The primary purpose of this study was to investigate the effects of three levels of personalization—individual, group, and non-personalized—on the achievement of seventh grade Hispanic boys and girls on mathematics word problems. Subjects were 123 seventh grade Hispanic students from a rural junior high school near Phoenix, Arizona. A pretest was employed to control for potential initial achievement differences by treatment or sex. Subjects were blocked by sex on the basis of their pretest scores, then randomly assigned within each block to the three treatments. A biographical inventory was administered three weeks prior to the administration of treatments to collect, in addition to biographical data, information regarding favorite objects, places, persons, and events for use in personalizing the word problems. The three treatments were administered during regularly scheduled math classes. Achievement was measured the following day using a 16-item constructed response posttest immediately after completion of the review material. An attitude scale was administered to measure the student's level of interest in the lesson and perceived level of difficulty, perception of the importance of the lesson, the level of importance attached to seeing one's name in print, preference for personalized or non-personalized problems, and preference for word or number problems. Analyses of the data indicate that personalization had a significant effect on student achievement on two-step problems, and that there was no significant difference in the overall performance of males and females. Attitude data generally favored individualized personalization over the other two treatments. (29 references) (GL)
Title:

Personalizing Math Word Problems

Author:

Cecilia L. Lopez
Introduction

Personalizing the context of instruction has been found to facilitate the reading comprehension of fourth-grade children who had been previously identified as poor readers (Bracken, 1982). There is also evidence that personalization increases the continuing motivation of high school students to return to task (Herndon, 1987).

Why should personalizing the context of instruction affect learner outcomes? Miller and Kulhavy (in press) hypothesized that personalization improves memory by increasing the associative strength during encoding of the personalized material and closely related content in a prose sentence. In a second experiment extending personalization to connective text, Miller and Kulhavy (in press) permitted subjects to encode information in terms of their own unique experiences, again incorporating the example that best fit the referent stimulus. In both studies, they found that incorporating personalized representations during encoding led to significantly greater recall of related information.

According to the National Assessment of Educational Progress (1975), solving story problems is a major weakness in the mathematical achievement of students. In analyzing the results of the National Assessment of Educational Progress (NAEP), Carpenter, Coburn, Reys, and Wilson (1975) found that only 31% of a national sample of 13 year olds could correctly solve a one-step word problem involving a 1-digit divisor and a 4-digit dividend. As a reaction to the findings of the NAEP, the National Council of Supervisors of Mathematics (1977) placed solving word problems first on its list of the ten most basic skills in mathematics.

Since learning to solve word problems is well documented as being difficult (Ballew & Cunningham, 1982; Carpenter et al., 1975; Marshall & Smith, 1987; Wright & Wright, 1986), a number of researchers have attempted to explain the causes of such difficulties. De Corte, Verschaffel, & De Win (1985) found that ambiguously stated problems, and not a lack of computational skills per se, significantly contribute to the inability of children to comprehend the text of word problems.

Both reading ability (Ballew & Cunningham, 1982) and computational ability (Muth, 1984) have been found to be significant factors in contributing to skill in solving word problems.

Personalizing of word problems appears to ameliorate
some of the difficulties youngsters have in solving such problems. Kintsch (1986) and Hudson (1983) report that word problems containing abstract situations or actions are twice as difficult (39% as opposed to 79%) for children to solve as are word problems incorporating actions or situations that are concrete and familiar to the child. Wright and Wright (1986) administered personalized test items to fourth graders and found significant improvement in students' ability to select correct arithmetic operations, but little effect on their computation skills. Ross and Anand (1987) and Anand and Ross (1987) found that a personalized context of instruction embedding familiar items, such as learner interests and background, increased the achievement of fifth and sixth graders to solve verbally stated math problems. However, subjects in both studies by Ross and Anand were enrolled in a university affiliated school. Thus, it may be inappropriate to generalize their findings to children attending schools in lower socioeconomic areas.

Personalizing the context of instruction has been explored both with groups and with individuals. In a program on conditional syllogisms, Herndon (1987) personalized the context of practice items based on the most common interests of the subject pool, rather than on individual interests. Herndon (1987) found that personalization significantly affected the continuing motivation of high school seniors to return to task, but did not significantly improve their achievement. López (1989) personalized word problems, using individual responses to a biographical inventory. Her results yielded significant differences favoring the personalized treatment in both achievement and student attitude, but no significant differences in continuing motivation.

