AD-A055 586 IBM THOMAS J WATSON RESEARCH CENTER YORKTOWN HEIGHTS N Y F/G 5/10
PREDICTING INSUFFICIENT LEARNING OF A COMPLEX PROCEDURE.(U)
JUL 77 D LYON, J C THOMAS N00014-72-C-0149
UNCLASSIFIED RC-6627 NL

| OF |
ADA
055586

END
DATE
FILMED
8 -78
DDC

RC 6627 (#28546) 7/11/77
Computer Science 19 pages

Research Report

PREDICTING INSUFFICIENT LEARNING OF A COMPLEX PROCEDURE

Don Lyon and John C. Thomas

IBM Thomas J. Watson Research Laboratory
P. O. Box 218
Yorktown Heights, New York 10598

11 Jul 77

22p.

RC-6627

N00014-72-C-0149

IBM Research Division
San Jose · Yorktown · Zurich

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) PREDICTING INSUFFICIENT LEARNING OF A COMPLEX PROCEDURE*		5. TYPE OF REPORT & PERIOD COVERED Interim Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Don Lyon** John Thomas		8. CONTRACT OR GRANT NUMBER(s) N00014-72-C-0149
9. PERFORMING ORGANIZATION NAME AND ADDRESS		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-197-020
11. CONTROLLING OFFICE NAME AND ADDRESS International Business Machines T.J. Watson Research Center P.O.Box 218 Yorktown Heights, NY 10598		12. REPORT DATE
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Office of Naval Research Code 455 Arlington, Virginia		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Human Learning, Procedural Learning, Predicting Learning, Learning Time, Business Systems, CAI, Comprehension Measures

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Thorough learning of complex procedures is required in the design, production, and purchase of software. This study examines the utility of three sources of information for predicting errors in such learning. Twenty-eight subjects were asked to learn thoroughly a complex order handling and invoicing system. The subjects were then tested by having them fill out the invoice associated with a present order. During this simulation they were allowed reference

20. to necessary data files but were not allowed reference to any procedural descriptions. The subject's prior knowledge and the subject's own estimate of the thoroughness of his or her learning were both predictive of overall performance, while study time was not predictive.

RC 6627 (#28546) 7/11/77
Computer Science 19 pages

PREDICTING INSUFFICIENT LEARNING OF A COMPLEX PROCEDURE

Don Lyon and John C. Thomas

IBM Thomas J. Watson Research Laboratory
P. O. Box 218
Yorktown Heights, New York 10598

This research was supported in part by the Engineering Psychology Programs, Office of Naval Research.

Predicting Insufficient Learning of a Complex Procedure

This paper reports an exploratory look at the utility of three sources of information which could be used to predict the errors an individual will make in learning a complex procedure. Both those who design and those who buy software are required to learn the details of complex procedures; errors in such learning can result in a program which does the wrong job. To the extent that such mistakes can be predicted, techniques for reducing their frequency can be devised.

We are interested in predicting errors under conditions similar to those which a businesperson selecting an application program might encounter. The procedure to be learned is an invoice processing system taken from an application program written in Business Definitions Language (BDL) (Hammer, Howe, Kruskal, and Wladowsky, 1975). We wish to ensure that each learner is motivated to learn the system well and free to study as long as necessary. The errors to be predicted are those which subjects exhibit when asked to fill out a sample invoice immediately after learning the system. Such errors may be the result of either initial miscomprehension or quick forgetting; in this initial study, we do not try to distinguish between these error sources. Our purpose is just to determine how useful certain kinds of information are for estimating how many (and, to a lesser extent, what kind of) errors a particular learner will make.

This research is part of an effort to develop a computer-based research system to aid in selecting application programs by engaging in a tutorial dialogue with a user. (Malhotra and Sheridan, 1976). This system is an example of individualized instruction; potentially, information specific to the current user could be used to ensure adequate learning. It is, therefore, information about the learner rather than information about the characteristics of the study material that we examine in this paper.

One such source of predictive information is learning time. If, within the range of times which motivated learners will choose, there were a strong relation between study time and performance, then one could hope to develop a metric to predict the time necessary for a given task. If someone spent too little time, a tutor could recommend more study. We could find no existing studies of the extent to which subject-controlled time predicts performance in learning complex material.

