1991

DOES OUR COMPLEX WRITING LOWER TEST SCORES ON MATHEMATICS WORD PROBLEMS?

William (Bill) H. Williams
CUNY Hunter College

Sandra P. Clarkson
Professor

DOES OUR COMPLEX WRITING LOWER TEST SCORES ON MATHEMATICS WORD PROBLEMS?

S. P. Clarkson and W. H. Williams, Hunter College of CUNY
Department of Mathematics and Statistics, New York, NY, 10021

ABSTRACT: This paper describes one of a series of studies underway at Hunter College to determine whether students' reading proficiency level affects their performance on mathematics "word" problems. Based on this study, we reached some specific conclusions:
1. Reading ability is a separate, quantifiable factor which impacts the performance of all students on mathematics word problems.
2. Less complex writing leads to better results on word problems for all students.
3. Less complex writing leads to even more improvement in test results for “weaker” readers [those needing reading remediation] than for “average” readers [those exempting reading remediation].
KEYWORDS: MATHEMATICS; READABILITY; READING; WORD PROBLEM; COLLEGE STUDENTS

1. MATH SCORES DEPEND DIRECTLY ON READING ABILITY

Many college, pre-calculus mathematics courses test much more than mathematics alone. Solving "word" problems in the language of mathematics first requires a clear understanding of the language in which the problems are written, followed by the related, but different, ability to translate the verbal language into symbolic mathematical expressions. Only then does the specific ability to manipulate the symbolic "words" of mathematics come into play. The overall result is that if a "word" problem is done incorrectly, there is no unambiguous way to conclude that the fault lies in any well-defined area of mathematics rather than in reading proficiency.

Four important factors affect performance on mathematics "word" problems. They are the level of mathematics complexity, the level of writing complexity, and the student's proficiency in ***both*** reading and mathematics.

Since the number of students in college with English as second language (ESL) is growing, the interrelationships between mathematics and language are of proportionally increasing importance. At Hunter College, approximately 40% of the students have English as a second language. Often, other Hunter students lack English proficiency. Would these students fare better if additional reading courses were mandated prior to their taking mathematics? This study suggests they would. Students may be able to perform strictly manipulative mathematics, but may not be able to solve word problems or understand more complex mathematical concepts due to weak reading skills. Furthermore, this problem is not necessarily confined to students who are clearly reading-deficient. It may also exist for students who have achieved an "acceptable" level of reading expertise. After all, there are varying levels of complexity in English.

This paper presents the results of one of a series of studies underway at Hunter College to determine how students' reading proficiency level affects their performance on mathematics "word" problems. Based on this study, we reached some specific conclusions:

1. Reading ability is a separate, quantifiable factor which impacting the performance of all students on mathematics word problems.
2. Less complex writing leads to better results on word problems for all students.

3. Less complex writing leads to even more improvement in test results for weaker readers than for average readers.

The results of this research appear to confirm the results of Mestre and Gerace (1981) who found, by different methodology, that their “additional reading training” improved students’ mathematics scores. Similar **quantitative** studies on college students do not seem to have appeared elsewhere in the literature. Furthermore, the magnitude of the reading effects observed in this study are larger than the differences typically reported in studies of the weaker performances of urban students on mathematics tests. This raises the possibility that a significant percentage of those reported differences may be the result of minorities and ESL students, for one reason or another, not having been taught to read well enough to deal with the material required by mathematics and science courses.

2. REVIEW OF THE LITERATURE

Studies with elementary school students indicate that reading ability and computation proficiency are factors important to success in solving word problems (Balow 1964; Cohen and Stover 1981; Glennon and Callahan 1968; West 1977). However, studies trying to relate the two abilities have offered conflicting results. Although there appears to be an improvement in student achievement when reading instruction is included within the instruction of mathematics (Aiken 1972), the effect of reading is not always so apparent. On the Metropolitan Achievement Test, reading ability does not seem to be a factor in the failure rate of students. In several studies, fully 49% of the errors were determined to be computational, not based on a misreading of the problem (Knifong and Holtan 1976, 1977).

There is no clear or consistent agreement on what determines the "readability level of mathematics material," although there is an extensive literature on the subject [see particularly *Syntax Variables and Reading Difficulty* by Jeffrey Barnett in Task Variables in Mathematical Problem Solving, Goldin and McClintock, eds. 1984). Some studies indicate that the (measured) "readability level" of the material appears to influence problem difficulty (Thompson 1967; Linville 1969); another found that readability level does not affect problem solving performance (Paul, Nibbelink, and Hoover 1986). Certain factors, however, clearly influence problem difficulty.

1) Vocabulary: Certainly the difficulty level of word problems in mathematics is influenced by the **vocabulary** used in the problems. In some situations, that vocabulary is fixed--for example, in physics-based word problems, we would necessarily be using words like *acceleration* and *velocity*. These words have specific mathematical meanings. Certainly, specific knowledge of *mathematical* vocabulary is important. However, even non-mathematical vocabulary can influence problem difficulty. This non-mathematical vocabulary includes verbal "cues" or key words (Steffe 1967; Jerman and Rees 1972; Nesher and Teubal 1975; DeCorte, Verschaffel and Verschueren 1982). Instruction in vocabulary has been shown to raise scores on verbal problem solving tasks (Henney 1971; VanderLinde 1974).