The rationale for using group instead of individually personalized context is one of expediency and efficiency. Incorporating group interests involves the development of a lesson or lessons that can be used for the entire class. Individual interests require that separate lessons be generated for each student. The group personalization should take considerably less time to generate than the individual one. Consequently, it is important to determine if the group method is as effective as the individualized one.

Research on sex differences in mathematics achievement has produced mixed results (Marshall, 1984). Maccoby and Jacklin (1974) found no sex differences until age 14, but differences favoring males after that age. Results from
the 1979 NAEP data showed no significant sex-related differences among 9, 13, and 17 year olds in solving word problems (Linn & Petersen, 1986).

In a review of the 1978 NAEP results, Valverde (1984) found that Hispanic children were 9 percentage points below the national average (55% to 46%) in mathematics achievement at age 9. By age 17, Hispanics had dropped to 12 percentage points below the national average (60% to 48%). Matthews, Carpenter, Lindquist, Silver (1984) reported similar results from the 1982 NAEP data.

Despite these findings, few studies have examined ways to improve the mathematical performance of Hispanic students. In one such study, Hannafin (1983) found a trend toward superior performance for Hispanic sixth graders receiving an instructional system designed to provide hierarchically sequenced sets of computational skills over Anglos who had received traditional instruction. This trend occurred only on instructional tasks which were relatively novel or independent of prerequisite skills. López (1989) investigated the effect of individualized personalization on the performance of Hispanic eighth graders on mathematics word problems. Her data revealed significant differences favoring personalized over non-personalized instruction on both student achievement and attitudes.

The primary purpose of this study was to investigate the effects of three levels of personalization (individual, group, non-personalized) on the achievement of seventh grade Hispanic boys and girls on mathematics word problems. Data were also collected on subjects' attitudes toward the instruction. A pretest was employed to control experimentally for potential, initial achievement differences by treatment or sex. Subjects were blocked by sex on the basis of their pretest scores, then randomly assigned within each block to the three treatments.

**Method**

**Subjects**

One hundred twenty-three seventh grade Hispanic students from a rural junior high school with a primarily Hispanic enrollment near Phoenix, Arizona were the final group of subjects for this study. Subjects were from 10 math classes taught by four teachers.
Materials

Biographical inventory. A biographical inventory based on the work of Anand and Ross (1987) and Ross and Anand (1987) was designed for this experiment. Respondents were asked questions regarding their background and interests. Topics included the names of friends and favorite places, foods, objects, activities, and events.

Instructional lesson. Three parallel versions of a unit on division of whole numbers were developed. The non-personalized version, serving as a template, was written first. The individual and group versions were then written with nouns, pronouns, prepositional phrases, and dates inserted as appropriate in the original non-personalized context. Familiar items, including birth dates and city of residence, were derived from Biographical Inventory items.

Variations in treatments were manipulated by changing the referents of example and practice problems. However, the numerical values and methods of measurement (e.g., number or weight) were identical for all treatments.

The non-personalized treatment used the standard referents found in math textbooks. An example of a typical non-personalized referent follows:

Jim expects 36 friends to come to a party. If each friend will get a 6-ounce (oz) serving of soda, how many 12-oz cans of soda are needed?

In the individual treatment, nouns and pronouns were replaced or modified with the familiar items students provided in the Biographical Inventory:

Gabriel expects 36 friends to come to his birthday party on March 7. If each friend will get a 6-ounce (oz) serving of Dr. Pepper, how many 12-oz cans of Dr. Pepper are needed?

The instructional lessons for the individual condition were computer generated so as to provide a feasible way of personalizing the lesson context for individual experiences and interests.

The group treatment was based on the responses of the majority. Nouns, pronouns, or noun modifiers were replaced with the most familiar objects, places, and persons the majority of students provided in the
Biographical Inventory. An example of how variations in the group treatment were manipulated follows:

Miguel expects 36 friends to come to his birthday party in December. If each friend will get a 6-ounce (oz) serving of Pepsi, how many 12-oz cans of Pepsi are needed?

