The Role of Rewording and Anomalous Information in the Comprehension and Solution of Statistics Word Problems

Queena N. Lee-Chua
Ateneo de Manila University
Mathematics Department
Loyola Heights, Quezon City
Philippines

Abstract:

Many studies have shown that certain concepts in statistics lend themselves to fallacies, and that students often tend to rely more on fuzzy ideas rather than on mathematically rigorous interpretations of these concepts. This study investigates whether anomalous information causes an increase in questions generated by college students while they solve statistics word problems. 40 science undergraduates are presented with different versions of each of 10 statistics problems: (a) original, (b) deletion of critical information, (c) addition of contradictory information, (d) addition of salient irrelevancies, and (e) addition of subtle irrelevancies. Results show that most of the transformed versions trigger more questions than the original (p < 0.01), with the deletion versions triggering more questions than do others. The role of rewording word problems is discussed in light of these findings, and cognitive models (such as the obstacle hypothesis) are used to explain pertinent differences. Implications of these findings for pedagogy are discussed, and recommendations for research conclude the paper.
Introduction

The study of statistics is double-edged: on the one hand, it seems deceptively close to everyday experience, where terms such as “randomness,” “chance,” “equally likely,” and so forth are frequently used by laypeople without apparent regard for their statistical definitions (Kosonen & Winne, 1995). The media, ratings boards, athletes, politicians, and others often loosely use these terms to justify one action over another, and in doing so, they fall prey to fallacies, including widely documented ones such as the representativeness heuristic (Kahneman & Tversky, 1973) and the outcome approach (Konold, 1989). These errors “betray the naivete of stochastic intuitions” (Pratt, 1998, p. 3) among many students and educated adults, and are the focus of much research in mathematics education and cognitive psychology.

On the other hand, the mathematical idea of statistical events is a more rigorous one. Terms are well-defined, rules formulated and proven, and reason preferred over intuition. Yet because of this dual nature of stochastics, students often find the topic difficult. “Thinking mathematically demands more. It presupposes that the learner has a more or less rich pool of intellectual tools at their [sic] disposal…[which are] available to a mathematician—and precisely those lacking in the naïve learner’ (Pratt, 1998, p. 2).

Undergraduate students majoring in the natural sciences are assumed to have more sophisticated mathematical tools than their humanities and social science counterparts, but they also fall prey to stochastic misconceptions. Lee-Chua (2000) has documented some instances in the Philippines. In introductory probability courses, most of the third-year science students manage to solve “simple” word problems (with one or at most two pieces of information, which can be directly inputted into a formula or equation). For
instance, they can readily solve a problem such as “A truth serum given to a suspect is known to be 90% reliable when the person is guilty. If you are guilty and you are given the serum, what is the probability that you will go free?” (From knowledge of the formula, or “intuition,” as the students claim, they will usually answer: \(1 - 0.9 = 0.1\), or 10%).

However, when additional pieces of information are given, and the problem becomes more complicated, they tend to make several errors (Lee-Chua, 2000). These deficiencies become especially evident when students attempt to comprehend and solve more complicated word problems. Why is this so? “Solving [statistics] problems calls for more than simple transfer of knowledge; it requires genuine problem solving” (Kosonen & Winne, 1995, p. 33).

Rewording and Anomalous Information

Researchers have discovered several ways to avoid errors and enhance comprehension, including using analogies (Fast, 1997), structure-emphasizing examples (Quilici & Mayer, 1996), and high-explanatory texts (Myers et al., 1983). Another method is to encourage students to generate salient questions, which plays a significant role in rewording the problem itself. The renowned psychologist and educator Jean Piaget has long posited the benefits of active learning and problem solving, including rewording (1952). Through the years, cognitive psychologists have discovered that the latter is essential to several cognitive processes such as learning of complex material, understanding of text, and solving of nonroutine word problems (Miyake & Norman, 1979).
Let us discuss the importance of rewording in the latter case—the solution of word problems. Solution first requires comprehension, and comprehension in turn requires the ability to construct from the worded problem a meaningful conceptual representation (which many cognitive psychologists have termed a “schema”) upon which problem-solving strategies may take place. “The successful problem solver is thus able to build an accurate mental representation that intervenes between the problem text and application of the solution strategies” (Davis-Dorsey et al., 1991, p. 61).