2) Grammatical structure: The **grammatical structure** of the problem statement affects its difficulty. Clearly, “grammatical structure” is not a quantifiable concept. It is more like a swarm of gnats--individually a nuisance, collectively, debilitating. Some of the identifiable problem areas have been discussed in the literature. Both the number of words (Jerman and Rees 1972) and the sentence length (Jerman and Mirman 1974) affect the difficulty level of problems. The

position of material can influence the difficulty level also. Placing the question first appears to focus the student on what is desired and appears to make problems easier (Williams and McCreight 1965). The position of certain other content and the placement of punctuation can also influence problem difficulty (Nesher 1982; Riley, Greeno, and Heller 1983). In word problems, when the "subject" is referred to (later) in the problem by a pronoun or other name, the literature indicates that the student's ability to solve such problems decreases (Dutka 1979; Barnitz 1979). When the reference is ambiguous or at some "distance" from the original subject name, the bilingual student has a particularly difficult time. This usually causes a strain on the short-term memory of the student, often requiring rereading and a longer "processing" time (Segalla 1973). The passive voice is a problem for students, particularly bilingual students.

3) Problem setting: Lower ability students seem to perform better on word problems set in a familiar **context** (buying groceries, playing games) (Lyda and Church 1964). In a study by Brownell and Stretch (1931), there is additional support given to the idea that context familiarity influences problem difficulty. Caldwell and Goldin (1979), showed that concrete problems (those describing real life situations) were substantially less difficult than problems in an abstract (mathematical) setting.

4) Mathematics content: The words themselves and how they are arranged are not the only source of difficulty in a word problem. The **mathematical computation** required for a solution also affects the difficulty. The magnitude of numbers (Suppes, Loftus and Jerman 1972; Houlihan and Ginsburg 1981), the number and type of operations and steps (Suppes et al. 1972; Whitlock 1974; Searle, Lorton, and Suppes 1974; Sherard 1974) and the sequence of operations (Berglund-Gray and Young 1940) all influence the difficulty level of word problems. Not only does number size influence computational difficulty, it even influences the ability to choose the correct operation (Bell, Swan, and Taylor 1981; Fischbein, Deri, Nello, and Marino 1985). And, although this is a conceptual problem, rather than a computational one, even the order that the numbers appear in can influence the difficulty level (Burns and Yonnally 1964). And some studies indicate that the complex relationship between the verbal description of the problem solving situation and the equation representing it may also interfere with solving the problem (Hiebert 1982)--that is, when the “words” of the problem do not follow an order that can be directly translated into an equation, the problem is more difficult for students.

3. THE STRUCTURE OF THE EXPERIMENTAL EXERCISES

In any mathematics class, instructors introduce new vocabulary and provide instruction in the necessary mathematics. They provide exposure to different types of problems and solve them for the students. They also suggest, or assign, a reasonable textbook for student reference and use during the course. However, when the time comes to test these same students, instructors often sit down and write up several questions, type and duplicate them without stopping to consider that one of the most important uses of language in mathematics is in assessing student knowledge of that same mathematics.

To aid with language-related deficiencies, one might think that students should be tested in their own language. But the literature shows that the more closely the language used in testing parallels the language used in instruction, the better the students do. One study (Llabre and Cuevas 1983), suggests that students should be tested in the same language used in class; that language produces the highest test results, particularly among Hispanic students.

But given that we are testing in English, is there anything we can do about the specific, written style of the exams that can help our students? In preparing for this study, we wanted immediate practical application to the classroom. Since "test banks" are provided with most contemporary texts, we wanted to find some simple changes that could be made to already-written exam questions that would result in higher student scores. To do this, we needed to be able to measure writing complexity. Early popular measures of readability (Dale-Chall 1948; Spache 1953; Harris and Jacobson 1973; Fry 1968, 1977;) and the Cloze technique (Taylor, 1953; Kane, 1970) have been applied to mathematics text materials. These procedures are time-consuming to apply. And while there is agreement that the readability level of materials affects performance, there is no agreement as to what “levels” are appropriate for college students. In fact, little attention seems to be paid to the reading level of materials in college texts. With the appearance recently of various computer-based programs for analyzing English, see for example Grammatik (Smye 1987) and Writers Workbench (Cherry 1982; Cherry et al. 1983), it has become somewhat easier to classify materials, if only roughly, for the purposes of determining the reading effect on college students. Specific formulas for measuring English complexity are used in these computer-based programs, see for example Flesch, Flesch-Kincaid, Gunning-Fog, and Fog Indices (Bunde 1975; Cramer 1978; Kent 1983; Sullivan 1980). For the purposes of this research, we decided to use a simple, straightforward approach to classifying the reading difficulty of the problems--a procedure that any instructor can use.

