Student Performance and Attitudes Using Personalized Mathematics Instruction

☐ Heng-Yu Ku
   Howard J. Sullivan

This study investigated the effects of personalized instruction on the achievement and attitudes of Taiwanese students on two-step mathematics word problems. A total of 136 fourth-graders in a Taiwanese public school participated in the study. Subjects initially completed a Student Survey on which they chose their favorite foods, sports, stores, classmates, and other selections. The most popular items were then used to create personalized math word problems for the pretest, personalized instructional program, and posttest. Subjects were blocked by ability based on their pretest scores and were randomly assigned within ability levels to either a personalized or nonpersonalized version of the print-based instructional program. After finishing the program, subjects completed a student attitude survey and the posttest. A repeated-measures univariate analysis of variance revealed that subjects in the personalized treatment made significantly greater pretest-to-posttest gains than those in the nonpersonalized treatment. Subjects also performed significantly better on the personalized pretest and posttest problems than on the nonpersonalized problems. Personalized subjects and higher-ability students both had significantly more positive attitudes toward the instructional program than did their nonpersonalized and lower-ability counterparts.

☐ Data from the National Assessment of Educational Progress (1992a, 1992b) indicate that mathematics word problems are difficult for students at all age levels in elementary and secondary schools. A major cause of the difficulty appears to be student inability to convert the problems into the math operations that must be performed to solve them (Hart, 1996). Some researchers have also noted that lack of familiarity with word problem structures may also contribute to poor student performance (Mayer, 1982; Rosen, 1984).

Comprehending a problem well enough to determine the correct operations to perform is, of course, an essential skill for solving math word problems. Several authors have argued that students are better able to comprehend and solve word problems when the problems are framed in a real-world context (Bransford, Sherwood, Hasselbring, Kinzer, & Williams, 1992; Brown, Collins, & Duguid, 1989; Choi & Hannafin, 1995). The practice of presenting problems in a context that is designed to reflect the expressed real-world interests or preferences of students is often referred to as personalization.

Personalizing mathematics word problems, such as incorporating personal background information into the problem content, can lead to improvements in performance (Anand & Ross, 1987; Davis-Dorsey, Ross, & Morrison, 1991; Lopez & Sullivan, 1991, 1992). Anand and Ross tested the effect of using computer-assisted instruction to personalize mathematics instruction for elementary school children. Students who received personalized instruction scored significantly higher on math word problems involving rule recognition and transfer than those whose instruction was not personalized.

The authors claimed that personalized contexts increased students' comprehension and motivation by helping them interpret important information in the problem statement. Davis-Dorsey et al. found that both second-grade and fifth-grade students made significant achievement gains from rewording and personalization of the context in math word problems. In two separate studies with rural eighth-grade Hispanic American students, Lopez and Sullivan found significant overall achievement effects for personalization on one-step and two-step mathematics word problems.

Several studies have found that student attitudes are more positive when student interests and preferences are incorporated into instruction in order to personalize it. Ross and his colleagues (Ross, 1983; Ross, McCormick, & Krisak, 1986; Ross, McCormick, Krisak, & Anand, 1985) employed personalization in a series of adaptive instruction studies. Favorable attitude results were obtained when preservice teachers received education-related materials and nursing students received medical-related materials, and poorer results were attained when each group received the other's materials. Herndon (1987) found that high school students who received instruction based on common group interests had significantly more favorable attitudes and higher return-to-task motivation than students whose instruction was not interest-based. Cordova and Lepper (1996) and Hart (1996) also found more favorable attitudes or motivation toward personalized than toward nonpersonalized instruction.

Researchers have offered two theory-based explanations for the effectiveness of personalized instruction in studies where it has yielded better results than nonpersonalization. One is that students' greater familiarity with personalized problem situations and content may enable them to solve problems more easily by reducing their cognitive load (Lopez & Sullivan, 1991, 1992; Miller & Kulhavy, 1991). This position is supported by d'Ailly, Simpson, and MacKinnon's (1997) statement that "self-referencing facilitates general encoding processes and decreases the load on working memory during problem solving" (p. 566).