However, this is easier said than done. Individual differences in schemata influence how information is absorbed, translated, and comprehended, thus students faced with the same word problem may possess different interpretations of it. A salient task in this case is to more clearly define the interrelations of elements in their conceptual frameworks, which is the main role of rewording (Davis-Dorsey, et al., 1991). Thus, problem rewording exists to make more explicit the various interrelations among the pieces of information presented in a problem.

Through strategic rewording of word problems, teachers can encourage students to generate questions. Several other methods have been found to encourage question generation, from social strategies such as providing a comfortable climate in the classroom for lively question-and-answer sessions (Van der Meij, 1988) to cognitive strategies such as identifying knowledge deficits (Otero & Campanario, 1990). An effective method, which is used in this paper to trigger questions, is rewording problems to employ anomalous information (Graessser & McMahen, 1993; Schank, 1986).

Schank’s anomaly model (1986) is particularly salient—he posits that normally each learner has built a huge storehouse of explained events in long-term memory through the
years. When the learners encounter an unexplained (anomalous) event, they tend to generate questions which would lead them to comprehend it. Once they are able to make sense of the problem, it ceases to become an anomaly, and new learning occurs.

This theory of learning through question generation of anomalous events is tempting for many cognitive psychologists and mathematics educators, who have used it in making sense of learning deficits (Miyake & Norman, 1979). Yet “empirical tests [of this theory] are conspicuously lacking” (Graesser & McMahan, 1993, p. 137). Graesser and McMahan (1993) have tried to address this deficiency through the utilization of reading fables and solving elementary algebra problems, and Lee-Chua (2000) has done investigations of this in probability word problems. However, empirical investigations of Schank’s model of learning in other fields (such as statistics) are sorely lacking.

Thus, this study was designed to further address such deficiencies. Following parts of Graesser and McMahan’s (1993) method, the author modified their experiment to suit the classroom setting. In particular, this study investigated the views of students majoring in mathematics and computer science at the Ateneo de Manila University when they encountered anomalous information in statistics word problems. Questions that guided this inquiry were:

1. Do students generate questions when they encounter anomalous information in statistics word problems?
2. What type of anomalous information triggers the most questions? Why?
3. How does rewording facilitate comprehension of statistics word problems?
4. What implications does this investigation hold for pedagogy and research?
Note that this study is limited to an investigation of cognitive factors in the comprehension and solution of word problems. Further experiments need to be done to assess the significance of social factors.

Methodology

Sample

Since this study centers on cognitive factors, the intrusion of social factors is kept to a minimum. This necessitates, among others, that the subjects be enrolled in the same field of study (mathematics and computer science) and handled by the same professor, in this case, the author. 40 students participated in the study, and all were in the third year of a four-year degree program.

Instrumentation

Ten quantitative statistics problems were selected from textbooks (e.g., Walpole, 1982). The problems were word problems that required a conceptual understanding of basic statistics concepts, such as the notion of randomness, sampling distributions, estimation, and hypothesis-testing. The problems emphasized reasoning and inferring rather than mechanical “plugging and number-crunching.”

For instance, consider the following problem from Walpole (1982, p. 232). This is hereafter referred to as the “original version”:

“A manufacturing firm claims that the batteries used in their electronic games will last an average of 30 hours. To maintain this average, 16 batteries are tested each month. If the computed t value falls between –t_{0.025} and t_{0.025}, the firm is satisfied with its claim. What conclusion should the firm draw from a sample that has a mean of 27.5 hours and a standard deviation of 5 hours? Assume the distribution of battery lives to be
approximately normal.” (The solution here requires knowledge of the sampling
distribution of means.)

Anomalous information is presented in the following manner. Aside from the original
version, there were four other versions of each of the ten problems, which are defined as
follows:

1. Deletion version: This involves removing a critical piece of information, which
   makes it impossible for the student to solve the problem. For instance, when the
   information “the sample has a mean of 27.5 hours and a standard deviation of 5
   hours” is replaced by “the sample has a high mean and a low standard deviation,”
   then the problem cannot be solved.

2. Contradiction version: This involves adding a statement that directly contradicts
   one or more of the premises in the original problem. For instance, when the clause
   “and at the same time the computed t value is greater than $t_{0.10}$” is added to the original
   condition, then a contradiction occurs.

3. Salient irrelevancy version: This involves adding a statement that is both
   quantitatively irrelevant to the solution of the problem and thematically irrelevant to
   the semantic context. For instance, a salient irrelevant piece of information would be
   “the manufacturer also makes other sorts of batteries, which last approximately 100
   hours.”

