Error Recognition in Calculus Problems: What Characterizes Expertise?

Alisha M. Crawley '95
Illinois Wesleyan University

Running head: ERROR RECOGNITION IN CALCULUS PROBLEMS

Error Recognition in Calculus Problems: What Characterizes Expertise?

Alisha M. Crawley

Illinois Wesleyan University

Acknowledgments

From the beginning to the end of this thesis project, I have received help from numerous people. First and foremost, I would like to thank my project advisor, Dr. Lionel Shapiro. Without Lon's help, the completion of this project would have been impossible. He provided me with encouragement when I needed it. He was always there to answer my questions and listen to my ideas. Also, his expert computer knowledge was invaluable.

I would also like to thank Dr. Johnna Shapiro, Dr. James Dougan, and Dr. Lawrence Stout for agreeing to be on my committee. I appreciate the time commitment they have made for me. I especially thank Dr. Dougan for his help on the statistical analysis.

I thank Mike Antholine for helping me with the data collection. I also thank him for his understanding when his participants decided not to show up at their assigned time.

Finally, I would like to thank all of my friends for putting up with my mood swings as I was trying to finish this project. I especially would like to thank Susan Reynolds for keeping me sane. Without her, this project would probably not have been completed.

Table of Contents

Abstract

Previous research in the area of expert-novice comparisons of mathematical problem solving has focused on the differences in categorization of and performance on math problems. These studies have led to the conclusion that while solving or categorizing problems, experts focus on deep processing and novices focus on surface structure. Other research dealing with true/false multiplication equations has shown that adults (considered experts in multiplication) can reject false answers before processing the equation. This study attempts to extend these findings by looking at the differences between experts and novices in the recognition of errors in true/false calculus verification expressions. The participants were professors (experts) and students (novices). The experiment consisted of participants answering 68 true/false calculus expressions (equations or conditionals) at three levels of difficulty. Reaction time, accuracy, and level of confidence were recorded. Experts were found to be quicker and more accurate overall. The experts were not able to reject the false problems more quickly than accepting the true ones. However, there was still some support for the hypothesis that experts are not only quantitatively better at task performance, but qualitatively different from novices in the type of processing they employ.

Error Recognition in Calculus Problems: What Characterizes Expertise?

Expertise theory has previously shown that experts process problems differently than novices do. This has been found across various domains. On the surface, it has been found that experts are faster than novices at the tasks in their domain. For example, reaction time is the defining characteristic of expertise in domains such as typewriting and mental calculation. Gentner (1988) used reaction time to define who his expert typists were in his study. Staszewski (1988) also used reaction time to define expert mental calculators. In tasks such as these, most people can complete the task, but the expert is the one who can do it quickly. The problem now is to find ways the problems are processed that would justify categorization of expertise by speed of processing.

In the domain of problem solving, reaction time is not necessarily a factor defining expertise, but experts usually are faster at solving problems in their domain than novices are. An interesting finding by Chi, Glaseri, and Rees is that although experts spend less time overall solving problems, they spend a greater proportion of their problem solving time representing problems than novices do (as cited in Lesgold et al., 1988). This leads to the question of how the experts are representing the problems.

In physics and mathematical expertise research, it has frequently been found that problem solvers represent problems in one of two ways, surface or deep structure. Using surface structure to categorize problems means using information found in the problem statement itself without taking anything else into account. Categorizing by deep structure means using the underlying theories and principles used in solving the problem. Not surprisingly, it has been found that experts generally use deep structure to categorize problems, whereas novices use surface structure. This has been found in both physics (Hardiman, Dufresne, and Mestre, 1989), and mathematics (Schoenfield, 1985). This coincides nicely with the idea that experts take a proportionately longer amount of time to represent problems than novices. Experts are spending representation time to begin solving the problem, but novices are not thinking that deeply yet. In fact, Chi, Feltovich, and Glaser (1981) found that novices use surface structure to define how to solve problems. Because trying to solve a problem using surface structure often leads to errors and "dead-ends", novices usually have longer reaction times and lower accuracy than experts. However, all is not lost for novices. It has been found that as expertise in physics increased, the method of solving went from surface to deep strategies (Anzai, 1985). From this body of research, experts are found to differ from novices in problem representation and problem solving.

How might this be used in other types of problem solving tasks? For example, what would happen if the task was determining whether answers to statements such as "3 x 3 = 7" are true or false, instead of actually categorizing and solving a particular problem? Researchers in arithmetic problem solving have been studying this type of problem. There have been some interesting findings concerning the rejection of false problems.