The second explanation is based on interest theory (Mayer, 1998). Mayer notes that students exert more effort and are more successful solving problems that interest them than problems that do not. Several researchers have cited greater student interest and motivation as reasons for better performance under personalized instruction (Cordova & Lepper, 1996; Lopez & Sullivan, 1992; Ross & Anand, 1987). The reduced-cognitive-load and increased-interest explanations appear to be compatible with one another rather than being alternative or competing explanations.

The success of personalization of instruction in studies in the United States caused the first author of this article to become interested in personalization as a potential technique for use in mathematics instruction in his home country of Taiwan. Two of the recent themes of education reform in Taiwan are (a) incorporating real-world examples into the learning context, and (b) changing the method of assessment (Ministry of Education, 1998). The National Council of Teachers of Mathematics in Taiwan recently proposed a new approach that favors a student-centered problem-solving approach in mathematics education. Personalization could be an appropriate method for making math instruction more meaningful and motivating to Taiwanese students under this new approach.

Taiwan has a much more homogenous culture than the United States. Its population of more than 22 million people and its total land area of less than 13,000 square miles make it one of the most densely populated countries in the world (U.S. Department of Commerce, 1999). There is an average of approximately 40 students per classroom in Taiwan's elementary schools. Students are under strong pressure to perform well at school, and many of them rely heavily on memorization to master subjects, including math (Cheng, Chang, Chiang, & Guo, 1998). The newer perspective on mathematics encourages problem solving and making closer connections between mathematics and the real world.

The authors' initial study of personalization in Taiwan investigated the effects of group personalization of math word problems on the achievement and attitudes of 72 fifth-grade students (Ku & Sullivan, 2000). Personal informa-

tion and preferences provided by students were incorporated into the mathematics word problems for students receiving the personalized treatment. Personalization of instruction did not yield a significant achievement difference over nonpersonalization, but this lack of a reliable difference appeared to be attributable to a ceiling effect that limited the improvement of higher-ability students. However, students did make significant pretest-to-posttest gains across both treatments and scored significantly higher on personalized than on nonpersonalized posttest items. Significant two-way interactions reflected greater pretest-to-posttest improvement for lower-ability than for higher-ability students and a greater difference between scores on personalized than on nonpersonalized posttest problems for lower-ability students. Student attitudes were significantly more favorable toward the personalized instruction.

The Ku and Sullivan (2000) study left at least two important questions unanswered regarding the effects of personalization. One question was if personalization of mathematics word problems for Taiwanese students would produce a significant achievement difference over nonpersonalization when posttest achievement is not limited by a ceiling effect. A second question was if personalizing mathematics test items would yield higher achievement on the test items even when students did not receive personalized instruction. The Ku and Sullivan study was not specifically designed to address the latter question, but data from the study suggested that simply stating test items in personalized form might yield significantly better performance on them, irrespective of whether students receive personalized instruction or not. If that is the case, student performance on mathematics tests generally may be improved by using contexts on the test items that are personalized, or at least familiar to most students.

The present study was designed to address the unanswered questions from the Ku and Sullivan (2000) research and to investigate the stability of the other findings from that study and other research. The likelihood of a ceiling effect, which confounded interpretation of the personalization versus nonpersonalization issue in the earlier research, was reduced by conduct-

ing this study with fourth-grade students instead of fifth graders as in the earlier study. In addition, this study was conducted with 136 students, or nearly twice as many as the earlier one, to increase the power of the statistical analyses. These changes were designed to provide a clearer answer to the question of whether or not instruction incorporating personalized math word problems would yield significantly better student performance than instruction using only nonpersonalized items.

The question of whether merely stating test items in personalized form would yield better performance than would stating them in nonpersonalized form, irrespective of the type of instruction students receive, was addressed by including an equal number of personalized and nonpersonalized items on the pretest, as well as on the posttest. In the previous study, the pretest had only 8 items, all of which were nonpersonalized, and the posttest had 12 items, 6 personalized and 6 nonpersonalized. Thus, comparisons of performance on personalized and nonpersonalized test items could be made only on the posttest. Interestingly, subjects in both the personalized and nonpersonalized instructional treatments performed significantly better on the personalized items than on the nonpersonalized items on the posttest, suggesting that simply stating test items in personalized form results in better performance on these items. Including personalized and nonpersonalized items on the pretest, as well as on the posttest, in the present study was designed to permit more direct and stable comparisons of the effects of the two types of test items alone and in combination with the instructional treatments.