The lesson covered procedures in solving one-step and two-step word problems involving division of whole numbers. Instruction on the strategy for solving both item types (one-step and two-step problems) contained the rule and its application with appropriate examples. Practice problems, three for each item type, followed. Each of the eight problems was placed on a separate page with enough open space to allow students to work the problem directly on the page. Answers to all practice problems were provided at the end of the lesson booklet.

Review. Prior to administration of the posttest, the seven-page review over the lesson content was administered. Three versions of the review were developed (individual, group, and non-personalized) to parallel the lesson content. The review contained a summary of rules and procedures for each item type and four practice problems, two for each item type. As in the instructional material, answers to all problems appeared at the end of the booklet.

Procedure

The Biographical Inventory was administered to students three weeks prior to administration of treatments. Besides biographical data, information regarding favorite objects, places, persons, and events were obtained.

Subjects were blocked by sex then randomly assigned within blocks to one of three levels of personalization (individual, group, or non-personalized).

The three versions of the instructional lesson were administered on a within-class basis in regularly scheduled math classes three weeks after the administration of the Biographical Inventory. The cover sheet on each booklet contained the student’s name, class period, and task related instructions which the experimenter read to all subjects. Students were told that they were helping to evaluate a new lesson. Additionally, they were told that after completing the lesson and review, they would take a test and that their test scores would count toward their
The review was given on the following day. The posttest was distributed to each student immediately after completion of the review material.

**Criterion Measures**

**Achievement.** The 16-item constructed-response posttest served as the criterion measure for achievement. The test covered the same mathematics operations as the instructional materials. It consisted of eight one-step and eight two-step word problems. None of the 16 problems involved remainders. All problems were in the non-personalized form. The inter-item reliability for the posttest, calculated in an earlier study by López (1989), is .77.

**Attitudes and continuing motivation.** This 4-point scale, containing 11 items, assessed level of interest; level of difficulty; number of familiar persons, places, or things; perception of the importance of the lesson; level of importance attached to seeing one's name in print; how much subjects liked the lesson; preference for personalized and non-personalized problems; and preference for word and number problems. Open-ended questions assessed what subjects liked most and least about the lesson.

**Design and Data Analysis**

The experimental design was a 3 Level of Personalization (Individual/ Group/ Non-Personalized) x 2 Sex (Male/ Female) x 2 Item Type (One-Step/ Two-Step) factorial design, with item type as a within-subjects variable. Achievement data were analyzed by multivariate analysis of variance (MANOVA) repeated measures. Attitude data were analyzed by separate analyses of variance (ANOVA).

**Results**

**Achievement**

Table 1 shows the mean scores by level of personalization, sex, and item type. The mean overall posttest scores by level of personalization were 9.49 for the individualized treatment, 9.61 for group, and 7.73 for non-personalized. The multivariate analysis yielded a significant difference for treatments, F(2, 117) = 3.61, p < .03. Univariate analyses revealed that the significant
treatment effect occurred for two-step problems, $F(2, 117) = 4.01, p < .021$, but not for one-step problems.

Post hoc contrasts using Newman Keuls were performed to determine the source of differences between treatments. Both the individualized mean and the group mean were significantly higher at the .05 level than the mean for the non-personalized treatment. These differences are reflected in the two-step problem means of 3.23 for individualized subjects, 3.44 for group subjects, and only 2.06 for non-personalized subjects. The individualized and group means did not differ significantly from each other.

Overall mean scores by type of test items were 6.02 for one-step items and 2.89 for two-step items, a highly significant difference, $F(1, 117) = 288.43, p < .001$. The mean posttest score for females was 9.06 and the mean for males was 8.71, a non-significant difference.

**Attitudes**

The mean attitude scores, based on a score of 1 for the most positive response to each item and 4 for the most negative, are shown in Table 2. The table reveals that the overall mean scores across all seven items were 1.69 for the individualized treatment, 1.82 for the group treatment, and 1.87 for non-personalized instruction. Analysis of variance of these means yielded a highly significant difference, $F(2, 117) = 3.12, p < .003$. Post hoc contrasts using Newman Keuls revealed that the overall mean for the individualized treatment was significantly higher than those for both the group and the non-personalized treatments. The group and non-personalized means did not differ significantly from each other.