4. Subtle irrelevancy version: This involves adding a statement that is quantitatively
   irrelevant to the solution of the problem, but \emph{thematically relevant} to the semantic
   content. A subtle irrelevancy is often harder to detect than a salient irrelevancy. For
   instance, a subtle irrelevant piece of information would be “and last month, 13
batteries passed the test” added to the original information “16 batteries are tested each month.”

The five different versions of this sample statistics word problem are presented in Table 1. The various versions of four other problems (with notes on the transformations) are presented in Tables 2 to 5.

For each of the ten problems, there are four transformed versions. Two raters were trained in the transformations, and they had perfect agreement on them.

Each subject received a ten-page booklet with one version of each problem on each page. The assignment of versions was counterbalanced across subjects so that for each problem, eight subjects were assigned to each of the five conditions. The order in which the ten problems was presented in each booklet was randomized for each subject.

Procedure

The subjects were instructed that they would be generating questions while solving word problems in statistics. Specifically, they were told to read and solve the problems. They were further instructed to write down any questions which came to mind while they were attempting to solve the problems. To further encourage question generation, no time limits were given for the task.
Table 1
Five Different Versions of a Statistics Word Problem on Means

<table>
<thead>
<tr>
<th>Version</th>
<th>Word Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>A manufacturing firm claims that the batteries used in their electronic games will last an average of 30 hours. To maintain this average, 16 batteries are tested each month. If the computed $t$ value falls between $-t_{0.025}$ and $t_{0.025}$, the firm is satisfied with its claim. What conclusion should the firm draw from a sample that has a mean of 27.5 hours and a standard deviation of 5 hours? Assume the distribution of battery lives to be approximately normal.</td>
</tr>
<tr>
<td>Deletion</td>
<td>A manufacturing firm claims that the batteries used in their electronic games will last an average of 30 hours. To maintain this average, 16 batteries are tested each month. If the computed $t$ value falls between $-t_{0.025}$ and $t_{0.025}$, the firm is satisfied with its claim. What conclusion should the firm draw from a sample that has a high mean and a low standard deviation? Assume the distribution of battery lives to be approximately normal.</td>
</tr>
<tr>
<td>Contradiction</td>
<td>A manufacturing firm claims that the batteries used in their electronic games will last an average of 30 hours. To maintain this average, 16 batteries are tested each month. If the computed $t$ value falls between $-t_{0.025}$ and $t_{0.025}$ and at the same time is greater than $t_{0.10}$, the firm is satisfied with its claim. What conclusion should the firm draw from a sample that has a mean of 27.5 hours and a standard deviation of 5 hours? Assume the distribution of battery lives to be approximately normal.</td>
</tr>
<tr>
<td>Salient Irrelevancy</td>
<td>A manufacturing firm claims that the batteries used in their electronic games will last an average of 30 hours. To maintain this average, 16 batteries are tested each month. If the computed $t$ value falls between $-t_{0.025}$ and $t_{0.025}$, the firm is satisfied with its claim. The manufacturer also makes other sorts of batteries, which last approximately 100 hours. What conclusion should the firm draw from a sample that has a mean of 27.5 hours and a standard deviation of 5 hours? Assume the distribution of battery lives to be approximately normal.</td>
</tr>
<tr>
<td>Subtle Irrelevancy</td>
<td>A manufacturing firm claims that the batteries used in their electronic games will last an average of 30 hours. To maintain this average, 16 batteries are tested each month. If the computed $t$ value falls between $-t_{0.025}$ and $t_{0.025}$, the firm is satisfied with its claim. What conclusion should the firm draw from a sample that has a mean of 27.5 hours and a standard deviation of 5 hours? Assume the distribution of battery lives to be approximately normal.</td>
</tr>
</tbody>
</table>
Table 2