Once we identified the "test bank" problems, we measured the complexity level of each of them, using various software programs, and then rewrote each problem with a less complex writing style. The rewritten questions contained ***exactly*** the same mathematics content as the original problem. But the English was simplified by making three types of changes. We first made all sentences active. The use of passive sentences seems to immediately present a problem for readers, even for native English users. Next, we shortened the sentences. Using "if-then" constructions and long, complex sentences doesn't make the mathematics more difficult, but it does make the English less accessible. Then we verified that the chosen problem settings were not culturally specific. Using baseball as an example may be reasonable for native-born students, but for immigrants, the choice may result in making a mathematics problem impossible. After rewriting the problems, we again checked the difficulty level. The result of this procedure was that we had pairs of questions, which were identical mathematically, but of different levels of readability. [See the appendix for examples of the questions.]

In reviewing the variability of the various quantitative complexity measurement systems over the test questions, we decided to use these measurement systems only to classify the paired questions into two classes, "standard" and "revised". There were two related reasons for this: first, the numerical measures of the individual questions are based on relatively little writing, often a paragraph; and second, the different software systems did not always yield a consistent numerical ordering of the difficulty of the questions. However, ***all*** of the measurement systems ***did*** agree that the "revised" versions were written in simpler English than the original, "standard" questions. Consequently, the decision was made to simply classify the different versions of the questions into two groups.

Before conducting the test, we introduced one additional factor. An unclear referent was introduced into one of the revised questions. This reference was unclear in the strict sense, although the meaning was unambiguous in the context of the problem. Nevertheless, this introduced an additional factor which could potentially confuse the student; *and most importantly, every one of the complexity measurements missed this unclear referent entirely.*

Since the "revised" problems are easier to read and understand, these problems should, theoretically, be easier to solve. The basic hypothesis of this study is that student scores on three of the revised problems would be higher than the scores on the corresponding standard problems. For the problem with the unclear referent, the results are not so clearly predicted. Would the easier writing or the unclear referent dominate? A review of earlier research on the use of such grammatical constructions (Barnitz 1979) indicates that elementary school students found unclear referents more difficult, so it was our overall expectation that college students would ***not*** do well on this particular question either.

4. THE STUDENT SUBJECTS

This study was conducted in a one-semester, first course in statistics in which word problems are common. Most of the students taking this course choose it as one of the four mathematics/science courses required for the basic distribution requirement at Hunter College. Many students are liberal arts majors and the remainder take the course as a requirement in health science or nursing programs.

Some standard information was available on the students; specifically, age, sex and scores on incoming placement tests in mathematics, physical science, reading and writing. About one half (32) of the 67 students were required to take the reading placement test at least twice. This Hunter placement test assesses reading comprehension and has a passing score comparable to the reading comprehension of an 11th grade student. Students who do not pass the reading test on the first attempt are required to take a reading course followed by the reading test each semester until they achieve a passing score. The mathematics placement test assesses skills in arithmetic and basic algebra and, as in the reading sequence, students who require two or more attempts at the mathematics test also must take and complete a mastery-based mathematics course. Students who pass the placement tests on the first attempt are exempted from the associated courses.

5. THE METHODOLOGY

Two groups of students within the class were arranged to be as statistically alike as possible. And analysis of the two groups did confirm that the groups were alike in all of the available student characteristics and, most importantly, this included the reading and mathematics placement tests. Furthermore, repeated analysis of in-class student test scores confirmed that there were indeed no significant differences between the two groups. (Also, a subsequent test of the difference between the two groups on the test totals of the questions discussed in this paper was associated with a p value of 0.27). Differences between the two groups were consistently quite small.

Two banks of test questions, bank "C" and bank "E", were used. The references "C" and "E" are simply taken from the first letter of the first question on each of the versions. One of each of the paired questions was placed on bank "C" and the other one on bank "E"; so if bank "C" received the standard version of a question, then bank "E" received the revised version. Each bank was given two revised questions and two standard problems plus a control problem, which was identical on both of the banks. Thirty-five students were given bank "E" and thirty-two students bank "C".

Although an attempt was made to balance the overall difficulty of the "C" and "E" question sets, significant differences between the "C" and "E" groups on the ***individual*** questions ***were expected*** . Further, since the matched pairs of experimental questions had identical mathematics, the expected differences could be directly attributed to differences in writing complexity.

In this study, no male/female differences were found to be statistically significant. Since significant male/female differences have been found by the authors in other Hunter College studies, this result may have been due to the small number of males (14) compared to the number of females (53) in the study. [Note that Hunter College has a student population that is approximately 75% female.] In this paper, male/female differences are not discussed further. No significant age-dependent results were found either.

6. THE OVERALL EFFECT OF WRITING COMPLEXITY

Since the two different versions of each problem were identical in mathematical content and the groups of students given the "C" and "E" test banks were alike in backgrounds, differences in the test scores for the two versions of the individual questions could be ascribed to the differences in the remaining factor, writing complexity. The student test score means, grouped by test bank and question, are displayed in Table 1.