In a true/false verification problem, Zbrodoff and Logan (1990) hypothesized that verification is not always production plus comparison. Supporting this theory is the finding that there is faster than average rejection in certain types of problems. For example, Ashcraft and Stazyk (1981) found that people reject extreme splits such as "2 x 2 = 13" quickly. They believe this happens because people know that it is not conceivable for the answer to "2 x 2" to be so large. Krueger (1986) and Krueger and Hallford (1984) found that when certain arithmetic rules are broken, rejection comes quickly. So, if "2 x 2 = 5" were to be presented, there would be quick rejection because it is impossible to multiply two even numbers and come up with an odd number. The same rule is applicable to addition problems. Another type of problem that usually warrants a quick rejection is one in which signs are switched such as "3 + 4 = 12" or " 3 x 4 = 7" (Zbrodoff and Logan, 1986).

For the sake of argument, let's call all of the participants in these studies experts at arithmetic. They are not necessarily experts at calculation like the expert mental calculators in Staszewski's (1988) study. However, they have the experience to be labeled experts at addition and multiplication. They have a certain way of processing problems that includes quick rejection of false problems based on arithmetic rules and tricks. They were obviously not compared to novices, but their strategies could be useful in determining how experts in higher level problem solving accept and reject problems.

However, there are certain other types of problems that facilitate fast acceptance of true problems and slow rejection of false problems. Campbell (1987) found that certain false answers can prevent the participant from rejecting the problem quickly enough. We replicated this study for the purpose of checking to see what kind of effects we would find in the multiplication problems. It may be the case that there is a different set of thought processes involved in arithmetic as opposed to upper level mathematics.

Since most people are able to reject certain types of false arithmetic problems quickly, experts in the field of calculus should be able to reject certain types of false calculus problems quickly. They may be able to do this because of knowing certain rules or tricks, much like the odd-even rule of arithmetic. They might be able to do it quickly because of their deeper representation of the problem itself. However, there is much stronger evidence that the experts will be able to reject the false equations more quickly than accepting true equations.

Lesgold et al. (1988) found that medical experts diagnosing X-ray images will ignore unimportant surface inconsistencies, and focus right on the problem. Novices are distracted by unrelated flaws, leading to longer latency to finding the real problem and sometimes inaccurate diagnoses. The background essentially disappears for these experts, and their attention is automatically focused on the real problem within the X-ray image. In anagram solving tasks, Novick and Cote (1992) found that experts at anagrams saw the answers pop out at them (as cited in Matlin, 1994). In this case, the foreground is jumping out, instead of the background diminishing. In calculus problem solving, one of these two ideas might also be happening. The difficulty of the calculus task is that the problem space is not visual but mental. Therefore, calculus verification would be a combination of the two. However, the answer is not what is popping out at the expert. The inconsistency, or the mistake in reasoning of the problem, is what is popping out. Therefore, rejection of false problems should come more quickly than acceptance of the true problems as a result of the extraneous background disappearing and the mistake popping out. Novices, however, should have to search for the mistake and therefore take a longer overall.

In a calculus verification task, it should be found that experts are faster and more accurate than novices. It should also be found that experts are faster at rejecting false problems then accepting true ones. This finding is expected because, this apparently happens in arithmetic with certain problems, experts encode problems deeply, and inconsistencies in the statement might "pop out" at the expert.

Methods

Participants

Forty-nine participants from Illinois Wesleyan University participated. Two professors served as experts, and 47 undergraduate students served as novices. There were two groups of novices. Group 1 consisted of 25 students who had completed the equivalent of 0 or 1 semester of the three-semester calculus sequence at Illinois Wesleyan University. The mean age of group 1 was 19.27. There were 17 males and 8 females in this group. They had completed an average of .7 semesters of calculus. The second group consisted of 22 students who had completed 2 or 3 semesters of the sequence. The mean age of group 2 was 19.5. There were 9 males and 13 females in this group. They had completed an average of 2.86 semesters of calculus. Group 3 consisted of two experts, both males, with the mean age of 39. They had an average of 22 years of experience with calculus.

Materials

Stimuli for the calculus study consisted of 68 true/false mathematical expressions taken and modified from Purcell & Varberg (1987), a typical undergraduate calculus book (see appendix A for example expressions). 34 of the stimuli were true statements, and 34 were false. 34 of the expressions were equations, and 34 were if/then statements converted to an equation-type of form with an "implies" sign ($\Rightarrow$). There were 30 typical first-semester calculus problems, 26 second-semester calculus problems, and 12 third-semester calculus problems. Thus, most of the participants should be familiar with some of the problems. There was also a Likert scale to assess the participant's confidence that s/he answered the problem correctly.

Stimuli for the multiplication study consisted of 36 true/false multiplication expressions from Campbell (1987) (see appendix A for examples). 18 of the stimuli were true, and 18 were false. 18 were hard and 18 were easy. The experiment was run at a Macintosh Centris 610 using the SuperLab (1989) program. The Kaufman Brief Intelligence Test (K-BIT) (Kaufman & Kaufman, 1990) was used to ensure similar intelligence across groups (see Appendix B). The participants also completed surveys designed to check their level of calculus knowledge (see Appendix C). All participants signed a human-participant consent form (see Appendix D).