Five Different Versions of a Statistics Word Problem on the Normal Distribution

<table>
<thead>
<tr>
<th>Version</th>
<th>Word Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>The average grade in a statistics exam was 82 and the standard deviation was 5. All students with grades from 88 to 94 received a grade of B. If the grades are approximately normally distributed and 8 students received a B grade, how many students took the exam?</td>
</tr>
<tr>
<td>Deletion</td>
<td>The average grade in a statistics exam was 82. All students with grades from 88 to 94 received a grade of B. If the grades are approximately normally distributed and 8 students received a B grade, how many students took the exam? (The clause “the standard deviation was 5” is deleted.)</td>
</tr>
<tr>
<td>Contradiction</td>
<td>The average grade in a statistics exam was 82 and the standard deviation was 5. All students with grades from 88 to 94 received a grade of B, those with 95 to 100 received a grade of A, and those below 87 a grade of C. If the grades are approximately normally distributed and 10 students received an A, 8 students received a B, and 5 students a C, how many students took the exam? (The data on the number of students receiving the three different grades contradict the range described, and the fact that the grades are normally distributed.)</td>
</tr>
<tr>
<td>Salient Irrelevancy</td>
<td>The average grade in a statistics exam was 82 and the standard deviation was 5. All students with grades from 88 to 94 received a grade of B. If the grades are approximately normally distributed, 8 students received a B grade, and half of them were satisfied with their grade, how many students took the exam? (The clause “half of them were satisfied with their grade” is a salient irrelevancy.)</td>
</tr>
<tr>
<td>Subtle Irrelevancy</td>
<td>The average grade in a statistics exam was 82 and the standard deviation was 5. All students with grades from 88 to 94 received a grade of B. If the grades are approximately normally distributed and 8 students received a B grade, and fewer students received an A, how many students took the exam? (The clause “fewer students received an A” is a subtle irrelevancy.)</td>
</tr>
</tbody>
</table>
# Table 3

## Five Different Versions of a Statistics Word Problem on Bayes’ Rule

<table>
<thead>
<tr>
<th>Version</th>
<th>Word Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Original</strong></td>
<td>A truth serum given to a suspect is known to be 90% reliable when the person is guilty and 99% reliable when the person is innocent. If the suspect was selected from a group of suspects of which only 5% have ever committed a crime, and the serum indicates that he is guilty, what is the probability that he is innocent?</td>
</tr>
<tr>
<td><strong>Deletion</strong></td>
<td>A truth serum given to a suspect is known to be 90% reliable when the person is guilty and 99% reliable when the person is innocent. If the suspect was selected from a group of suspects only a few of which have ever committed a crime, and the serum indicates that he is guilty, what is the probability that he is innocent? (The original information of 5% committing a crime is changed to “only a few have committed a crime.”)</td>
</tr>
<tr>
<td><strong>Contradiction</strong></td>
<td>A truth serum given to a suspect is known to be 90% reliable when the person is guilty and 99% reliable when the person is innocent. Furthermore, 1% of the guilty are judged innocent by the serum. If the suspect was selected from a group of suspects of which only 5% have ever committed a crime, and the serum indicates that he is guilty, what is the probability that he is innocent? (The clause “1% of the guilty are judged innocent by the serum” contradicts the original previously stated information.)</td>
</tr>
<tr>
<td><strong>Salient Irrelevancy</strong></td>
<td>A truth serum given to a suspect is known to be 90% reliable when the person is guilty and 99% reliable when the person is innocent. If the suspect was selected from a group of suspects of which only 5% have ever committed a crime, 25% of the subjects are scared to die, and the serum indicates that he is guilty, what is the probability that he is innocent? (The clause “25% of the suspects are scared to die” is a salient irrelevancy.)</td>
</tr>
<tr>
<td><strong>Subtle Irrelevancy</strong></td>
<td>A truth serum given to a suspect is known to be 90% reliable when the person is guilty and 99% reliable when the person is innocent. If the suspect was selected from a group of suspects of which only 5% have ever committed a crime, and only 1% have ever been found guilty. If the serum indicates that he is guilty, what is the probability that he is innocent? (The clause “only 1% have ever been found guilty” is a subtle irrelevancy.)</td>
</tr>
</tbody>
</table>
### Table 4