Another potential source of information is the learner's own assessment of how much he or she has learned. To the extent that learners' self-assessments are accurate, learning times will be unpredictive of performance since the motivated learner will take as much time as necessary. In a previous study (Thomas, 1976) in which subjects engaged in a dialogue to learn about the same order-handling system used in this experiment, subjects generally judged fairly accurately their own ability to simulate the system. And Thomas and Gould (1975) reported a strong relation ($r=.86, p<.005$) between a subject's mean confidence rating and a test of his learning of a query language. Laboratory studies using nonsense material (cf. Blake, 1973) have also demonstrated fairly accurate self-assessment of memory.

Finally, we are interested in the predictive utility of a measure of subjects' prior knowledge about invoicing systems. We know that a proper measure of relevant prior knowledge is a reasonably good predictor of learning performance when all subjects have the same amount of study time (e.g., Mayer, Stiehl, and Greeno, 1975). But the extent to which subjects can compensate for low prior knowledge by studying longer when strongly motivated to do so is unknown. It makes sense to investigate such a question only if all of the information needed to understand the procedure is given in terms understandable to every subject. While we cannot be sure that this was the case here, we tried to ensure maximum comprehensibility by pilot testing the procedural description, including a small glossary of key concepts, and excluding those few subjects for whom the material was clearly too difficult.

We are interested in both the strength of the relationship between prior knowledge and learning performance and the extent of correspondence between particular pieces of prior knowledge and parts of the to-be-learned procedure. An error classification scheme will be introduced to aid in the search for such correspondence.

Method

Subjects. Twenty-eight subjects from a part-time employment agency were used. There were four men and 24 women. Data from five of the subjects were not analyzable; two of these dropped out of the experiment, and the other three were not native English speakers and had inordinate difficulty mastering the written material.

Procedure. The experiment required completion of four tasks: (1) *forms guessing,* (2) *learning an invoicing system,* (3) *predicting performance on sample invoice* and (4) *completing a sample invoice*. Subjects were first asked to write down the number of months experience they had had in billing or invoicing work. They were then given the forms guessing task, the instructions for which are in Appendix A. Subjects were given a general idea of what the purposes of orders and invoices would be for a particular small business. (The same hypothetical business, a carpet importer called the Magic Carpet Company, was used throughout the experiment). They were asked to guess what the order and invoice forms would look like for Magic Carpet Company, and they were offered a bonus according to how much of the information needed on a real order and invoice they were able to guess. It was emphasized that visual appearance per se was unimportant; that we were interested in the kinds of information displayed. Subjects were given thirty minutes for the task, and no one needed more time.

After completion of the forms guessing (and a ten-minute break), subjects were given a description of an invoice handling system (together with sample documents and files), and were told to study the system until they knew it completely. They were told that after studying the description, they would be asked to demonstrate their knowledge by actually filling out an invoice. They were offered an "all-or-none" bonus: if the test invoice was filled out completely correctly, they would receive the bonus; if even one mistake were made, they would get no bonus. This was made abundantly clear in the instructions (reproduced in Appendix B). There was no limit on the amount of time they could spend studying.

After a subject had finished studying the description, a short questionnaire was given which asked for a prediction of whether or not the test invoice would be correctly filled out. Then, the test invoice was given, and an unlimited time was allowed to fill it out (times required by each subject to study the description and to fill out the invoice were recorded). After filling out the invoice, the subject answered a questionnaire about the strategies used to remember the material (the first 7 subjects did not receive this questionnaire, but were debriefed orally). Then the completed invoice was checked for errors. If an error was found, the subject was asked to explain how the incorrect number was obtained. The explanation of the erroneous process was noted (and, for sixteen subjects, tape recorded). As stated in the instructions, arithmetic errors were not counted as errors for purposes of granting the bonus, and they are not analyzed here.

It is important to note that the number of errors credited to a subject was the number of unique erroneous procedures, not the number of wrong invoice fields. The invoice called for computation of charges on two different items, but the same error made on both items was only counted once. Errors propagated through the remaining calculations were likewise counted only once.

Results

General Overview. Several qeneral features of performance on this task are of interest. First, though this is a fairly complex invoicing system whose BDL program (cf. Hammer, Howe, Kruskal and Wladawsky, 1975) involves 114 operations, there were, on the whole, few errors made. The 23 subjects made 59 errors, an average of 2.57 errors per subject (s.d.=2.53). Six

(26 per cent) of the subjects made no errors at all. Another six made only one error. The largest number of errors made by a subject was eight.