Procedure

Participants first signed the consent form and completed the participant survey. They were then given the K-BIT (Kaufman & Kaufman, 1990). After they had completed the K-BIT, they were asked to sit in front of the computer for instructions. They were instructed to hit one key labeled "true" or one key labeled "false" as quickly and accurately as possible when the arithmetic expression flashed on the screen. They were instructed to keep one finger on the "true" key, and one finger on the "false" key at all times. All participants used their right hand. When the subject read and understood the instructions, s/he pressed the "true" key. There was then a 500 millisecond (ms) inter-stimulus interval (ISI), after which the problem was displayed. When the participant answered, the problem was removed from the screen. After all of the arithmetic problems were finished, instructions appeared for the calculus expressions. When the subject read and understood the instructions, s/he pressed the "true" key. There was then a 500 millisecond (ms) inter-stimulus interval (ISI), after which the problem was displayed. When the participant answered, the problem was removed from the screen. There was then a second 500 ms delay, after which the Likert scale appeared with the instructions for the participant to choose how confident s/he was that s/he got the problem right on a scale of 1 (not at all confident) to 5 (very confident) (see Figure 1). The Superlab program recorded the participant's response to each problem, whether it was correct, the reaction time in milliseconds, and the confidence rating.

Results

The calculus results were analyzed using a 3 (expertise level) x 3 (problem difficulty) x 2 (problem type) x 2 (truth value) analysis of variance (ANOVA), with reaction time as the dependent variable. The results are shown in table 1. We were interested in the following effects. A significant effect was found for the interaction of group by truth ($F[2, 3286] = 3.311$, $p < .05$). A main effect was found for expertise level ($F[2, 3286] = 30.64$, $p < .04$). A Bonferroni/Dunn post-hoc test revealed that there was a significant difference between all three groups for expertise level (see Figure 2 and Table 2).

Other analyses include a separate analysis of reaction time on a subset of the problems where the confidence level was higher than three. The only significant effect was of expertise level ($F[2, 1693] = 9.179$, $p < .0001$). Also, an analysis was done on accuracy, as shown in Table 3. There was a significant main effect of expertise level ($F[1, 3286] = 40.127$, p

< .0001). Post-hoc tests showed that the difference between all three levels of expertise was significant ($p < .0001$) (see Figure 3).

Results of the multiplication study were analyzed using a 3 (expertise level) x 2 (truth value) x 2 (problem difficulty) ANOVA. The results revealed significant main effects of truth ($F[1, 1748] = 4.223$, $p < .05$) on reaction time. The mean of the true problems was 1372.374 ms, and the mean of the false problems was 1640.972 ms (see Table 4).

K-BIT scores revealed that all three groups were within one standard deviation of each other.

Discussion

The finding that experts are faster than novices in general supports the theory that experts complete tasks in their domain faster than novices do. It is very interesting to note that the experts had faster reaction times than both levels of novice. This seems to be a straightforward finding. However, the lower novices had significantly lower reaction times than the higher novices. But the low novices were only at about 50% accuracy. This indicates that there was a lot of blind guessing on the part of the lower novices. However, the experts were much faster than the lower novices, and were much more accurate. This supports the theory that experts are solving the problem in ways that are qualitatively different from novices. They are able to encode and to solve the problems accurately even more quickly than naive subjects are able to guess blindly!

For this particular study, however, we are interested in the interaction of expertise level and truth value. We expected to find that as expertise level increased, there is an increase in the difference in reaction times of true and false problems, with false problems taking a shorter amount of time than the true problems. However, we did not find this effect. There are a few reasons that this might be.

First and foremost, a major flaw in this study is the small sample size of experts. Since there are only two experts in our study, it is impossible for us to draw any definite conclusions. Because of the nature of expertise, it is hard to find experts to participate in studies. Even when they are found, there are problems. For example, Larkin, McDermitt, Simon & Simon have found that in the physics domain, novices use backward inferences to solve the problems, while experts use forward inference. This is almost always agreed upon in literature (as cited in Priest and Lindsay, 1992). However, when Priest and Lindsay (1992) used a much larger sample size, they found that both experts and novices use forward inference. Sample size might need to be taken into account. However, the experts in their study were graduate students, whereas the experts in our study are professors, who have many more years of experience than graduate students. In problem solving tasks, it is widely agreed upon that acquiring expertise takes time and practice. Perhaps their finding is a result of the use of graduate students, not sample size. Because of this problem, however, it is important to take sample size into consideration. Perhaps with a greater sample size, we would have seen the desired effect.

Another potential explanation is that experts may have different schemes for processing different types of problems. For example, we found the same thing as Campbell in that the true problems were actually faster than the false problems in the multiplication study. This is different from previous research in the field of arithmetic. So, perhaps the type of problem used is a factor. To see if this may be true, the interaction of problem type and expertise in the calculus study was considered. It was found that this was not a significant finding.

