# Using a Triad Judgment Task to Examine the Effect of Experience on Problem Representation in Statistics

Mitchell Rabinowitz (mrabinowitz@fordham.edu)<br>
Graduate School of Education, Fordham University<br>
New York, NY 10023 USA

Tracy M. Hogan (tehogan@fordham.edu)<br>
Graduate School of Education, Fordham University<br>
New York, NY 10023 USA

## Abstract

This research investigated whether the differences found between novices and experts in using surface and deep structures to categorize problems applied to the domain of statistics. Also explored was whether the methodology of a triad judgment task was reliable in discriminating how beginning and advanced students represent statistics problems. The task was designed in which source problems shared either structural features (t-test, correlation, or chi-square) or surface similarity (story narrative) with the target problem. Graduate students (N = 101) with varying levels of experience in the domain of statistics were asked to chose which source problem "goes best" with the target problem for each triad. Students with advanced experience in statistics tended to represent the problems on the basis of deep, structural features while beginning students tended to rely on surface features. Discussion on the effectiveness of the methodology employed and potential uses in other domains is presented.

## Introduction

Students learning statistics are required to learn a set of interacting skills. First, they need to become familiar with statistical procedures and how to use them (computing formulas). Second, they need to be able to recognize when to use those statistical procedures. The first set of skills is procedural in nature, i.e., they need to learn formulas and know how to execute the computation (or the statistical packages). The latter type of skill is representational, i.e., they need to be able to perceive and represent features within contexts that suggest which procedures should be used.

Previous research (Adelson, 1981; Chi, Feltovich & Glaser, 1981; Chase & Simon, 1973; Hardiman, Durfresne & Mestre, 1989; Schoenfeld & Herrmann, 1982) has shown that experts and novices within a domain represent problems within that domain on the basis of a different set of features. Bransford, Brown & Cocking (1999) report that this difference, in part, lies in knowledge organization. Expert knowledge centers on core concepts and big ideas found within the domain while novices rely on isolated facts and do not connect these facts in a way that allows them to generate inferences. For example, Chi and colleagues (1981) found that participants with advanced experience in physics sorted problems in their discipline on the basis of structural features, including the laws and principles of physics. When asked to sort the same problems, novices represented, and subsequently sorted the problems on the basis of surface features, such as the object being manipulated in the problem.

Quilici and Mayer (1996) argue that while surface features are generally more salient than structural features for novices, successful analogical transfer is dependent upon the recognition of structural similarities among problems. Consequently, they investigated the role of examples in how students learn to categorize statistic word problems. Their findings suggest that exposure to examples influences inexperienced students' structural schema construction, particularly when the example problems emphasize structural characteristics versus surface characteristics. Quilici and Mayer contend that their study merits further research concerning the conditions under which students rely on surface features or structure features in categorizing problems. In that Quilici and Mayer's participants were limited to those with little or no knowledge about statistics, further research concerning the effect of experience on problem representation is warranted.

This study was designed to replicate the expert/novice difference in perception and representational skill in the context of statistics problems. The purpose of this study was two-fold. First, the study investigated whether the differences found between novices and those with advanced experience in statistics use surface and/or deep structures to categorize problems applied to the domain of statistics. Second, this research explored whether the methodology of a triad judgment task was reliable in discriminating how beginning and advanced students represent statistics problems. Consequently, this study extended Quilici and Mayer's research (1996) by determining if those with advanced training in statistics do indeed cue in on the structural features of a statistical word problem.

To complete this extension, a triad judgment task was designed and administered to individuals with varying levels of statistical experience. According to Hardiman, Dufresne & Mestre (1989), the triad judgment task offers several advantages over the traditional sorting task used in previous research (Chi et. al, 1981; Quilici & Mayer, 1996). First, participants are able to concentrate on individual problem sets rather than being presented with a stack of cards to sort simultaneously. Second, the task allows for large-group administration and ease in scoring. The design of the triad task in this study was similar in nature to that employed by Hardiman and colleagues' research (1989). However, it differed in that this research examined problem representation in the domain of statistics while theirs was grounded in the field of mechanics.