Five Different Versions of a Statistics Word Problem on Proportions

<table>
<thead>
<tr>
<th>Version</th>
<th>Word Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>A vote is to be taken among the residents of a town and a suburb to determine whether a mall is to be constructed. The proposed site is within town limits, so many voters in the suburb feel that the proposal will pass because of the large proportion of town voters in favor. To determine if there is a significant difference in the proportion of town voters and suburban voters favoring the proposal, a poll is taken. If 120 of 200 town voters favor the proposal and 240 of 500 suburban residents favor it, would you agree that the proportion of town voters favoring the proposal is higher than the proportion of county voters? Use a 0.025 level of significance.</td>
</tr>
<tr>
<td>Deletion</td>
<td>A vote is to be taken among the residents of a town and a suburb to determine whether a mall is to be constructed. The proposed site is within town limits, so many voters in the suburb feel that the proposal will pass because of the large proportion of town voters in favor. To determine if there is a significant difference in the proportion of town voters and suburban voters favoring the proposal, a poll is taken. If 120 of 200 town voters favor the proposal, would you agree that the proportion of town voters favoring the proposal is higher than the proportion of county voters? Use a 0.025 level of significance. (The clause “240 of 500 suburban residents favor it” is deleted.)</td>
</tr>
<tr>
<td>Contradiction</td>
<td>A vote is to be taken among the residents of a town and a suburb to determine whether a mall is to be constructed. The proposed site is within town limits, so many voters in the suburb feel that the proposal will pass because of the large proportion of town voters in favor. To determine if there is a significant difference in the proportion of town voters and suburban voters favoring the proposal, a poll is taken. If 240 of 500 town voters favor the proposal and 120 of 200 suburban residents favor it, would you agree that the proportion of town voters favoring the proposal is higher than the proportion of county voters? Use a 0.025 level of significance. (The original proportions of town and suburban voters are interchanged.)</td>
</tr>
<tr>
<td>Salient Irrelevancy</td>
<td>A vote is to be taken among the residents of a town and a suburb to determine whether a mall is to be constructed. The mall is supposed to provide employment to 80 town and 50 suburban residents. The proposed site is within town limits, so many voters in the suburb feel that the proposal will pass because of the large proportion of town voters in favor. To determine if there is a significant difference in the proportion of town voters and suburban voters favoring the proposal, a poll is taken. If 240 of 500 town voters favor the proposal and 120 of</td>
</tr>
</tbody>
</table>
A vote is to be taken among the residents of a town and a suburb to determine whether a mall is to be constructed. The proposed site is within town limits, so many voters in the suburb feel that the proposal will pass because of the large proportion of town voters in favor. To determine if there is a significant difference in the proportion of town voters and suburban voters favoring the proposal, a poll is taken. If fewer than half of the residents in each area show up, then the results are void and another poll will be taken. If 120 of 200 town voters favor the proposal and 240 of 500 suburban residents favor it, would you agree that the proportion of town voters favoring the proposal is higher than the proportion of county voters? Use a 0.025 level of significance.

(The sentence “If fewer than half of the residents in each area show up, then the results are void and another poll will be taken” is a subtle irrelevancy.)
assigns to $x$ the value 0 if neither ball is red, the value 1 if the first ball only is red, the value 2 if the second ball only is red, the value 3 if both balls are red, and the value 4 if the first ball is red and the second one black. Using a loss function of the form $L(y; x) = |y - x|$, find the risk function $R(y; x)$.

(The clause “the value 4 if the first ball is red and the second one black” contradicts the assumption for the value of 1.)

**Salient Irrelevancy**

Suppose that an urn contains 3 balls, of which $x$ are red and the remainder black. There are more balls with a radius of 5 inches than 2 inches. We wish to estimate $x$ by selecting two balls in succession without replacement. Let $y$ be the decision function that assigns to $x$ the value 0 if neither ball is red, the value 1 if the first ball only is red, the value 2 if the second ball only is red, and the value 3 if both balls are red. Using a loss function of the form $L(y; x) = |y - x|$, find the risk function $R(y; x)$.

(The clause “There are more balls with a radius of 5 inches than 2 inches” is a salient irrelevancy.)

**Subtle Irrelevancy**

Suppose that an urn contains 3 balls, of which $x$ are red and the remainder black, and $x$ is less than 5. We wish to estimate $x$ by selecting two balls in succession without replacement. Let $y$ be the decision function that assigns to $x$ the value 0 if neither ball is red, the value 1 if the first ball only is red, the value 2 if the second ball only is red, and the value 3 if both balls are red. Using a loss function of the form $L(y; x) = |y - x|$, find the risk function $R(y; x)$.

(The clause “$x$ is less than 5” is a subtle irrelevancy.)