However, there may be a potential source of error in the fact that there are other differences in the problems themselves. The problems in this study were from chapter reviews, intended to make sure the students have learned the fine points of the chapter. The novices might not have even recognized that these were trick problems. Thus, a different way to look at these problems would be to have different levels of error as an independent variable. Perhaps experts are better at picking out deep or conceptual errors, whereas novices would be better at picking up on surface errors.

One other possible source of error is that not all novices are alike. Some may be better problem solvers, and therefore more expert-like in their processing. Mathematics is something that follows the phrase "use it or lose it". A different measurement may be using two different novice groups. All the novices would have the same level of formal calculus instruction. This instruction would have to be recent. The characteristic that would distinguish the two groups would be grades in the calculus classes. Then, if the hypothetical interaction is found, it would give even more support to the hypothesis that as novices become experts, they shift to expert thought processes, which would agree with the study by Anzai (1985).

This line of research may give us more understanding into the expert mind. Perhaps teaching methods could be modified to try to have children learning to think like the experts. Dufresne, Gerace, Hardiman, Mestre (1992) and Mestre, Dufresne, Gerace, and Hardiman (1993) have already found that teaching students to solve physics problems using expert-like strategies leads to more expert-like problem solving. Schoenfield (1985) found that this is also true for mathematical problem solving.

One possible application of this study is to take the findings and base education ideas on them. In this study, we found more support that the experts are quantitatively better than novices in solving problems in that they were able to accurately process problems more quickly than novices can guess. However, this study was not able to find exactly what these quantitative differences are. Consequently, this study would not contribute to the educational theories. However, this study could be added to the many others in this field, and reviewed with the idea of educational implications in mind. The study of the qualitative differences of processing between novices and experts is a step toward applying this type of research to a real-world situation, education.

References

Ashcraft, A. M. H., & Stazyk, E. H. (1981). Mental Addition: As three verification models. Memory and Cognition, 9, 185-196.

Anzai, Y. (1985). Learning and use of representations for physics expertise. In K. Anders Ericsson & Jacqui Smith (Eds.), Toward a General Theory of Expertise (pp 64-92). Cambridge: Cambridge University Press.

Campbell, J. I. D. (1987). Production, verification, and priming of multiplication facts. Memory and Cognition, 15, 349 - 364.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. Cognitive Science, 5, 121-152.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), Advances in the psychology of human intelligence (pp 7-75). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dufresne, R.J., Gerace, W.J., Hardiman, P.T., & Mestre, J.P. (1992). Constraining novices to perform expertlike problem analyses: Effects on schema acquisition, The Journal of the Learning Sciences, 2, 307-331.

Gentner, D. R. (1988). Expertise in Typewriting. In M. T. H. Chi, R. Glaser, & M.J. Farr (Eds.), The Nature of Expertise (pp 1-21). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hardiman, P.T., Dufresne, R., & Mestre, J.P. (1989). The relation between problem categorization and problem solving among experts and novices. Memory & Cognition, 17, 627-638.

Kaufman, A.S. & N.L. Kaufman (1990). Kaufman Brief Intelligence Test, American Guidance Service, Inc. Circle Pines: Minnesota.

Krueger, L. E. (1986). Why 2X2=5 looks so wrong: On the odd-even rule in product verification. Memory & Cognition, 14, 149.

Krueger, L. E. & Hallford, E. W. (1984). Why 2+2=5 looks so wrong: On the odd-even rule in sum verification. Memory & Cognition, 12, 171-180.

Larkin, J., McDermott, J., Simon, D. & Simon, H. (1980). Models of competence in solving physics problems. Cognitive Science, 3, 317-345.

Lesgold, A. , Rubinson, H., Feltovich, H., Glaser, R., Klopfer, D., Wang, Y. (1988). Expertise in a complex skill: diagnosing X-ray pictures. In M. T. H. Chi, R. Glaser, & M.J. Farr (Eds.), The Nature of Expertise (pp 1-21). Hillsdale, NJ: Lawrence Erlbaum Associates.

Matlin, M. W. (1994). Cognition (3rd ed.). Fort Worth: Harcourt Brace.

Mestre, J.P., Dufresne, R.J., Gerace, W.J., & Hardiman, P.T. (1993). Promoting skilled problem-solving behavior among beginning physics students, 30, 303-317.

Novick, L. R., & Cote, N. (1992). The nature of expertise in anagram solution. Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society (pp. 450-455). Hillsdale, NJ: Erlbaum.

Priest, A.G. & Lindsay, R.O. (1992). New light on novice-expert differences in physics problem solving. British Journal of Psychology, 83, 389-405.

Purcell, E.J. & Varberg, D. (1987). Calculus with analytic geometry (3rd ed.). New Jersey: Prentice-Hall.