There was a fair amount of variability in times required to study the material; times ranged from 34 to 114 minutes (s.d.=20.1). Times required to complete the test invoice ranged from 20 to 75 minutes (s.d.=12.9). There was a small, insignificant positive correlation (r=.27) between study time and the time necessary to complete the invoice. Finally, there was insignificant small positive correlations between number of errors on the test invoice and 1) reading time, (.27), 2) completion time (.14), 3) total time (.27). This lack of sizeable correlations, together with the considerable variability in times, would be expected if people who differ in the skills necessary to do the task adjust their time to their skill level. Not only is study time a poor predictor of number of errors, but examination of Figure 1 (a scatterplot of number of errors as a function of study time), shows that there is no study time below which only unacceptable performances are observed (if one error is considered acceptable). Note that the relation is not linear but triangular. It may be summarized by saying that those who take little study time probably will do well, but the performance of those who take larger amounts of time cannot be predicted. This finding might be restricted to motivated learning situations, however. It might also be restricted to situations in which the entire mass of material to be learned is in front of the subject.

Self-Assessment of Learning. Sixteen of the subjects were given a short, two-question questionnaire immediately after studying the material and before being given the test invoice. The first question asked for the following prediction:

> Do you think that you will fill out the sample invoice completely correctly (except for arithmetic errors)?

A small bonus was offered for making the correct prediction. The second question presented four alternative statements about the subject's understanding of the material; each subject marked the single statement which best applied. Here are those statements:

(1) I believe that I know how to fill out the sample invoice correctly.

(2) I didn't understand some parts of the description, so I don't expect to be able to fill out the invoice. I don't think more studying will help.

(3) I understood the description, but I honestly doubt that I can remember everything long enough to fill out the invoice. I don't think more studying will help.

(4) I could be pretty sure of learning the procedure if I took more time, but I don't think it's worth it. I'm tired of studying.

(5) None of the above.

Four of the 16 subjects predicted that they would complete the invoice perfectly, and three of those four did so (the other made only one error. Two subjects did not feel that any of the statements were appropriate and they made 8 errors and 2 errors. Ten subjects predicted that they would make errors and all did so, averaging 3.7 errors. (Thus, it appears that, at least when motivated, people can be accurate in assessing what they know).

Error Classification. In this experiment, seven common errors accounted for nearly half (27) of the total; Appendix C gives a list of the exact errors made and the number of people who made each one. Remember that these are particular errors, not classes of errors, so that an individual can only make an error once on each item line in the invoice (errors which were made on both item lines of the test invoice were only counted once).

Table 1 gives a classification of these errors. An error is put in columns 1 or 2 if there is clear evidence from the recorded protocol that an operation was performed with incorrect operator(s) or operand(s). Column 3 lists errors in which there is no indication that the person knew that an operation was necessary. Column 4 contains cases in which it can be shown that a wrong file field was referenced while the errors in column 5 resulted from failure to recognize the necessity of referring to file information. Column 6 contains data type errors (like mistaking a percentage for a dollar amount, and vice versa). The errors in column 7 are unique in that they are errors which have no effect on the invoice total price. They are errors in which a person does some part of the procedure correctly, but merely places a wrong value in an invoice field. An example is placing the extended price discount *percentage* instead of epd *amount* in the epd field. Finally, column 8 and 9 contain intrusion errors in which a person fabricates some operation or file reference.

Insert Table 1 Here

There are several findings of interest concerning the distribution of errors. First, note that 25 of the 59 errors were complete omissions of parts of the process. Errors of this type are especially crucial; if someone remembers that an operation is necessary, but cannot remember how to perform it, there is at least a chance that known constraints and common sense (e.g., knowing that a discount is likely to be a small fraction of a price) can help reconstruct the correct procedure. Or, the person may seek outside help. Conversely, if a person completely forgets about an operation, he is not so likely to detect his difficulty.

Second, there are in these data fewer cases of incorrect operator than incorrect operand in calculations. There are, of course, slightly more operands than operators in the calculations of the procedure, but not five times as many. Perhaps there are more 'common sense' constraints placed on possible operators than on possible operands. This requires further experimental research.