Questions 1, 3 and 5 all resulted in differences ***in the direction predicted*** . Students who received questions with less complex writing scored higher. In questions one, and five, the student "C" group scored higher and in question 3, the students who received bank "E" scored higher. ***In each case***, regardless of which student group received the less complex writing, that group scored higher and the differences were associated with quite small p values. Further strengthening the hypothesis of the study, the difference in test results on the control question, (Q4), was associated with a large p values, and indeed, the "C" and "E" group scores on this question are very similar.

As discussed earlier, the second question was different. The revised version of this question involved both differences in writing complexity and an unclear referent. Additionally, this question had two parts. The first part contained very little English and could be considered a control question. And indeed, the version-to-version difference in the test scores (Q21) was very small. But as suspected, the second part of the question, (Q22), was a different matter. The unclear referent had a devastating, significant impact on the students who received it. Students scored significantly lower even though this was the revised version of the question and was written in "easier" English.

Overall, the results in Table 1 confirm the expectation of the study and are statistically significant. Under the hypothesis, the *i*th less complex reading question has an expected response a_i larger than the more-complex version. Their sum, α, may be interpreted as a "reading effect," which occurs by chance with a probability of only 0.02. Clearly, the more complex writing impacted ***all*** of the students, and by a significant amount.

	Q1	Q2	Q21	Q22	Q3	Q4	Q5
Overall	**15.04**	**15.43**	**3.34**	**12.09**	**16.78**	**10.67**	**14.70**
C	Revised 16.00	Standard 16.66	Standard 3.44	Standard 13.22	Standard 16.09	Control 10.22	Revised 15.59
E	Standard 14.17	Revised 14.31	Revised 3.25	Revised 11.06	Revised 17.40	Control 11.09	Standard 13.89
p value	0.13	0.07	0.59	0.05	0.14	0.51	0.14

Table 1: Question Means by Test Bank--All Students

7. THE DIFFERENTIAL EFFECT OF READING ABILITY

An analysis of variance of individual student total test scores revealed the interesting result that ***only*** the number of times a student took the reading placement test was a significant factor, ($p = 0.0043$). Students who required two or more attempts to pass the qualifying reading test scored significantly lower than those who passed the reading test the first time. Furthermore, the number of attempts required to pass the mathematics placement test was ***not*** significant. Mathematics deficiencies had apparently been corrected by the required mastery-based remedial course, but reading deficiencies had not.

One well may believe that students lacking reading or mathematics "ability"--those having to take placement exams twice--are simply less capable than others. But this is not a correct assumption. Mathematics skills required as prerequisite for statistics courses are easy to identify and are taught in the remedial mathematics courses required for students failing the mathematics test once. Once these skills have been acquired, they are practiced again and again in the statistics course itself. On the other hand, reading skills required for the statistics course are not necessarily a part of the reading instruction of the remedial reading courses. The specific statistical vocabulary may well be taught, but the accompanying verbiage can range from straightforward, clear language to very technical, highly convoluted sentences and ideas. So, a student who can comprehend material that is written on an eleventh grade level, an assumption made reasonable by the passing level on the reading exam, cannot necessarily comprehend material that is not specifically written for students on that level. The mathematics needed for statistics is set; the reading needed for statistics is not monitored or graded. For this reason, students entering Hunter College with reading deficiencies are much more at risk than students with mathematics deficiencies. It seems entirely reasonable to expect that similar difficulties appear at many colleges and universities.

The mean of total student test scores (total of all five questions), classified by the number of attempts at the reading (R1 or R2) and mathematics (M1 or M2) placement tests, are shown in Table 2. A "1" in the designation indicates that the student exempted the remedial courses. A "2" in the designation indicates that the student took the remedial course and then retook the placement exam until they passed it. There are a number of important points. First, the marginal effect due to reading background shows a major, statistically significant, drop from 86.84 to 66.98, a twenty point decline (nearly 30%) for those students with lesser reading ability. In contrast, the marginal mathematics difference, while ordered as expected, is not nearly as large and is not statistically significant.

The conditional differences are also interesting. For M1 and M2 students separately, the differences between R1 and R2 readers are large and are significant. But for R1 and R2 students separately, the results have a different character; specifically, for R1 students alone, the difference between M1 and M2 students is not significant, while this same difference is significant for R2 students. Students in the M2/R2 grouping scored seriously lower on this test. The overall interaction in Table 2 is not significant, but the very large drop for the R2/M2 students is, at least potentially, pedagogically very important.

Since the unclear referent in question 2 was seen to have such a major effect on R2 readers, it is fair to ask if the results of Table 2 depend entirely on it. Consequently, the same analysis was repeated, excluding question 2. The results, which are not included here, are very much the same as those in Table 2. The same differences are significant, and in particular, the effect of the complex writing on the R2/M2 student is statistically significant, although the impact is not quite so dramatic.