Schoenfield, A. H. (1985). Mathematical Problem Solving. San Diego: Harcourt Brace Jovanovich.

Staszewski, J. J. (1988). Skilled memory and expert mental calculation. In M. T. H. Chi, R. Glaser, & M.J. Farr (Eds.), The Nature of Expertise (pp 71-127). Hillsdale, NJ: Lawrence Erlbaum Associates.

SuperlaLab (Version 1.4) [Computer software]. (1989). Cedrus Corporation.

Zbrodoff, N. J., & Logan, G.D. (1990). On the relation between production and verification tasks in the psychology of simple arithmetic. Memory & Cognition, 18, 83-97.

Zbrodoff, N. J., & Logan, G.D. (1986). On the autonomy of mental processes: A case study of arithmetic. Journal of Experimental Psychology: General, 115, 118-130.

Appendix A

Example Calculus and Multiplication Problems

Calculus

True equation:	$\pi^{\sqrt{2}} = e^{\sqrt{2}\ln\pi}$
False equation:	$[f(x)\cdot g(x)]' = f'(x)\cdot g'(x)$
False conditional:	$y = \pi^5 \Rightarrow D_x y = 5\pi^4$
True conditional:	$y = (x^3 + x)^8 \Rightarrow D_x^{25} y = 0$

Multiplication

True easy:	3 x 4 = 12
False easy:	5 x 6 = 35
True hard :	4 x 9 = 36
False hard:	7 x 8 = 48

Appendix B

K-Bit Answer Sheet

Subtest 1 VOCABULARY
Part A – Expressive Vocabulary

Item	Response	Score (circle)	Item	Response	Score (circle)
1. bed	______	1 0	26. screw	______	1 0
2. fork	______	1 0	27. saddle	______	1 0
3. frog	______	1 0	28. binoculars	______	1 0
4. hammer	______	1 0	29. globe	______	1 0
5. bus	______	1 0	30. cactus	______	1 0
6. umbrella	______	1 0	31. cash register	______	1 0
7. window	______	1 0	32. battery	______	1 0
8. piano	______	1 0	33. magnifying glass	______	1 0
9. ladder	______	1 0	34. anchor	______	1 0
10. drum	______	1 0	35. antlers	______	1 0
11. leaf	______	1 0	36. escalator	______	1 0
12. lamp	______	1 0	37. tweezers	______	1 0
13. flashlight	______	1 0	38. fire extinguisher	______	1 0
14. owl	______	1 0	39. life jacket	______	1 0
15. smoke	______	1 0	40. fire hydrant	______	1 0
16. feather	______	1 0	41. funnel	______	1 0
17. bridge	______	1 0	42. outlet	______	1 0
18. water fountain	______	1 0	43. compass	______	1 0
19. pencil sharpener	______	1 0	44. hexagon	______	1 0
20. calendar	______	1 0	45. thermostat	______	1 0
21. lock	______	1 0			
22. doorknob	______	1 0			
23. penguin	______	1 0			
24. hook	______	1 0			
25. ruler	______	1 0			

Ceiling Item = highest item administered
Errors = total number of items scored 0

Ceiling Item ( )
minus Errors ( )
equals Raw Score [ ]

Comments and Observations

2

Subtest 1 VOCABULARY
Part B – Definitions

Item	Response	Score (circle)	Item	Response	Score (circle)
Sample A	brown		21. confound	______	1 0
Sample B	antique		22. tenacious	______	1 0
1. school	______	1 0	23. hypochondriac	______	1 0
2. rainbow	______	1 0	24. chaos	______	1 0
3. sunshine	______	1 0	25. caricature	______	1 0
4. delicious	______	1 0	26. apathetic	______	1 0
5. plow	______	1 0	27. indelible	______	1 0
6. chimney	______	1 0	28. enamored	______	1 0
7. daisy	______	1 0	29. clandestine	______	1 0
8. uncle	______	1 0	30. omnipotent	______	1 0
9. Eskimos	______	1 0	31. lethargic	______	1 0
10. lie	______	1 0	32. contingent	______	1 0
11. gallant	______	1 0	33. rancor	______	1 0
12. experiment	______	1 0	34. valedictorian	______	1 0
13. evacuate	______	1 0	35. acquiesce	______	1 0
14. jeopardy	______	1 0	36. enigma	______	1 0
15. gratitude	______	1 0	37. serendipity	______	1 0
16. elevate	______	1 0			
17. sufficient	______	1 0			
18. indicate	______	1 0			
19. vitamin	______	1 0			
20. relinquish	______	1 0			

Ceiling Item ( )
minus Errors ( )
equals Raw Score [ ]