Though the number of errors in each cell of this classification is small, it may nevertheless be of interest to look at one particularly important comparison in terms of this error classification. On the short questionnaire which subjects filled out before completing the test invoice, there was a question which allowed people to say whether or not they had understood the description. (Seven subjects were asked approximately the same question orally.) Seven of the 23 subjects said they hadn't understood; these subjects accounted for 36 of the errors. Table 2 shows the distribution of error types for subjects who said they did and subjects who said they did not understand. There is one rather large difference between the two groups; they differ in number of missing computations. If a statistical test were performed, this difference would not be significant because of the number of such comparisons made in the experiment. But it suggests further study of the possibility that one's feeling of understanding a complete procedure may be more dependent on knowing *what* subprocedures are to be done than knowing *how* they are to be done.

Insert Table 2 Here

Prior Knowledge. The total number of fields on the Magic Carpet order and invoice is 43; the average number of these fields guessed by the subjects was 15.2. The gross number of order and invoice fields which a given person was able to guess was a poor predictor of the number of errors he or she would make on the test invoice (r=-.37). However, there are a number of sources of noise in the gross measure. Some fields, for example, were repeated, being on both the order and on the invoice. Counting such fields twice arbitrarily gives them more predictive weight than they may deserve. Also, on the test invoice, the customer name and address and the invoice data were filled in by the experimenter. Since subjects could not have made an error on names, addresses and dates if they had wanted to, it seemed unwise to include these in the measure of prior information. (One given item, customer number, was included in the final measure of prior information because it was thought that a subject's use of a customer number might imply his or her consideration of the possibility of a file with information about customers). The revised measure of amount of prior information was the number of so-called 'critical fields' guessed. A list of the critical fields is given in Appendix D. It is important to note that these fields were selected *a priori* for the above reasons; they were not selected to maximize the relationship between critical fields and subsequent errors. Nevertheless, this relationship proved to be a very strong one (r=-.78).

One possible explanation of this is that those people who already have in memory a given task-relevant concept will learn more about its application in this particular procedure than will those who were not previously acquainted with it. We can call this the specific transfer hypothesis. If this hypothesis were true, one might expect to find a rough correspondence between an individual's performance on a given part of the invoice and whether or not he or she guessed the existence of that part. Surprisingly, no such correspondence exists in this data. For example, ten people guessed the need for a tax computation on the invoice. These people did make fewer errors on the tax-related parts of the invoice than the other twelve people (5 and 14 errors, respectively). However, on the parts of the invoice *not* concerned with taxes, the tax-guessers made only four errors, while the others made 28. Thus, it appears that the tax-guessers were just generally less likely to make errors than the others. This evidence does not, of course, rule out the specific transfer hypothesis: perhaps the correspondences of interest are at a more abstract level.

Two other hypotheses which could account for the predictive power of the critical fields measure are:

(1) Those who guess more critical fields are just more intelligent than those who do not and the more intelligent learn better.

(2) Those who are already familiar with some relevant concepts can focus their effort on the new ideas; those who do not have so many relevant preexisting concepts must learn more than is new. This hypothesis assumes that these latter people do not or cannot compensate sufficiently for the additional required effort by taking more time.

Nothing in the present experiment is relevant to distinguishing between these two hypotheses. Indeed, they may be two theoretical frameworks for viewing the same phenomenon.

Summary

A number of things have been learned about error prediction in this particular situation:

(1) Number of errors could not be predicted from study time.

(2) When they were motivated to do so, people could accurately predict whether or not they would make errors.

(3) Errors of complete omission of parts of the procedure were disturbingly common, comprising 42 percent of the errors made.

(4) A subject's ability to guess what information the invoice would contain was a very good predictor of how thoroughly he or she would subsequently learn the procedure.

(5) There was no direct correspondence between the particular fields guessed and the fields upon which errors were made.

There were also some statistically insignificant trends which may merit further investigation. For example, people appeared to be more likely to get the operands of a calculation wrong than the operators. Also, errors of omission of parts of the procedure seemed to be particularly common among those who said they did not understand the procedure; other kinds of errors were more evenly distributed between those who said they did and those who said they did not understand.

This experiment was an exploratory one, and it suggests some useful extensions. For example, knowing that people can predict their overall performance in this situation pretty well, can they also predict the *kind* of errors they are likely to make? The extent to which people can do this has a bearing on the usefulness to them of flexible learning situations such as a tutorial dialogue.