	Math 1 (M1)	Math 2 (M2)	
Reading 1 (R1)	85.54	88.14	86.84
Reading 2 (R2)	73.39	60.57	66.98
	79.46	74.36	76.91

Table 2: Mean Student Test Scores: Reading Level by Math Level

The conditional results in Table 2 for R1 and R2 students suggest additional calculations. In particular, the analysis displayed in Table 1 was repeated for the R1 and the R2 readers separately and is shown in Tables 3 and 4. The differences between these two tables are striking.

The R1 students were systematically impacted by the more complex writing and consistently performed better on the less complex questions than the more complex ones. However, the individual p values are noticeably larger than those in the combined results of Table 1, and furthermore, the reading parameter, α, has a probability of a larger value equal to 0.50. So while the group of better readers was apparently systematically impacted by the more complex writing, larger studies of this particular reading group are indicated and are underway at Hunter College.

The impact on the R2 students was much greater; they scored lower than the R1 students on every question, whether the writing was more complex or not. Every mean in Table 4 is lower than the corresponding mean in Table 3. But more importantly, every paired C/E question difference for R2 students is larger than the corresponding difference for the R1 readers, is in the direction hypothesized, and is associated with a smaller p value. Everywhere the R2 readers were more seriously affected by the more complex writing than the R1 students. Finally, for the R2 students, the reading effect, α, had a probability of a larger value of only 0.02.

Within question two, the students who had taken remedial reading were seriously impacted by the unclear referent in the second part of question, Q22. The C/E version difference is much larger than that for the control part of the question. In contrast, the unclear referent did not have the devastating effect on the R1 students that it had on the R2 students.

All in all, the weaker reading students were heavily impacted by the more complex writing and certainly much more heavily than the R1 students.

	Q1	Q2	Q21	Q22	Q3	Q4	Q5
C	Revised 17.33	Standard 17.94	Standard 3.67	Standard 14.27	Standard 17.53	Control 10.47	Revised 16.33
E	Standard 15.95	Revised 17.15	Revised 3.90	Revised 13.75	Revised 18.00	Control 11.50	Standard 16.25
p value	0.23	0.28	0.35	0.33	0.36	0.61	0.48

Table 3: Question Means by Test Bank--R1 Students

	Q1	Q2	Q21	Q22	Q3	Q4	Q5
C	Revised 14.82	Standard 15.53	Standard 3.24	Standard 12.29	Standard 14.82	Control 10.00	Revised 14.94
E	Standard 11.80	Revised 9.87	Revised 2.40	Revised 7.47	Revised 16.60	Control 10.53	Standard 10.73
p value	0.13	0.02	0.16	.002	0.20	0.76	0.05

Table 4: Question Means by Test Bank--R2 Students

8. CONCLUDING DISCUSSION

There are four important factors affecting performance on mathematics "word" problems: the level of mathematics complexity; the level of writing complexity; and the student's proficiency in ***both*** reading and mathematics. In this paper, we have focused on the complexity of the writing used in statistics word problems and the reading backgrounds of incoming students at Hunter College. These factors were separated by giving students pairs of questions containing identical mathematics but different levels of writing complexity.

All students in this study were impacted by the questions that contained the more complex writing and, on average, scored lower on them. The questions which were rewritten to have less complex writing led to better mathematics test results for all students. But weaker readers were significantly impacted and scored about 25% lower than the better readers.

Since the number of students in college with English as a second language (ESL) is growing, the importance of the interrelationships between mathematics and language is also growing. Also, other students, particularly urban students, often lack English proficiency. Would these students fare better if a higher level of reading competence were obtained prior to taking mathematics courses? This study suggests they would.

An alternative conclusion is that all students deserve better written text and test materials, particularly in statistics. It is important to remember that the standard problems used in this study were taken directly from available texts, some of which have a more complex writing style than any of the questions used here. Distressingly, some sections in one popular text tested at the grade 22 level! There are indeed much more complex levels of writing complexity than we have used in this study, raising the possibility that at some level of writing complexity almost all students will be impacted. Future reports by the authors will address this point.

Finally, we note that the magnitude of the depressing effects of complex writing on the test results observed in this study are larger than the differences typically reported in studies of the weaker performances of urban students on mathematics tests. Although a more extensive study of the magnitude of reading effects needs to be conducted, the magnitude of the differences observed here suggest that it is possible that these urban students are scoring below average in mathematics and science simply because, for whatever reason, they do not have sufficient reading skills to understand the written materials used in assessing their mathematics skills

BIBLIOGRAPHY

Aiken, L., Jr. (1972), "Language Factors in Learning Mathematics," *Review of Educational Research*, 42, 359-385.

Balow, I. (1964), "Reading and Computational Ability as Determinants of Problem Solving," *The Arithmetic Teacher*, 11, 18-22.

Barnett, J. (1984), “Syntax Variables and Reading Difficulty ,” *Task Variables in Mathematical Problem Solving*, in G.A. Goldin and C.E. McClintock, eds., The Franklin Institute Press, (p. 25-29).