**Results**

**Number of Questions Generated**

Though no time limits were set, all subjects finished the ten questions within 90 minutes. The number of questions that each subject generated for each problem was scored. A clause was counted as a question if it contained a query (e.g., “Exactly how high is a high mean and exactly how low is a low standard deviation?”) or a non-interrogative statement pointing out anomalies (e.g., “There is not enough information to solve this problem.”). Two raters were trained to identify questions with a high degree of reliability (Cronbach’s alpha exceeded .90).
The mean number of questions generated per problem were computed. These means and standard deviations are shown in Table 6. As predicted by the anomaly model, the mean number of questions generated in the four transformed versions combined (2.62) was (highly) significantly greater than the number generated in the original condition (1.82), as measured by t-test analyses, p < 0.01. Pairwise comparisons of each transformed version indicated all the four versions differed significantly from the original version, p < 0.01 or p < 0.05. Moreover, the deletion version generated significantly more questions than the three other transformations. As for these three, only the subtle irrelevancy version differed significantly from the two others (p < 0.05 versus p < 0.01 for the rest).

Table 6
Data Analysis of Total Number of Questions Generated While Solving Statistics Word Problems

<table>
<thead>
<tr>
<th>Version</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Probability (Transformed Version vs Original)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>1.82</td>
<td>1.40</td>
<td>------</td>
</tr>
<tr>
<td>Deletion</td>
<td>3.01</td>
<td>1.33</td>
<td>0.01</td>
</tr>
<tr>
<td>Contradiction</td>
<td>2.61</td>
<td>1.35</td>
<td>0.01</td>
</tr>
<tr>
<td>Salient Irrelevancy</td>
<td>2.55</td>
<td>1.39</td>
<td>0.01</td>
</tr>
<tr>
<td>Subtle Irrelevancy</td>
<td>2.29</td>
<td>1.39</td>
<td>0.05</td>
</tr>
<tr>
<td>Combined (4) transformations (without original)</td>
<td>2.62</td>
<td>1.37</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Transformation-Relevant Questions

To investigate these differences further, another analysis was performed. Differences between the number of questions generated by the original version and those generated by the other versions can be further borne out by analyzing transformation-relevant questions for the different versions. A question was considered transformation-relevant if it addressed any of the anomalies (e.g., “Why is the fact that 13 batteries passing the test last month included if it only distracts us from what the problem wants us to solve?”). Two judges were trained to distinguish between transformation-relevant and transformation-irrelevant questions with a high degree of reliability (Cronbach’s alpha exceeded 0.90).

The mean number of transformation-relevant questions generated per version is shown in Table 7. Results are mostly consistent with those of Table 6, except for the subtle-irrelevancy case. The mean number of transformation-relevant questions generated in the four transformed versions combined (2.37) was significantly greater than the mean number of questions generated in the original version (1.82), $p < 0.05$. Pairwise comparisons indicated that the deletion ($p < 0.01$), contradiction ($p < 0.05$) and salient irrelevancy ($p < 0.05$) versions differed significantly from the original version, whereas the subtle irrelevancy version did not. Moreover, the deletion version produced a significantly higher number of transformation-relevant questions (2.82) than either the contradiction (2.35) or salient irrelevancy (2.32) versions.
Table 7
Data Analysis of Transformation-Relevant Questions Generated While Solving Statistics Word Problems

<table>
<thead>
<tr>
<th>Version</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Probability (Transformed Version vs Original)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>1.82</td>
<td>1.40</td>
<td>-----</td>
</tr>
<tr>
<td>Deletion</td>
<td>2.82</td>
<td>1.10</td>
<td>0.01</td>
</tr>
<tr>
<td>Contradiction</td>
<td>2.35</td>
<td>1.25</td>
<td>0.05</td>
</tr>
<tr>
<td>Salient Irrelevancy</td>
<td>2.32</td>
<td>1.21</td>
<td>0.05</td>
</tr>
<tr>
<td>Subtle Irrelevancy</td>
<td>1.99</td>
<td>1.47</td>
<td>Not significant</td>
</tr>
<tr>
<td>Combined (4) transformations</td>
<td>2.37</td>
<td>1.26</td>
<td>0.05</td>
</tr>
<tr>
<td>(without original)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discussion

The Role of Rewording

In generating questions, students are engaged in the process of mental problem rewording. In so doing, they begin to clarify for themselves the nature of the problem, to piece together the various pieces of information given (anomalous or otherwise), and to create a more meaningful problem schema. Successful problem solvers utilize successful rewording procedures (which includes deletion of contradictory information, and the ignoring of salient and subtle irrelevancies).

The findings of this study support the anomaly model (Schank, 1986), which predicts that overall, students will ask more questions when they encounter anomalous rewording
of statistics word problems, rather than when there are no such anomalies. These anomalies include deletion of critical information, contradictory information, salient irrelevant information, and subtle irrelevant information. The deletion versions consistently elicited more questions than the other versions, and the subtle irrelevancy versions triggered the fewest questions.