References

Blake, M. Prediction of recognition when recall fails: Exploring the feeling-of-knowing phenomenon. *Journal of Verbal Learning and Verbal Behavior, 12,* 311-319 (1973).

Hammer, M., Howe, W., Kruskal, V., and Wladawsky, I. A very high level language for data processing applications. IBM Technical Report RC5583, August, 1975.

Malhotra, A. and Sheridan, P., Experimental Determination of Design Requirements for a Program Explanation System. IBM Technical Report, RC5831, January, 1976.

Mayer, R., Stiehl, C., and Greeno, J., Acquisition of understanding and skill in relation to subjects' preparation and meaningfulness of instruction. *Journal of Educational Psychology,* 1975, (3), 331-350.

Thomas, J. and Gould, J., A psychology study of Query By Example. *AFIPS Conference Proceedings,* 1975 National Computer Conference. AFIPS Press: Montvale, New Jersey 1975.

Thomas J., A method for studying natural language dialogue, IBM Technical Report, RC5882, Febraury 1976.

Acknowledgements.

We would like to thank Lance Miller for his excellent technical comments and Martha McCrae for her help in the data analysis.

Appendix A

INSTRUCTIONS: PART ONE

In this experiment, you will be learning about how a typical small business goes about billing its customers. We will take as an example a business (called the Magic Carpet Company) that imports carpets and sells them to furniture stores in several states (these furniture stores will be referred to as "customers"). Magic Carpet Company gets filled-out order forms from its various customers, telling it what each customer wants to buy. The carpets are then sent to the customer, and an invoice is made out for each customer. The invoice lists all of the costs of what was ordered, and is sent to the customer for payment.

In the first part of this experiment, we want you to try to guess what these two forms, the order form and the invoice, would look like for Magic Carpet Company. (Assume that the various kinds of carpet can be ordered by number, so that a description of them on the order form is unnecessary). We suggest that you try to put yourself in the place of someone who has to use these forms: what kinds of information would you need to have? Draw out your versions of these forms on the paper provided. We have examples of a typical order and invoice for Magic Carpet Company, and you will be paid a bonus according to how close your forms are to ours. By "close", we do not mean similar in appearance: we mean containing similar information.

It is very important for the success of the experiment that you be as complete as you can; try to include everything on the forms that you think would be on the real thing.

Appendix B

INSTRUCTIONS: PART TWO

In this part of the experiment, you will get a written description telling exactly how Magic Carpet Company makes out its invoices. You will also get a blank order form, an invoice form, a list of abbreviations and some other sample documents. Your task is to study the description and sample documents until you are sure you could fill out an invoice given an order form, *without referring to this description and without making a single error*. When you feel you have learned the process completely, you will be asked to fill out a test invoice. (You will be able to refer to the sample documents and the list of abbreviations while doing the test invoice.)

In this experiment, we are interested in finding out how correct people are when they are sure they know something. Therefore, it is very important that you study the description until you are sure you can fill out an invoice perfectly. If you fill out the test invoice without a single error (except for obvious addition or multiplication fluffs), you will receive an extra two hours' pay. *If you make even one mistake, you will not get this bonus.* Take as much time to study the materials as you need to learn them completely.

LIST OF EXACT ERRORS MADE

Appendix C

	Subject Error	Number of People Made Each Error
C J	Doubled state and local tax for Item 03	2
T A	Used Fed Tax amount as a percentage for Item 03	2
C G	Didn't double Federal tax amount	3
	Multiplied State and Local tax % by GEXTPRICE. Not EXTPRICE	2
	Added GEXTPRICES to get GP1	2
	Doubled State and Local tax for Item 03 in Totals	2
N G	Put % instead of amount in EPD	7
	Used Fed Tax amount as a percentage	2
	Used liability code as State and Local tax percentage	5
	Forgot chain discounts formula	5
L S	Used listprice as unit price	2
	Percentage instead of amount in EPD	7
	Assumed EXTPRICE=GEXTPRICE	1
	Did not double Federal tax amount	3
	Used liability code as State and Local tax amount	5
	Set GP1 - sum of extended prices and taxes	1
	Forgot chain discount formula	5
S J	Put CUPCD in unitprice field	1
	Put percentage instead of amount in EPD	7
	EXTPRICE = GEXTPRICE - 1.00	1
	Did not double Federal tax amount	3
	Used Liability as State and Local tax amounts	5
	Did not double extended weight	2
	Did not know that chain discounts were percentages	2
	Did not add special charges	3
H S	Used listprice as unit price	2
	Used customer price code as extended price discount	1
	Did not check tax liabilities on items for State and Local tax	3
V F	Used EPD as percentage instead of amount	7
	Doubled extended price discount amount	1
	Used liability codes as State and Local tax percentage	5
	Applied special charges to GP2 instead of total due	1
L K	Did not check item liabilities for State and Local tax	3
S S	Added GEXTPRICES to get GP1	1
	Did not know that customer discounts were percentages	2
Y W	Used liability codes as State and Local tax percentages 5	5
M I	Used DCPCT for extended price discount	1