Barney, L. (1972), "Problems Associated With the Reading of Arithmetic," *The Arithmetic Teacher*, 19, 131-133.

Barnitz, J. (1979), "Reading Comprehension of Pronoun-Referent Structures by Children in Grades Two, Four, and Six," *Technical Report No. 117*. Bolt, Beranek and Newman, Inc., Center for the Study of Reading. Cambridge, Mass.

Bell, S., Swan, M., and Taylor, G. (1981), "Choice of Operation in Verbal Problems With Decimal Numbers," *Educational Studies in Mathematics,* 12, 399-420.

Berglund-Gray, G. and Young, R. (1940), "The Effects of Process Sequence on the Interpretation of Two-Step Problems in Arithmetic," *Journal of Educational Research*, 34, 21-29.

Brownell, W., and Stretch, L. (1931), "The effect of unfamiliar settings on problem-solving," *Duke University Research Studies in Education (No. 1)*. Durham, N.C.: Duke University Press.

Bunde, G. (1975), "An Effectiveness Evaluation between Manual and Automated Readability Counting Techniques," *CNETS Report 5-75,* Naval Education and Training command, Pensacola, Fla.

Burns, P. and Yonnally, J. (1964), "Does the Order of Presentation of Numerical Data in Multi-Steps Arithmetic Problems Affect Their Difficulty?" *School Science and Mathematics*, 64, 267-270.

Caldwell, J., and Goldin, G. (1987), "Variables Affecting Word Problem Difficulty in Secondary School Mathematics," *Journal for Research in Mathematics Education*, 18, 187-196.

Cherry, L. (1982), "Writing Tools." *IEEE Transactions on Communications*, 30.

Cherry, L., Fox, M., Frase, L.., Gingrich, P., Keenan, S..., and Macdonald, N.. (May/June 1983), "UNIX Writer's Workbench Software: Computer Aids for Text Analysis," *Bell Laboratories Record*, 10-16.

Cohen, S.A., and Stover, G. (1981), "Effects of teaching sixth-grade students to modify format variables of math word problems," *Reading Research Quarterly*, 16, 175-200.

Cramer, E. (1978), "A Quick Guide to Readability Formulas," *Curriculum Review* v17, P. 416.

Dale, E., and Chall, J.S. (1948), "A Formula for Predicting Readability. Pt. I," *Educational Research Bulletin*, 27, 11-20.

Dale, E., and Chall, J.S. (1948), "A Formula for Predicting Readability. Pt. II," *Educational Research Bulletin*, 28, 37-54.

DeCorte, E., and Verschaffel, L., and Verschueren, J. (1982), "First graders' solution processes in elementary word problems," In A. Vermandel (Ed.), *Proceedings of the Sixth International Conference for the Psychology of Mathematical Education* (pp. 91-96). Antwerp: Universitaire Instelling Antwerpen.

DeCorte, E., and Verschaffel, L. (1987), "Word Problems in the Primary Grades: Implications of Recent Research," Invited address at the 65th Annual Meeting of the National Council of Teachers of Mathematics.

Dutka, J. (1979), "Anaphoric Relations, Comprehension and Readability," Paper presented at the Annual International Conference on Processing of Visible Language.

Fischbein, E., Deri, M., Nello, M.S., and Marino, M.S. (1985), "The Role of Implicit Models in Solving Verbal Problems in Multiplication and Division," *Journal for Research in Mathematics Education*, 16, 3-17.

Fry, E. (1977), "Fry's readability graph: Clarification, validity, and extension to level 17.," J*ournal of Reading*. 21, 242-252.

Fry, E. (1968), "A Readability Formula That Saves Time," *Journal of Reading* 11, 513-516.

Glennon, V.J., and Callahan, L.G. (1968), *Elementary school mathematics: A guide to current research (3rd ed.)*. Washington, D.C.: Association for Supervision and Curriculum Development, 94-95.

Harris, A., and Jacobson, M. (1974), "Some Comparisons Between the Basic Elementary Reading Vocabularies and Other Word Lists.," *Reading Research Quarterly* 9, 87-109.

Henney, M. (1971), "Improving Mathematics Verbal Problem-Solving Ability Through Reading Instruction," *The Arithmetic Teacher*, 18, 223-229.

Hiebert, J. (1982), "The Position of the Unknown Set and Children's Solutions of Verbal Arithmetic Problems," *Journal for Research in Mathematics Education*, 13, 341-349.

Houlihan, D., and Ginsburg, H. (1981), "The Addition Methods of First- and Second-Grade Children," *Journal for Research in Mathematics Education*, 12, 95-106.

Jerman, M. (1973), "Problem Length as a Structural Variable in Verbal Arithmetic Problems," *Educational Studies in Mathematics*, 5, 109-123.

Jerman, M., and Mirman, S. (1974), "Linguistic and Computational Variables in Problem Solving in Elementary Mathematics," *Educational Studies in Mathematics*, 5, 317-362.