When transformation-relevant questions are analyzed, a similar pattern is found. Students ask more transformation-relevant questions when they encounter anomalous rewording, with deletion versions again eliciting more questions than the others. This time, the subtle irrelevancy versions failed to elicit significantly more questions than the original.

The Obstacle Hypothesis

What accounts for these differences? The main theory behind this is the so-called obstacle hypothesis. Based on Schank’s model (1986), other researchers (Graesser, Person & Huber, 1992) have identified types of anomalies and have contrasted the obstacles presented by each type. Those anomalies which “present a [clear] obstacle in planning, problem solving, or understanding” (Graesser & McMahan, 1993, p. 138) are more easily detected than those which do not present such obstacles.

In our present study, the deletion and contradiction versions present serious obstacles to problem solving and understanding, whereas the two irrelevancy versions do not present as much of an obstacle as the former two. In fact, the subtle irrelevancy version (as its name implies) presents even less of an obstacle than its salient counterpart. Thus, the frequency of transformation-relevant questions should be higher in the deletion and contradiction versions than in the irrelevancy versions.
Thus, this study finds empirical support for the obstacle hypothesis outlined above. The deletion and contradiction versions combined did produce more transformation-relevant questions than did the two irrelevancy conditions, and particularly, the deletion versions consistently generated more questions than did the others. However, this was not the case with the contradiction versions, which in this study did not differ significantly from the salient irrelevancy versions.

Why is this so? The obstacle hypothesis can be modified through the so-called “contradiction detection” model (Ottero & Campanario, 1990). Many people tend to discount contradictions, or they may rationalize these away (e.g., attributing the contradiction to a typographical error). However, further studies focusing on the contradiction version should be conducted before definite conclusions can be drawn.

On the whole, this study supports an earlier study (Lee-Chua, 2000) on probability word problems, which is not surprising, since the study of statistics and probability are inextricably linked (Pratt, 1998; Walpole & Myers, 1995). This study also supports the major findings of Graesser & McMahen (1993), with certain minor inconsistencies. For instance, in one of their experiments, only the deletion versions differed significantly from the subtle irrelevancy versions, as compared to our present study, which found the contradiction versions to differ significantly as well.

How do we explain these seemingly inconsistent findings? These different findings can be attributed to three factors: (1) the different fields which were the focus of investigation (fables and algebra problems in their case, statistics word problems in this case), (2) the different levels of specialization of the subjects (beginner psychology students in their case, third-year mathematics and computer science students in this case),
and (3) the different variations in procedures (e.g., strict time limits in their case, no time limits in this case). All these factors can lead to minor inconsistencies. It should be noted that the major findings (e.g., more questions are generated when anomalies are detected) are consistent in both studies.

Conclusion

Pedagogical Implications

What are the implications of the findings of this study on present classroom instruction?

1. Since previous research has found question-generation and rewording to be correlated to students’ comprehension and learning, teachers should strive to encourage students to ask questions and to formulate problems in their own words. Problems in mathematics texts are usually worded in formal technical language, and students who are not well-versed in this language will not be able to comprehend well the nuances of the problems (Davis-Dorsey, Ross & Morrison, 1991). When students are encouraged to reword the problems in their own language, their understanding of the problems is facilitated. Furthermore, if they ask questions (and reword the problems), teachers can then ascertain if the students comprehend what is being asked of them.

2. Students can be encouraged to ask questions and reword problems by addressing social and cognitive factors. Social factors, such as fostering a receptive atmosphere in the classroom for asking questions, have already been explored by other papers (Graesser, Person & Huber, 1992; Miyake & Norman 1979; Van der Meij, 1988). Since the present study concentrates on cognitive instead of social factors, we will not discuss them here. The major contribution of this study is cognitive--the confirmation that student question-asking and problem-rewording can also be
accomplished by dealing with cognitive factors, using the method of anomaly
detection described in the Discussion section.

3. To develop active thinking and participation by the students, teachers themselves
should possess good problem-posing skills. Problems can be reworded or transformed
in such a way as to engender critical thinking among students, in order to avoid the
fallacies and errors mentioned in the Introduction. When students actively reword,
analyze, and comprehend a problem, they tend not to fall back upon superficial
procedures, such as routine number-crunching computations or naïve intuition, which
are done without in-depth comprehension. It should be noted that past Philippine
surveys have revealed that many elementary and secondary teachers do not possess
adequate knowledge in problem-solving formulation (Lee, 1991). Recent efforts by
government and academe (such as the public high school teachers assisted by the
Department of Education and the National Educators Academy of the Philippines in
their graduate studies) have focused on providing in-service teachers with the
opportunity to learn new teaching strategies based on cognitive approaches.