The judgment task required participants to identify which of two given source problems "goes best" with a target problem (Figure 1). The source problems shared either similar surface features or structural features with the target problem. Surface features were similar in that the story narrative shared common characteristics while similar structural features involved the requirement of the same statistical test (t-test, correlation and chi-square). Surface features included similar story characters (personnel expert, meteorologist, college dean and psychologist) and similar dependent/independent variables (words typed per minute/experience of typists, annual rainfall/average yearly temperature, grade point average/reading score, number of errors on a test/amount of sleep). The structural features included the nature of the independent variable (one group or two independent groups) and the nature of the dependent variable (continuous or categorical).

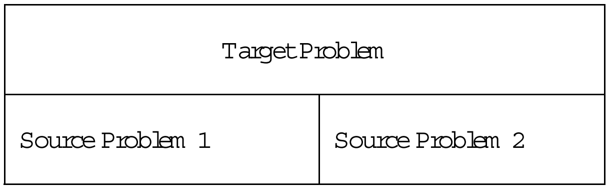

Figure 1
Structure of Triad Problems

Using the statistics word problems from Quilici and Mayer's study (1996), 18 triads were designed to investigate whether this judgment task would discriminate between those representing the problems using deep, structural features with those relying on surface features. To do this, we administered the task to students with varying levels of experience in the domain of statistics. We hypothesized that students with more advanced statistical experience would predominantly represent problems based on structural features while students with less statistical experience would tend to represent the problems based on surface features.

## Method

### Participants

The participants were 101 graduate students with a varied amount of experience in statistics. Those with no prior statistics courses totaled 27 participants, 33 participants completed one course, 13 finished two courses, 10 had completed three courses, six participants completed four courses, eight participants finished five courses, three participants had completed six courses and one participant completed eight courses. All individuals who volunteered to participate in this study earned course extra-credit.

### Problem Task

A triad judgment task was used to investigate the features that people use to represent common statistics problems. The task involved the presentation of three statistical problem statements- one target problem and two source problems. Participants were asked to read each set and judge which of the two source problems "goes best" with the target problem. Comparisons were based upon two features: surface and structure. Surface features were defined by the narrative characteristics (i.e., "After comparing weather data for the last 50 years, a meteorologist claims...") and structural features were defined by requisite statistical tests (t-test, correlation, chi-square).

There were three sets of comparison types that participants were asked to evaluate (Appendix). In the first comparison, one source problem shared only similar surface features to the target problem while the other source problem shared only similar structural features. Thus, Comparison One was considered Similar Narrative / Dissimilar Structure - Similar Structure / Dissimilar Narrative (SN/DS-SS/DN). In the second comparison, one source problem shared no similarities in either surface or structure while the other shared only similar structure to the target problem. Thus, Comparison Two was considered Dissimilar Narrative / Dissimilar Structure - Similar Structure/Dissimilar Narrative (DN/DS-SS/DN). In the third comparison, one source problem shared only similar surface features to the target problem while the other shared neither surface nor structural similarities. Thus, Comparison Three was considered Similar Narrative / Dissimilar Structure - Dissimilar Structure /

Dissimilar Narrative (SN/DS-DS/DN). Each participant was presented six triads per comparison for a total of 18 triads.

## Procedure

Participants were given a packet that contained the 18 triad problems and a cover sheet. On the cover sheet, the participants recorded background information including prior statistics courses, education level, and gender. Participants were tested during class and were given as much time as needed to complete the task.

## Scoring

A maximum score of 18 points, at six points per comparison type was possible. Participants scored one point per triad under Comparison One (SN/DS-SS/DN) and Comparison Two (DN/DS-SS/DN) if they selected on the basis of structural features. For Comparison Three (SN/DS-DS/DN), participants scored one point if they selected similar surface features in that neither comparison problems shared structural features with the target problem. Thus, a higher score implies a tendency towards choosing the structural dimension or the surface dimension where appropriate.