The problem contexts for group personalization are personalized for the entire group based on the most common interests and preferences of all group members, whereas the contexts in individualized personalization are normally personalized by computer for each individual based on the interests and preferences of that individual. Group personalization is easier for teachers and instructional designers to construct than individualized personalization. However, a disadvantage of the group procedure is that the interests and preferences of some individuals in the group may not be well repre-

sented in the instruction because the personalization is based on the group preferences. Group personalization was used in this study because it is easier to construct and for teachers to implement and because the limited availability of computers in the schools, as is the case in Taiwan and many other countries, also favors its use.

The present study investigated the effects of two levels of group personalization (personalized, nonpersonalized) on the achievement of fourth-grade Taiwanese students on two-step mathematics word problems. Mathematics ability level, as determined by scores on the pretest administered prior to the instructional phase of the study, was also included as a variable because it is an important factor in mathematics achievement and because of differential findings by ability level in earlier research on personalization (Dwyer, 1996; Ku & Sullivan, 2000; Lopez & Sullivan, 1991).

The primary research questions for the study were as follows:

1. Does personalization of instruction increase the achievement of Taiwanese students on mathematics word problems?

2. Does personalization of instruction have a differential effect on the performance of higher-ability and lower-ability students on mathematics word problems?

3. Do Taiwanese students perform better on personalized word problems than on nonpersonalized problems irrespective of the type of instruction (personalized or nonpersonalized) they receive?

4. Does personalization of mathematics word problems influence student attitudes toward the instruction on these problems?

METHOD

Subjects

A total of 136 fourth-grade Taiwanese students from four classes taught by different teachers at a public elementary school in Taiwan participated in this study. The school is located in a midlevel income and socioeconomic area in Fen-

gyuan, a city with a population of approximately 160,000 people.

Materials

Student survey. A 20-item student survey was used to determine the personal backgrounds and interests of the participants. Topics included the names of the students' favorite places, activities, sports, friends, convenience stores, foods, and so forth. Students gave two favorite responses for each survey item.

The survey was administered one week prior to the pretest. Responses to each survey item were tabulated by the experimenter and then used to design the personalized version of the instructional program and the tests.

Instructional program. Two parallel versions of an instructional program on two-step math word problems were developed in print form in Chinese. Taiwanese students learn addition and subtraction in the third grade and multiplication and division in the fourth grade. The word problems in the instructional program and the test items were taken directly from the fourth- and fifth-grade mathematics textbooks used by the participants. The program was administered after the students had studied one-step multiplication and division, and it covered two-step processes that they had not yet formally studied.

The nonpersonalized version of the instructional program was written first and included standard problem types from the students' math textbooks. The personalized version was then written by incorporating the most popular referents (places, foods, sports, names, etc.) from the Student Survey into the previously nonpersonalized version.

An example of a practice item in personalized and nonpersonalized form is provided below. The example item requires using multiplication followed by division.

Nonpersonalized: The teacher has 2 dozen cans of soft drink to be shared equally by 8 students. How many cans of soft drink does each student get?

Personalized: The teacher, Ms. Sue, has 2 dozen cans of milk tea to be shared equally by 8 students. How many cans of milk tea does each student get?

The instructional program covered procedures for solving two-step word problems involving four different combinations of multiplication and division operations (multiply-multiply, multiply-divide, divide-multiply, divide-divide) with whole numbers. A four-step strategy based on the work of Enright and Choate (1993) was incorporated into the instructional program for both treatments. The four steps, which had not been taught to the students previously, were:

1. Read the question.
2. Think through the problem. What must be found out? What steps are involved?
3. Choose the steps and do the math.
4. Check your answer.

Instruction for solving each of the four types of problems contained the four-part Enright and Choate (1993) strategy with two worked examples for each problem type. After the examples for each problem type, the instructional program contained two practice problems for the students to solve. Each pair of practice problems was placed on a page with enough open space to allow students to work out the problems. When students had completed the instructional program, the experimenter collected the programs and scored the practice problems. On the following day, the experimenter conducted a review in which he provided the answers to the practice problems and the explanations for them.