When the seven questionnaire items shown in Table 2 are considered individually, it can be seen that significant differences occurred for items 2, 3, 6, and 9. Post hoc contrasts showed that the individualized treatment mean was significantly different from the non-personalized mean ($p < .05$) on each of these items. Individualized subjects reported that the lesson was more difficult, there was more familiar content, they liked the lesson better, and they preferred personalized content. Means between the individual and group treatments and the group and non-personalized treatments did not differ significantly from one another on any of the four items.

Two additional items on the attitude questionnaire assessed preference for math word problems that were not
personalized and preference to study more math word as opposed to number problems. ANOVA followed by Newman Keuls revealed that subjects in the individualised treatment (2.72) had a significantly lower preference for non-personalised problems than subjects in both the group (2.42) and the non-personalised (2.39) treatments, $F(2, 117) = 5.05, p < .008$. The three treatment groups did not differ significantly in their preference to study more word problems.

Frequency of constructed responses to the open-ended questions about what subjects liked most and what they liked least were also tabulated. Student responses indicated that what they liked most was the personalization of the lessons. This factor was cited 24 times, 19 by subjects in the individualised treatment and 5 by those in the group treatment. Subjects cited specific types of computations, notably division and multiplication, as what they liked least. A total of 25 responses fell into this category, 13 from the non-personalized treatment.

**Discussion**

The present research revealed a significant effect for personalization on student achievement on two-step problems, the more difficult problems in the instructional unit. It seems likely that at least two factors may have contributed to this effect. One is that personalized subjects liked the lesson more, as indicated by responses of individualized subjects to question 6 on the attitude survey -- "How much did you like the lesson?" This could have caused them to expend greater effort on the lesson, which could explain the fact that the individualized subjects reported the instruction to be significantly harder (question 2 on the attitude survey). If subjects receiving personalized instruction liked the instruction more and invested more effort in it, they could reasonably be expected to demonstrate improved performance.

A second possible contributing factor is that personalizing the problems may have increased subjects' level of understanding of them. When the problems are personalized by using familiar names, objects, and situations, the problem situation may be more familiar to the subjects. They may therefore be more able to deduce the necessary mathematical operations.

An explanation from cognitive theory suggests how personalization may influence learning by affecting encoding and retrieval. Miller and Kulhavy (in press)
demonstrated that individualized personalization increases the retrievability of associate information by stimulating stronger and more memorable encoding. If personalization does provide the learner with more memorable illustrations of rule application, then more effective semantic processing of the problem situation may occur.

López (1989) found in an earlier experiment that individualized personalization of instruction was effective in improving the achievement of Hispanic children. The present results indicate that personalization is also effective with Hispanic students on a group basis as well as on an individual one. Thus, this finding extends to group personalization the earlier results obtained with individualized personalization by López (1989) and by Anand and Ross (1987) and Ross and Anand (1987).

Males and females in the present study did not differ significantly in their overall performance or their performance on particular types of problems. This finding is consistent with the results of the National Assessment of Educational Progress (NAEP) from 1973, 1978, and 1982. The NAEP, a comprehensive national test administered to a very large sample, mainly uses constructed-response items for word problems (NAEP, 1977), as did the present study. The NAEP data consistently yielded no significant sex-related differences among students at ages 9 and 13 in solving math word problems (Linn & Petersen, 1986; NAEP, 1983). Several studies published in the 1970s also reported no sex-related differences in mathematics achievement in the elementary grades (Hilton & Berglund, 1974; Maccoby & Jacklin, 1974).

Other researchers, however, report sex-related differences, generally favoring males in mathematics achievement. Fennema and Sherman (1977), Hilton and Berglund (1974), and Maccoby and Jacklin (1974) all report significant differences favoring males starting at grade 7. Benbow and Stanley (1982) report significant differences favoring males with mathematically gifted and highly motivated seventh and eighth graders. All of these results favoring males were from research using multiple-choice tests.

The variations in findings regarding sex-related differences may be a function of either the differing nature of the tests (constructed-response versus multiple choice) or of the differing subject populations. The most comprehensive sample across time, grade levels, number of subjects, and geographic sites was from the NAEP studies.
which consistently yielded no sex-related differences on primarily constructed-response items. That boys performed better than girls on several studies involving multiple-choice tests may relate to the fact that these tests require somewhat different cognitive operations than constructed-response tests. The correct answer can often be derived in the former by examining the response choices and estimating the most reasonable choice, whereas constructed-response tests require computation of the correct answer. It is possible that males are better at the estimating skill, but not on the computations required in constructed-response tests.