# Results

A correlation analysis was conducted to examine in greater depth the relationships between the level of experience (as measured by the number of statistics courses an individual completed) and the three comparison types. Findings suggest a significant relationship between number of courses and total score, r = .39, p < .01. This suggests that the more experience an individual has in statistics, the more likely they are to make more structural comparisons between two statistical passages.

While there was a significant correlation between the number of courses completed and total score on the triad judgment task, there were differences found among the three comparison types. Specifically, only Comparison One (SN/DS-SS/DN) and Comparison Two (DN/DS-SS/DN) were significantly correlated with the number of courses (r = .35, p < 01, r = .39, p < .01, respectively). These results, taken together, suggest that the more experience one has in statistics, the more likely he/she is to group statistical passages according to similar methodologies. As expected, there was no significant correlation between experience level and Comparison Three (SN/DS-DS/DN). If neither of the two source problems shared structural features with the target problem, individuals, regardless of experience, choose upon the basis of surface features.

In addition, to investigate whether individuals with more experience in statistics performed differently on the three comparison types as did novices, a repeated measures ANOVA was conducted. Individuals were grouped into three levels of experience in the domain of statistics. Level One included participants who had taken either zero or one course (n=60), Level Two reflected participants that had completed either two or three courses (n=23) and Level Three included participants that had completed four or more statistics courses (n=18). Means and standard deviations for scores on the three comparison types for each experience level are presented in Table 1. The significant interaction between experience level and comparison type suggests a relationship between the level of experience and the way the individual represents the particular statistical problem, F (2, 98) = 4.94, p < .01. Tukey's HSD test indicated that those in level three performed significantly different than those in levels one and two. The significant main effect of experience level indicates that individuals with more training in statistics represent statistical passages in ways that are more expert, F (2, 98) = 6.67, p < .01. The significant main effect of comparison type suggests that individuals, regardless of level of experience, do not respond in the same way to the different problems found in the triad judgment task, F (2, 98) = 44.89, p < .01.

Table 1: Means and Standard Deviations for Comparison Type by Experience Level.

| Level | n | Type I | | Type II | | Type III | |
|---|---|---|---|---|---|---|---|
| | | M | SD | M | SD | M | SD |
| One | 60 | 1.73 | 1.68 | 3.87 | 1.24 | 4.87 | 1.32 |
| Two | 23 | 2.13 | 1.94 | 4.04 | .88 | 4.52 | 1.20 |
| Three | 18 | 3.44 | 1.61 | 4.94 | 1.06 | 4.16 | 1.50 |

# Discussion

In this study, two questions were tackled. The first question was, How do beginning and advanced students in statistics compare in the way they represent statistical word problems? The analyses revealed several contrasts. It was shown that those with advanced experience tended to look for similar deep structures in the word problems presented within the triads. Conversely, the findings suggest that novices relied more heavily on the surface features to match a source problem with a target problem. However, when presented with comparisons types where neither of the source problems shared deep structural features with the target problem, all students, regardless of experience, selected on the basis of similar surface features.

The second question was, Can a triad judgment task be used to reliably discriminate how beginning and advanced students represent statistics word problems on either the basis of structural features or surface features? On the basis of earlier research (Chi, Feltovich & Glaser, 1981; Hardiman, Durfresne & Mestre, 1989), we reasoned that those with advanced training in statistics would make selections based on structural features while those with less training would select on the basis of surface features in a triad judgment task. Findings were consistent with our prediction. This suggests that the triad judgment task may indeed be a promising methodology to employ in studies where sorting tasks are traditionally used.

This study yields implications for educators of statistics. First, instruction in statistics should address the nature of problems and their structural components (e.g., type of data presented and the driving question of the problem). Second, learners should be provided with explicit instruction in recognizing similarities of problems based on core concepts, a skill requisite of experts (Bransford, Brown and Cocking, 1999).

This study certainly contributes to the relatively narrow research base of experts-novices in statistics, yet further studies are needed. Specifically, more studies are needed to explore the circumstances that promote the transition from using surface characteristics to deep structural features in representing problems.