Comments and Observations

3

Subtest 2 MATRICES

Item	Response	Score (circle)	Item	Response
Sample A	A		25. A	___
1. A	___	1 0	26. H	___
2. C	___	1 0	27. D	___
3. E	___	1 0	28. H	___
4. D	___	1 0	29. C	___
5. A	___	1 0	30. F	___
6. C	___	1 0	31. B	___
7. B	___	1 0	32. G	___
8. D	___	1 0	33. G	___
9. E	___	1 0	34. G	___
Sample B	C		35. C	___
Ages 6-10 go to item 10. Ages 11-90 go to item 15.			36. E	___
			37. C	___
			38. D	___
10. A	___	1 0	39. A	___
11. F	___	1 0		
12. B	___	1 0	40. H	___
13. F	___	1 0	41. A	___
14. C	___	1 0	42. H	___
			43. B	___
15. B	___	1 0	44. A	___
16. A	___	1 0		
17. H	___	1 0	45. B	___
18. C	___	1 0	46. B	___
19. G	___	1 0	47. A	___
			48. G	___
20. A	___	1 0		
21. D	___	1 0		
22. F	___	1 0		
23. E	___	1 0		
24. E	___	1 0		

Ceiling Item ( )
minus Errors ( )
equals Raw Score [ ]

Comments and Observations

4

See "Discontinue Rule" in Chapter 2 of the K-BIT manual for information on when to discontinue testing on each task.

See "Obtaining Raw Scores" in Chapter 3 of the K-BIT manual for information on how to calculate raw scores.

Suggestions for follow-up testing or other recommendations

For additional forms, call or write AGS, 4201 Woodland Road, Circle Pines, MN, 55014-1796; toll-free 1-800-328-2560, in Canada 1-800-263-3558. Ask for item # QT3352, K-BIT Individual Test Records (25/package).

Kaufman Brief Intelligence Test

SCORE PROFILE

See "Graphing Standard Score" in Chapter 3 of the K-BIT Manual for directions.

	Standard Score ± Band of Error	40	50	60	70	80	90	100	110	120	130	140	150	160
Vocabulary	±													
Matrices	±													
K-BIT IQ COMPOSITE	±													

percentile rank: 1 2 5 9 16 25 37 50 63 75 84 91 95 98 99

−4 SD −3 SD −2 SD −1 SD MEAN +1 SD +2 SD +3 SD +4 SD

Printed on recycled paper

K·BIT KAUFMAN BRIEF INTELLIGENCE TEST

by Alan S. Kaufman and Nadeen L. Kaufman

Individual Test Record

Name ___ Sex ___

Parents' names ___

Home address ___ Phone ___

Grade (if enrolled in school) ___ Highest grade completed (if no longer in school) ___

Current or previous occupation (if applicable) ___

Examiner ___

	YEAR	MONTH	DAY
Test date			
Birth date			
Chronological age			

SUBTESTS	Raw Score	Subtest Standard Score ± Band of Error ___% Confidence (Table C.1, C.3)	National % ile Rank (Table C.4)	Descriptive Category (Table C.4)	Other Data (please specify)
Expressive Vocabulary					
Definitions	+				
VOCABULARY		±			
MATRICES		±			
Sum of Subtest Standard Scores			Transfer sum to K-BIT IQ Composite below		

COMPOSITE	Sum of Subtest Standard Scores	Composite Standard Score ± Band of Error ___% Confidence (Table C.2, C.3)	National % ile Rank (Table C.4)	Descriptive Category (Table C.4)	Other Data (please specify)
K-BIT IQ COMPOSITE		±			

Subtest Standard Score Comparison	Vocabulary Standard Score	Matrices Standard Score	Standard Score Difference	Significance Level (please circle) Table C.5
				NS .05 .01

Appendix C

Surveys

Background Questionnaire - Expert

Please answer the following questions to the best of your ability.

1. Age (years): ________

2. Sex:
   ______ MALE
   ______ FEMALE

3. Please list some of your hobbies:

   ____

4. Please list all of the classes you are teaching this semester:

   ____

   ____

   ____

5. How much do you use calculus on a daily basis?
   ______ NEVER
   ______ RARELY
   ______ SOMETIMES
   ______ OFTEN
   ______ ALWAYS

6. How many years have you been working with calculus? ______

7. About how many years has it been since you've had formal instruction in calculus?

   ____

8. How good of a calculus user do you feel you are?
   ______ EXCELLENT
   ______ GOOD
   ______ AVERAGE
   ______ BELOW AVERAGE
   ______ POOR

Background Questionnaire - Novice

Please answer the following questions to the best of your ability.

1. Age (years): _______

2. Year in school: _______

3. Sex
   _____ MALE
   _____ FEMALE

4. Major: _____________ Minor: _____________

5. Please list some of your hobbies:

   ____

6. Right now, what career do you plan to go into? Please be as specific as possible.

   ____

7. Please list all of the college math courses you have taken and when you took them: (ex. Calculus 1, 2, linear algebra - freshman year, probability, statistics - sophomore year, etc.). Write (IP) after any classes in progress.