The attitude data generally favor individualized personalization. Subjects under individualized personalization had significantly more positive overall attitudes than those under the other two treatments. Individual subjects liked the lesson better, recognized more familiar content, preferred personalized problems, and had a lower preference for the prospect of non-personalized problems in future math content. The most plausible explanation for the unexpected finding that subjects in the individualized treatment reported the lesson to be more difficult would seem to be that the personalized subjects expended greater effort on the lesson because of the personalization.

An interesting aspect of the high attitude ratings for individualized personalization is that they were obtained with content that is generally unpopular with students. Story problems, especially multi-step problems, are considered by students to be difficult (Ballew & Cunningham, 1982; Marshall, 1984; Marshall & Smith, 1987) and boring (Cheek & Castle, 1981). Nevertheless, the attitudinal responses of personalized subjects in this study indicated that the instruction was interesting and that they liked it.

The present results have implications for educational practice generally. The simplest application is for instructors to learn the interests of their students, either through a formal survey of interest or more informal means, and incorporate them into instruction whenever possible. The data from this study suggest that use of common interests of the group in this manner is also effective. More sophisticated application could involve the design of computer software that incorporates individual or group interests into word problems and other practice material while maintaining the integrity of the content to be learned.
Both the present study and the earlier research by López (1989) indicate that personalization in mathematics instruction is effective in improving the performance of Hispanic children, especially with two-step math problems. Both studies were conducted over relatively brief time periods of two instructional lessons each. Further research should be conducted to investigate the extent to which positive effects would be sustained over an extended time period and with different subject matter. Research investigating these issues should help to clarify the conditions under which personalization of instruction is most effective in maximizing improved classroom performance.
References


Table 1

Means and Standard Deviations of Posttest Scores by Level of Personalization, Sex, and Test Item Type

<table>
<thead>
<tr>
<th>Level of Personalization</th>
<th>Individualized</th>
<th>Group</th>
<th>Non-Personalized</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-Step</td>
<td>2-Step</td>
<td>1-Step</td>
</tr>
<tr>
<td>Females</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>6.23</td>
<td>3.36</td>
<td>6.24</td>
</tr>
<tr>
<td>SD</td>
<td>1.85</td>
<td>2.48</td>
<td>1.70</td>
</tr>
<tr>
<td>Males</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>6.29</td>
<td>3.09</td>
<td>6.07</td>
</tr>
<tr>
<td>SD</td>
<td>1.49</td>
<td>2.36</td>
<td>1.58</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>6.26</td>
<td>3.23</td>
<td>6.17</td>
</tr>
<tr>
<td>SD</td>
<td>1.66</td>
<td>2.40</td>
<td>1.63</td>
</tr>
<tr>
<td>Questionnaire Item</td>
<td>Individ</td>
<td>Group</td>
<td>Non-Pers</td>
</tr>
<tr>
<td>------------------------------------</td>
<td>---------</td>
<td>-------</td>
<td>----------</td>
</tr>
<tr>
<td>1. Interesting-Boring</td>
<td>1.93</td>
<td>1.89</td>
<td>2.07</td>
</tr>
<tr>
<td>2. Easy-Hard</td>
<td>2.02</td>
<td>1.75</td>
<td>1.75</td>
</tr>
<tr>
<td>3. Familiar Content</td>
<td>1.53</td>
<td>1.81</td>
<td>2.16</td>
</tr>
<tr>
<td>4. Importance of Problem Solving</td>
<td>1.28</td>
<td>1.44</td>
<td>1.23</td>
</tr>
<tr>
<td>5. Importance of Name in Print</td>
<td>2.07</td>
<td>2.25</td>
<td>1.93</td>
</tr>
<tr>
<td>6. Liking for Lesson</td>
<td>1.63</td>
<td>1.94</td>
<td>2.02</td>
</tr>
<tr>
<td>9. Preference for Personalized Problems</td>
<td>1.37</td>
<td>1.69</td>
<td>1.93</td>
</tr>
<tr>
<td><strong>GRAND MEAN</strong></td>
<td>1.69</td>
<td>1.82</td>
<td>1.87</td>
</tr>
</tbody>
</table>

*Note.* Questions 7 and 8 were open-ended items.