4. Care must be taken when transforming word problems into the various versions in the
framework of this paper. Even if students providing additional information in
rewording may often facilitate comprehension and solution of word problems, we
have seen that too much information (especially when these are composed of subtle
and salient irrelevancies) may hinder the process. This finding is backed by previous
research (Davis-Dorsey, Ross & Morrison, 1991; Graesser, & McMahan, 1993; Lee-
Chua, 2000; Myers, Hansen, Robson, & McCann, 1983). Cognitive studies, such as
those from memory and learning research (Bruning, Schraw & Ronning, 1999) have
posed that when faced with novel ideas (such as learning new lessons), most students find it difficult to retain more than several discrete chunks of information at any one time. Encoding and retrieval processes may also be affected adversely (specific details are given in Craik & Lockhart, 1972; Andre, 1987). Mathematics (because of its abstract nature, Kintsch & Greeno, 1985) and statistics and probability (because of their overlaps with everyday usage, as discussed in the Introduction) are especially prone to learning problems and confusion.

5. With question-relevant information, however, student understanding is facilitated (King, 1992; Rosenshine, Meiser & Chapman, 1996). Comprehension improves because students practice selective attention, which has been aroused by thoughtful and relevant information. Student concentration increases, their focus on the word problem at hand (King, 1992).

6. Since statistics is linked to everyday usage, then teachers can take advantage of this by building on students’ informal knowledge. In their discussion of situated cognition, J. S. Brown, A. Collins and P. Daguid (1989) assert that mathematics instruction should not immediately attempt to define abstract concepts and procedures—without regard for the context that gives them meaning. “The implication is that, at least initially, learning should be linked with authentic problem situations that students understand well. Because procedures are built on comprehension, students will be able to apply their knowledge more flexibly” (Bruning, Schraw & Ronning, 1999, p. 343).
Recommendations for Research

Problem rewording, question generation, and anomaly detection are all rich sources for cognitive psychology and mathematics education research. Though some experiments have been published in recent years, there is still insufficient information in those areas. This paper now posits some modifications of the present investigation for future research:

1. Design experiments focusing on the differences between the two irrelevancy versions, and between the contradiction and salient irrelevancy versions (if any).
2. Modify some methodological features, such as putting a time limit on the task, or varying the subject sample, and compare results across various studies.
3. Replicate this study in different areas of mathematics, such as geometry word problems, calculus word problems, or linear algebra word problems.
4. Replicate this study with different groups of subjects (e.g., novice problem solvers and intermediate problem solvers), and compare the results across various studies.
5. Use other cognitive models (aside from Schank’s anomaly model) to explain the findings of this study.
6. Incorporate other variables (such as social factors) to form a better picture of question generation and the role of rewording. For instance, gender and culture come to mind. Do males tend to ask more questions than females? Do Western students tend to reword problems more than Asian ones? This framework would necessarily have to be expanded, and other variables operationalized.

Local and international studies (e.g., The Third International Mathematics and Science Study, or TIMSS) have revealed that mathematics learning in the Philippines lags behind most other countries (Nebres, 1998). The comprehension and solution of word problems
(statistics or otherwise) is one particular area of difficulty for students and teachers alike (Lee-Chua, 1995). Some nations in Southeast Asia have also encountered similar difficulties (Ogena & Golla, 1999). Rewording and anomaly detection may help learners become more successful in comprehending, and therefore, solving mathematical word problems.

References


The author acknowledges the support of the Endowed Professorial Chair for ME in honor of Ma. Beatriz Perez Rubio Delgado granted by the Loyola Schools, Ateneo de Manila University for the school year 2000-2001.

Biodata:

Dr. Queena N. Lee-Chua is an associate professor in the Mathematics Department of the Ateneo de Manila University, Philippines. She also teaches in the Psychology Department. Her main interests involve the psychology of teaching and learning mathematics (emphasizing cognitive and clinical aspects) and the popularization of mathematics and science. An author of several technical and popular science books, she also writes a weekly science column for a major daily, and hosts an educational television show.