## References

Adelson, B. (1981). Problem solving and the development of abstract categories in programming languages. Memory & Cognition, 9, 422-433.

Bransford, J.D., Brown, A.L., & Cocking, R.R. (Eds.). (1999). How people learn: Brain, mind, experience, and school. Washington DC: National Academy Press.

Chase, W.G., & Simon, H.A. (1973). Perception in chess. Cognitive Psychology, 4, 55-81.

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.

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

Quilici, J.L., & Mayer, R.E. (1996). Role of examples in how students learn to categorize statistics word problems. Journal of Educational Psychology, 88, 144-161.

Rabinowitz, M., & Glaser, R. (1985). Cognitive structure and process in highly competent performance. In F.D. Horowitz and M. O'Brien (Eds.), The gifted and talented: A developmental perspective. Washington DC: American Psychological Association.

Shoenfeld, A.H., & Hermann, D.J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. Journal of Experimental Psychology: Learning, Memory, & Cognition, 8, 484-494.

## Appendix

Comparison One: Similar Narrative/Dissimilar Structure - Similar Structure/Dissimilar Narrative (SN/DS-SS/DN)

Target: After examining weather data for the last 50 years, a meteorologist claims that the annual precipitation is more likely to be above average in years when the temperature is above average than when temperature is below average. For each of the 50 years, she notes whether the annual rainfall is above or below average and whether the temperature is above or below average.

Source 1: After examining weather data for the last 50 years, a meteorologist claims that the annual precipitation varies with the average temperature. For each of 50 years, she notes the annual rainfall and average temperature.

Source 2: A college dean claims that a group of good readers contains more honors students than a group of poor readers. For each of 100 first year college students, a reading comprehension test was used to determine whether the student was a good or poor reader and grade point average (GPA) was used to determine whether or not the student was an honors student.

Comparison Two: Dissimilar Narrative/Dissimilar Structure – Similar Structure/Dissimilar Narrative (DN/DS-SS/DN)

Target: A college dean claims that good readers earn better grades than poor readers. The grade point averages (GPA) are recorded for 50 first-year students who scored high on a reading comprehension test and for 50 first-year students who scored low on a reading comprehension test.

Source 1: A psychologist tests the hypothesis that people who are fatigued also lack mental alertness. An attention test is prepared which requires subjects to sit in front of a blank TV screen and press a response button each time a dot appears on the screen. A total of 110 dots are presented during a 90-minute period, and the psychologist records the number of errors for each subject. Twenty subjects are selected; half are tested after being kept awake for 24 hours and half are tested in the morning after a full night's sleep. Based on the number of errors on their test, each subject is also labeled as high or low in mental alertness.

Source 2: A personnel expert wishes to determine whether experienced typists are able to type faster than inexperienced typists. Twenty experienced typists (i.e., with 5 or more years of experience) and 20 inexperienced typists (i.e., with less than 5 years of experience) are given a typing test. Each typists average number of words typed per minute is recorded.

Comparison Three: Similar Narrative/Dissimilar Structure - Dissimilar Structure/Dissimilar Narrative (SN/DS-DS/DN)

Target: After examining weather data for the last 50 years, a meteorologist claims that the annual precipitation varies with the average temperature. For each of 50 years, she notes the annual rainfall and average temperature.

Source 1: After examining weather data for the last 50 years, a meteorologist claims that the annual precipitation is greater in years with below average temperature than in years with above average temperature. She notes the annual rainfall for each of 25 years that had above average temperatures as well as 25 years that had below average temperatures.

Source 2: A psychologist tests the hypothesis that people who are fatigued also lack mental alertness. An attention test is prepared which requires subjects to sit in front of a blank TV screen and press a response button each time a dot appears on the screen. A total of 110 dots are presented during a 90-minute period, and the psychologist records the number of errors for each subject. Twenty subjects are selected; half are tested after being kept awake for 24 hours and half are tested in the morning after a full night's sleep. Based on the number of errors on their test, each subject is also labeled as high or low in mental alertness.