Jerman, M., and Rees, R. (1972), "Predicting the Relative Difficulty of Verbal Arithmetic Problems," *Educational Studies in Mathematics*, 4, 306-323.

Kane, R.B. (1970), "The readability of mathematical textbooks: revisited," *Mathematics Teacher*, 1970, 63, 579-581.

Kent, T. (1980), "Readability and Audience Response: Unfogging the Fog Indexes," Paper presented at the Annual Meeting of the conference on College Composition and Communication. Washington, DC.

Knifong, J.D., and Holtan, B. (1977), "A search for reading difficulties among erred word problems," *Journal for Research in Mathematics Education*, 8, 227-230.

Knifong, J.D., and Holtan, B. (1976), "An analysis of children's written solutions to word problems," *Journal for Research in Mathematics Education,* 7 106-112.

Linville, W. (1969)*The effects of syntax and vocabulary upon the difficulty of verbal arithmetic problems with fourth grade students*. Unpublished doctoral dissertation, Indiana University.

Llabre, M.; and Cuevas, G. (1983), "The Effects of Test Language and Mathematical Skills Assessed on the Scores of Bilingual Hispanic Students," *Journal for Research in Mathematics Education*, 14, 318-324.

Lyda, W., and Church, R. (1964), "Direct, Practical Arithmetic Experience and Success in Solving Realistic Verbal 'Reasoning' Problems in Arithmetic," *Journal of Education Research*, 57, 530-533.

Mestre, J. and Gerace, W. (1981), "The Interdependence of Mathematical Skills, Grade Point Average, and Language Proficiency for Hispanic College Students," Paper presented at the Remedial and Developmental Mathematics in College: Issues and Innovations at CUNY. New York.

Nesher, P. and Teubal, E. (1975), "Verbal Cues on an Interfering Factor in Verbal Problem Solving," *Educational Studies in Mathematics*, 6, 41-51.

Nesher, P. (1982), "Levels of description in the analysis of addition and subtraction word problems," In T.P. Carpenter, J.M. Moser, and T.A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective*. Hillsdale, N.J.: Erlbaum 25-38.

Nesher, P. (1976), "Three determinants of difficulty in verbal arithmetic problems," *Educational Studies in Mathematics* 7, 369-388.

Paul, D., Nibbelink, W., and Hoover, H. (1986), "The Effects of Adjusting Readability on the Difficulty of Mathematics Study Problems," *Journal for Research in Mathematics Education*, 17, 163-171.

Riley, M.S., Greeno, J.G., and Heller, J.I. (1983), "Development of children's problem-solving ability in arithmetic" In H.P. Ginsburg (Ed.), *The development of mathematical thinking. New York,* Academic Press 153-196.

Searle, B., Lorton, P., and Suppes, P. (1974), "Structural Variables Affecting CAI Performance on Arithmetic Word Problems of Disadvantaged and Deaf Students," *Educational Studies in Mathematics*, 5, 371-384.

Segalla, A. (1973), "Using Structural Variables to Predict Word Problem Solving Difficulty for Junior College Arithmetic Students," paper presented at the American Educational Research Association annual meeting, New Orleans.

Sherard, W., III. (1974), *The effect of arithmetical operations on the difficulty levels of verbal problems*. Unpublished doctoral dissertation, George Peabody College for Teachers.

Smye, R. (1987), "Style and Usage Software: Mentor, not Judge," paper presented at the Annual Meeting of the Conference on College Composition and Communication. Atlanta.

Spache, G. (1953), "A New Readability Formula for Primary-Grade Reading Materials," *Elementary School Journal*, 53, 410-413.

Steffe, L. (1967), "The effects of two variables on the problem-solving abilities of first grade children," *Teaching report no 21*, Madison, Wis: Wisconsin Research and Development Center for Cognitive Learning, University of Wisconsin, 1967.

Sullivan, J. (1980), "The Power of the Flesch Reading Ease Formula," In Bruno, S. (ed.), *Unfurling Communication's Colors in the 80's*. Urbana, IL.

Suppes, P., Loftus, E., and Jerman, M. (1972), "Problem solving on a computer based teletype," *Educational Studies in Mathematics*, 4, 306-323.

Taylor, W.L. (1953), "Cloze procedures: A new tool for measuring readability." *Journalism Quarterly*, 1953, 30, 414-438.

Thompson, E. (1967), *Readability and accessory remarks: Factors in problem solving in arithmetic,* Unpublished doctoral dissertation, Stanford University.

VanderLinde, L. (1974), "Does the Study of Quantitative Vocabulary Improve Problem Solving?," *Elementary School Journal*, 65, 143-152.

West, T.A. (1977), "Rx for verbal problems: A diagnostic-prescriptive approach," *The Arithmetic Teacher*, 25(2), 57-58.

Whitlock, P. *An investigation of selected factors that affect ability to solve verbal mathematical problems at the primary level*. Unpublished doctoral dissertation, Fordham University, 1974.

Williams, M., and McCreight, R. (1965), "Shall We Move the Question?" *The Arithmetic Teacher, 12*, 418-421.