The instructional materials used in this study were in a similar format to those in the Ku and Sullivan (2000) study and covered the same instructional objective. The practice problems and examples were similar, though not identical, to those in the earlier study, and were personalized based on the interests and preferences of the 136 subjects in the present study. The Enright and Choate (1993) four-step problem strategy was taught and practiced by students more systematically in the present research than in the 2000 study. The pretest forms were modified for this study to include more items and equal numbers of personalized and nonpersonalized items, as described in the criterion measures section below. Additional items were added to the student attitude survey for this second study, and a teacher attitude survey was also added. All materials used in the study were in Chinese.
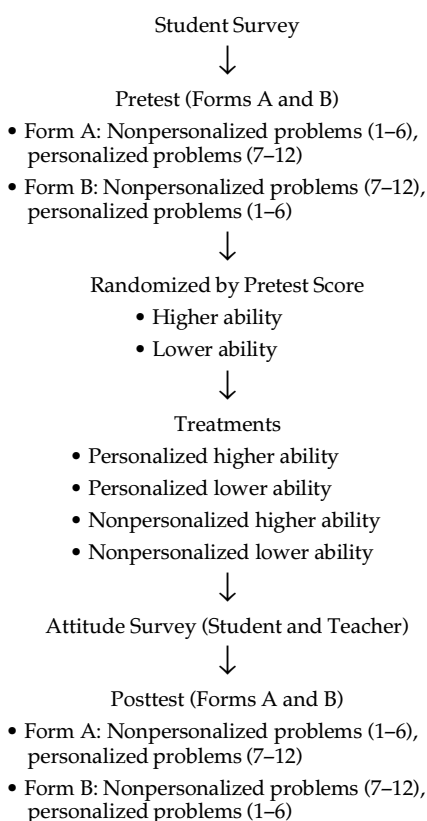
Procedures

The student survey was administered to participants two weeks prior to the treatment. The most popular choices (places, foods, sports, etc.) from the survey were subsequently used to convert the nonpersonalized problems into the personalized content for the pretest, the instructional program, and the posttest. The pretest was administered one week after students filled out the Student Survey.

After the pretests had been scored, the subjects were blocked within each class by their pretest scores into higher-ability and lower-ability groups, and were assigned within blocks to either the personalized or the nonpersonalized version of the instructional program. Sixty-eight subjects each, 34 higher-ability and 34 lower-ability, were in the personalized treatment and in the nonpersonalized treatment. The mean pretest score was 9.35 (SD = 1.65) for the 68 higher-ability subjects and 3.74 (SD = 2.11) for the 68 lower-ability subjects.

The experimental part of the study took place in three 40-min class periods (one on each of three different days) one week after the pretest. The experimenter served as the instructor for all treatments in regularly scheduled math classes in four different classrooms. The experimenter read instructions to all students and told them that they would be helping with the development of a new math program in Taiwan and that they should try their best to learn and to solve the problems. On the first day, subjects completed the instructional part of the program. On Day Two, the experimenter went over the eight practice problems with subjects and wrote the answers on the blackboard. After this review, students filled out the student attitude survey and teachers completed the teacher attitude survey. On the final day, subjects took the posttest. A diagram summarizing the experimental design is shown in Figure 1.

Figure 1 □ Summary of experimental design

Student Survey
↓
Pretest (Forms A and B)
- Form A: Nonpersonalized problems (1–6),
  personalized problems (7–12)
- Form B: Nonpersonalized problems (7–12),
  personalized problems (1–6)
↓
Randomized by Pretest Score
- Higher ability
- Lower ability
↓
Treatments
- Personalized higher ability
- Personalized lower ability
- Nonpersonalized higher ability
- Nonpersonalized lower ability
↓
Attitude Survey (Student and Teacher)
↓
Posttest (Forms A and B)
- Form A: Nonpersonalized problems (1–6),
  personalized problems (7–12)
- Form B: Nonpersonalized problems (7–12),
  personalized problems (1–6)

Criterion Measures

Pretest. A total of 24 problems were developed, in both personalized and nonpersonalized forms, for the pretest and posttest: 12 items, 3 from each of the four combinations of multiplication and division math operations, were randomly assigned to the pretest; and 12 were assigned to the posttest. Thus, both the pretest and the posttest contained 12 two-step math word problems involving whole numbers. The problems on the pretest were in random order within each problem type; those on the posttest were in the same order as the pretest, by problem type. Each test consisted of 3 problems involving a multiplication operation followed by a second multiplication operation (multiply-multiply), 3 problems involving multiplication followed by division (multiply-divide), 3 problems involving division followed by multiplication

(divide-multiply), and 3 problems involving division followed by a second division operation (divide-divide).

