of the lungs and an enlarged heart. Echocardiography shows an enlarged left atrium and ventricle. And ECG reveals atrial fibrillation.

Case of heart failure with interpreted lab data

A 70-year-old female is admitted into hospital because of increasing shortness of breath. History taking reveals that the patient has been very tired lately and tolerates her food badly. Sometimes she has chest pain, especially after dinner.

Physical examination shows a pale and tired woman. She has an irregular, unequal pulse of 100/min. The blood pressure is 110/70 mmHg and jugular venous pressure is elevated. The patient has wide-spread peripheral edema, and positive jugular venous pulsations. The heart is enlarged to all sides, and auscultation reveals a holosystolic murmur at the apex radiating towards the axilla. Lungs: at both sides rales at lung bases. Liver and spleen not palpable.

Laboratory results shows a low normal erythrocyte sinking rate (ESR), a slightly increased Hemoglobine-level and a slightly increased packet cell volume (PCV). Electrolytes normal. Kidney function normal. CPK low normal. Analysis of blood gasses shows metabolic compensation for respiratory alkalosis.

The thoracic X-ray shows congestion of the lungs and an enlarged heart. Echocardiography shows an enlarged left atrium and ventricle. And ECG reveals atrial fibrillation.