   ____

   ____

8. How much do you use calculus on a daily basis?
   _____ NEVER
   _____ RARELY
   _____ SOMETIMES
   _____ OFTEN
   _____ ALWAYS

9. How good of a calculus student do you feel you are?
   _____ EXCELLENT
   _____ GOOD
   _____ AVERAGE
   _____ BELOW AVERAGE
   _____ POOR

Appendix D

Consent Form

Illinois Wesleyan University
Department of Psychology
Consent Form for Research Participants

Title of Study: Experience and knowledge representation in mathematical problem solving
Principle Investigator: Lionel R. Shapiro, Ph.D.

This is a study of thinking and how thinking may change under different conditions. We are investigating whether factors such as experience change the way that people solve problems. As a participant, you may be asked some general information questions pertaining to your medical and educational background and then be given two tests: a brief intelligence test which takes approximately 30 minutes, and a test involving the solution of mathematical problems, which also takes approximately 30 minutes and is administered on a computer. (Please note that no computer expertise is required and that your use of the computer will consist only of pressing one of two keys.)

The intelligence test contains items related to your vocabulary and your ability to solve spatial problems. The computer test requires you to identify mathematical expressions as either correct or incorrect. You will be given several sets of these expressions and the time it takes you to solve them will be measured.

Your intelligence test score, as well as your solution times, will be kept completely confidential. Although the data collected today may be published in the future, your name will never be connected with your scores or with the study in published form.

There are no known risks involved with this study, and although some participants may find the problems challenging, most do not find the tasks uncomfortable.

There are no known direct benefits to you as a result of your participation in this study, but your participation may help others indirectly by providing us with information on the nature of thinking.

As a participant in this study, you have the right to ask questions pertaining to the clarification of your tasks, and to be informed of the nature of the study before you begin. Your participation is voluntary, and as such, you have the right to refuse to participate or to withdraw from the study at any time, with no penalty or loss of benefit. You will receive additional information about the study following your participation. You may, if you wish, receive a copy of this consent form.

By signing below, you acknowledge that you have read this consent form and you understand your rights in this study.

Name of participant (please print) ____________________

Signature of participant ____________________

Date signed ____________

Name of experimenter ____________________

Signature of experimenter ____________________

Date signed ____________

Name of witness ____________________

Signature of witness ____________________

Date signed ____________

Table 1

Analysis of Variance for Reaction Time for Calculus

Source	df	SS	MS	F-value
E	2	12724763634.103	6362381817.052	30.064*
T	1	41305657.271	41305657.271	.195
E x T	2	1401547335.765	700773667.883	3.311*
PT	1	41956638.517	41956638.517	.198
E x PT	2	268637530.166	134318765.083	.635
T x PT	1	352702967.913	352702967.913	1.667
E x T x PT	2	1474004025.705	737002012.853	3.482*
D	2	1617985744.810	808992872.405	3.823*
E x D	4	1186129629.404	296532407.351	1.401
T x D	2	655380452.882	327690226.441	1.548
E x T x D	4	341615234.843	85403808.711	.404
PT x D	2	382896562.054	191448281.027	.905
E x PT x D	4	255906919.039	63976729.76	.302
T x PT x D	2	42135925.601	.21067962.800	.100
E x T x PT x D	4	248650926.008	62162731.502	.294
Residual	3286	695419374332.719	211630972.104	

Note: E = Expertise Level, T = Truth Level, PT = Problem Type, D = Problem Difficulty

* $p < .05$.

Table 2

Mean reaction time and accuracy as a function of expertise level, for Calculus

Expertise Level	Reaction Time	Accuracy
Low Novice	13074.196 ms	.532
High Novice	16898.068 ms	.640
Expert	7293.089 ms	.948

Table 3

Analysis of Variance for Accuracy for Calculus

Source	df	SS	MS	F-value
E	2	18.413	9.207	40.127*
T	1	.521	.521	2.271
E x T	2	.114	.057	.249
PT	1	.902	.902	3.933*
E x PT	2	.385	.193	.839
T x PT	1	.112	.112	.489
E x T x PT	2	.613	.306	1.335
D	2	.138	.069	.301
E x D	4	.510	.128	.556
T x D	2	.099	.049	.216
E x T x D	4	1.198	.299	1.305
PT x D	2	.853	.426	1.859
E x PT x D	4	.882	.221	.961
T x PT x D	2	.215	.107	.468
E x T x PT x D	4	.146	.036	.159
Residual	3286	753.929	.229	

Note: E = Expertise Level, T = Truth Level, PT = Problem Type, D = Problem Difficulty

* $p < .05$.

Table 4

Analysis of Variance for Reaction Time for Multiplication

Source	df	SS	MS	F-value
E	2	22137226.939	11068613.470	6.298*
D	1	14648335.386	14648335.386	8.335*
E x D	2	7825734.005	3912867.002	2.227
T	1	7420788.580	7420788.580	4.223*
E x T	2	1215385.626	607692.813	.346
D x T	1	2141.601	2141.601	.001
E x D x T	2	697050.763	348525.381	.198
Residual	1748	3071847919.728	1757350.068	

Note E = Expertise Level, D = Problem Difficulty, T = Truth Level
* $p < .05$.