Appendix C - 2

	Did not know how to calculate extended price	1
	Multiplied Federal tax amount by extended weight	1
	Forgot to double extended weight	2
	Did not properly calculate GP1	1
	Did not properly do chain discounts	5
	Did not add special charges	3
C E	Put listprice in unit price field (but did not use it)	2
	Put percentage instead of amount in extended price discount	7
	Did not double GEXTPRICE for item 03	1
	Did not know chain discount formula	5
B M	Put percentage instead of amount in extended price discount	7
	Did not properly use chain discount formula	5
J C	Put percentage instead of amount in EPO	4
R G	Multiplied State tax % by GEXTPRICE	1
A V	Put listprice in unit price field	2
	Put percentage instead of amt. in extended price discount	7
	Used wrong State tax percentage	2
	Did not check item tax liabilities	3
	Used DCPCT for chain discounts	1
	Subtracted (instead of added) special charges	1

LIST OF CRITICAL FIELDS

Appendix D

ORDER

Customer Number
Item Number
Quantity Ordered

INVOICE

Quantity Shipped (or not shipped)
Unit Price
GEXTPRICE
Extended Price Discount
Extended Price
Federal Tax
State Tax
Local Tax
GROSSPRICE 1
Chain Discounts
GROSSPRICE 2
Total Federal Tax
Total State Tax
Total Local Tax
GROSSPRICE 3
Special Charges
Net Amount
Cash Discounts

TABLE 1

ERROR CLASSIFICATION

	Wrong Operator in Computation	Wrong Operator in Computation	Missing Computation	Reference to Wrong Field in File	Missing Reference to File Field	Wrong Data type	Wrong Item in Field	Extra Computation	Extra Reference to File Field	
# of Errors	2	10	14	1	11	9	7	4	1	(=59)

TABLE 2

ERROR CLASS BY WHETHER OR NOT S's SAID THEY UNDERSTOOD THE DESCRIPTION

	Wrong Operator in Computation	Wrong Operator in Computation	Missing Computation	Reference to Wrong Field in File	Missing Reference to File Field	Wrong Data type	Wrong Item in Field	Extra Computation	Extra Reference to File Field
Said they Understood (N=11)	0	6	2	0	5	4	2	3	1
Said they didn't understand (N=7)	2	4	12	1	6	5	5	1	0

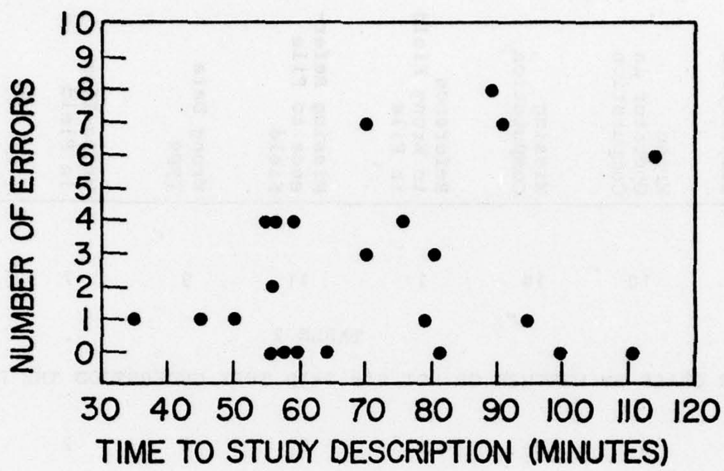

Figure 1. Scatterplot depicting number of separate errors on the invoice for each subject as a function of that subject's total study time.