APPENDIX: THE TEST QUESTIONS AND THEIR READABILITY

1C. [Revised] A college official thinks that students begin college at an earlier age than they used to. Two years ago, the average age was 18.75 years. He records the ages of a random sample of 40 students in this year's class. The average age is 18.36 years and the standard deviation 0.41 year.

(a) Define μ in this case.

(b) State H_0 and H_1.

(c) What decision should be made?

(d) Obtain an 90% confidence interval for μ.

1E. [Standard] An educational planner suspects there has been a significant decline in the mean age at which students begin college. Complete figures for two years ago indicate a mean age of 18.75 years. A random sample of 40 student records is selected from this year's freshman class at the state university. The mean age at orientation is 18.36 years and the standard deviation 0.41 year.

(a) Define μ in this case.

(b) State H_0 and H_1.

(c) What decision should be made?

(d) Obtain an 90% confidence interval for μ.

2C. [Standard] The owner of a chicken farm is interested in using a new chicken feed which it is claimed will significantly improve the gain in weight of chicks. Before investing heavily in the new and more expensive feed, she plans to test it out on 64 two-week-old chicks. She knows that the traditional feed produces a mean gain of 21 grams with a standard deviation of 2.4 grams between the second and third weeks of life.

(a) State the appropriate null and alternative hypothesis, defining μ.

(b) What would you advise the owner if the mean gain in weight for the 64 chicks during the week of the test is:

(i) 
(ii) 
(iii) 
(iv) 

2E. [Revised] A chicken farmer wants her chickens to gain weight more quickly. A new feed claims to do this. Before buying much of the new and more expensive feed, she plans to test it out on 64 two-week-old chicks. She knows that chicks eating the old feed gain a mean of 21 grams with a standard deviation of 2.4 grams from the second to the third weeks of life.

(a) State the appropriate null and alternative hypothesis, defining μ.

(b) What would you advise the owner if the mean gain in weight for the 64 chicks during the week of the test is:

(i) 
(ii) 
(iii) 
(iv) 

3C. [Standard] A tobacco manufacturer claims that the average amount of tar in his cigarettes is only 9.4 milligrams. A consumer group believes that this figure underestimates the true mean amount of tar. A random sample of n cigarettes of this brand are selected and the amount of tar per cigarette is determined. The sample mean and the standard deviation of the amount of tar are ☒ = 9.505 and s = .533 milligram.

(a) Define μ in this case.

(b) State H_0 and H_1.

(c) What conclusion should be reached if the data above were based on (i) n =40; (ii) n =80; (iii) n =200 cigarettes?

(d) Explain how your conclusion depends on the sample size, n.

3E. [Revised] A cigarette company says that the average amount of tar in their cigarettes is 9.4 milligrams. A consumer group thinks that this number is too low. They select a random sample of n cigarettes of this brand. They determine the amount of tar per cigarette. The sample mean of the amount of tar is ☒ = 9.505 milligrams. The standard deviation is s = .533 milligram.

(a) Define μ in this case.

(b) State H_0 and H_1.

(c) What conclusion should be reached if the data above were based on (i) n =40; (ii) n =80; (iii) n =200 cigarettes?

(d) Explain how your conclusion depends on the sample size, n.

5C. [Revised] A psychologist thinks that older children can complete a task faster than younger children. He wants to compare the variability of the times for the two groups. He asked a group of 12 five-year-olds and 20 six-year-olds to complete a task. Their times (in minutes) were:

		n		s
FIVE-YEAR-OLDS	I	12	15.8	3.9
SIX-YEAR-OLDS	II	20	10.2	2.8

(a) Compute a 90% confidence interval for μ_1, the mean of the 5-year-olds.

(b) Compute a 90% confidence interval for $\mu_{2,}$ the mean of the 6-year-olds.

(c) Do the two confidence intervals that you computed in parts (a) and (b) above tell you anything about the difference between the 5-year-olds and the 6-year-olds?

5E. [Standard] A psychologist is interested in learning how quickly six-year-olds complete a task as compared to five-year-olds. He is sure that on average the older children will take less time than the younger but is interested in comparing the two groups. A group of 12 five-year-olds and 20 six-year-olds were asked to complete the task, with the following results (in minutes):

		n		s
FIVE-YEAR-OLDS	I	12	15.8	3.9
SIX-YEAR-OLDS	II	20	10.2	2.8

(a) Compute a 90% confidence interval for μ_1, the mean of the 5-year-olds.

(b) Compute a 90% confidence interval for $\mu_{2,}$ the mean of the 6-year-olds.

(c) Do the two confidence intervals that you computed in parts (a) and (b) above tell you anything about the difference between the 5-year-olds and the 6-year-olds?

*Sandra Pryor Clarkson is an Associate Professor and William H. Williams is a Professor in the Department of Mathematics and Statistics, Hunter College of CUNY, 695 Park Avenue, New York, New York 10021. This work was partially supported by a PSC-CUNY grant.