Figure Caption

Figure 1. Likert scale for confidence judgments.

Figure 2. Reaction time as a function of expertise level in calculus (with error bars representing one standard error)

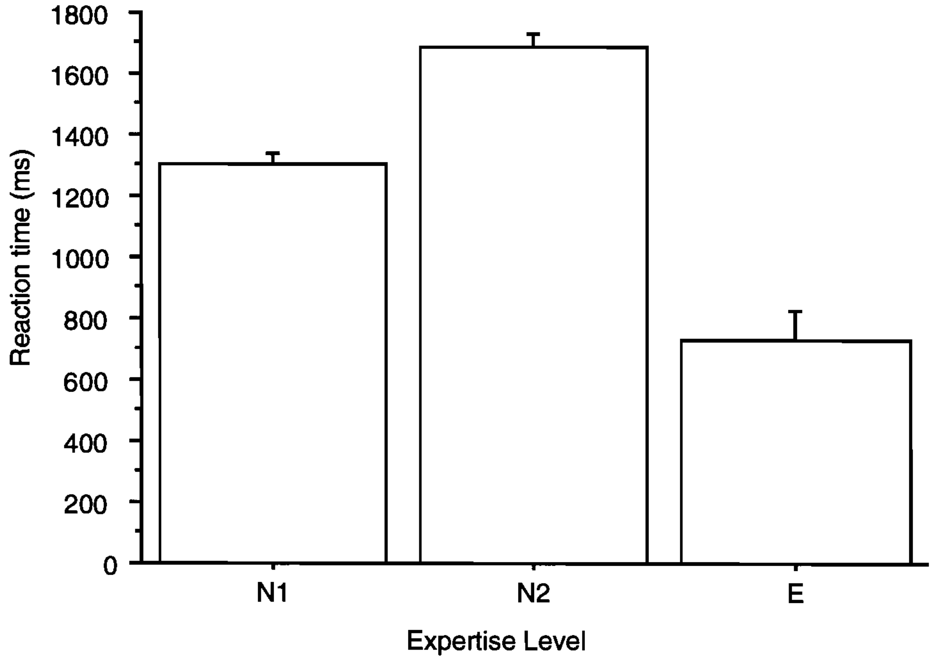
Figure 3. Accuracy as a function of expertise level in calculus (with error bars representing one standard error)

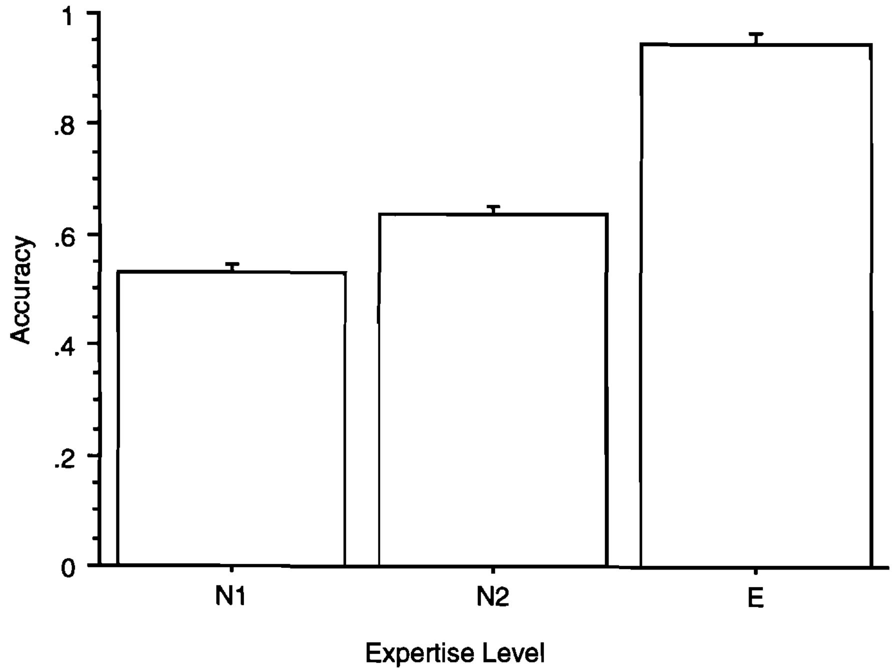
How confident are you of your answer on a scale of 1 to 5?

Not at all confident

Very confident

1 2 3 4 5


1800
1600
1400
1200
1000
800
600
400
200
0
Reaction time (ms)
N1
N2
E
Expertise Level



1
.8
.6
.4
.2
0
Accuracy
N1
N2
E
Expertise Level