The pretest was constructed and named as two different forms, Form A and Form B. The total of 24 items were randomly assigned as 12 pretest and 12 posttest items as described above. In Form A, problems 1–6 were written as nonpersonalized problems and problems 7–12 were written as personalized problems. In Form B, the nonpersonalized problems from Form A (problems 1–6 on Form A) were converted to the personalized problems 7–12, and the personalized problems from Form A (problems 7–12 on Form A) were converted to nonpersonalized problems 1–6. Thus, the same 12 items appeared in both personalized and nonpersonalized forms across the two test versions, with the 6 items in nonpersonalized form as items 1–6 on each form and the 6 personalized items as items 7–12 on each form. The answer to each problem was scored as correct only when the correct final answer to the problem was given. The KR-20 reliability coefficient was .83 for the pretest.

Posttest. As with the pretest, posttest problems were developed and administered in two forms. The first six problems on Form A of the posttest were in nonpersonalized form and the second 6 problems were in personalized form. The first six problems on Form B of the posttest were items 7–12 from Form A in nonpersonalized form and the final six problems on Form B (Items 7–12) were items 1–6 from Form A in personalized form. Thus, the problems on Form A and Form B of both the pretest and the posttest consisted of six nonpersonalized items followed by six personalized items. Like the answers on the pretest, each answer on the posttest was scored as correct or incorrect only based on the final answer. The KR-20 reliability coefficient was .87 for the posttest.

Subjects who received Form A on the pretest received Form B on the posttest, and those who received Form B on the pretest received Form A on the posttest. The overall mean proportions correct for all subjects across both tests were .68 for Form A and .69 for Form B.

Student attitude survey. A 10-item attitude sur-

vey served as the criterion measure for assessing student attitudes and motivation. Of the 10 items, 8 were four-choice Likert-type questions that assessed student attitudes and continuing motivation toward the instruction. These items dealt with such matters as how interesting and how easy the instructional program was, how much students learned from it, whether students could do two-step problems well, and whether they would like to do more problems like those in the program. Responses to these 8 items were assigned a score of 4 for the *most positive* response and a score of 1 for the *least positive* response. The 2 remaining items were open-ended questions dealing with student likes and dislikes about the instructional program. The KR-20 reliability coefficient for the 8 Likert-type items was .70.

Teacher attitude survey. A seven-item teacher survey, consisting of six four-choice Likert-type items and one open-ended question, was used to assess teacher attitudes toward the instructional program and the personalization strategy. Items dealt with topics such as the appropriateness of the program, its quality, whether it helped students learn, the effectiveness of personalization, and whether the teachers liked personalization. Teacher responses to the Likert-type items, like those of the students, were scored from 4 (*most positive*) to 1 (*least positive*).

Data Analysis

The data analysis for student achievement was a 2 (Treatment: Personalization and Nonpersonalization) × 2 (Ability Level: Higher Ability and Lower Ability) × 2 (Test Occasion: Pretest and Posttest) × 2 (Problem Type: Personalized and Nonpersonalized problems) repeated-measures univariate analysis of variance (ANOVA). Treatment and ability level were between-subjects variables in the analysis and test occasion and problem type were within-subjects variables. En route data on performance-on-practice items and on time in program were also collected and analyzed.

Attitude data were analyzed using a 2 (Treatment) × 2 (Ability Level) × 8 (Survey Items) mul-tivariate analysis of variance (MANOVA) for the overall survey means.

RESULTS

The results are reported in this section for achievement, en route practice, time in program, student attitudes, and teacher attitudes.

Achievement

The pretest and posttest results for the two levels of personalization, the two ability levels, the two test occasions, and the two problem types are shown in Table 1. The achievement data for each variable in the four-factor design are discussed below.

Treatment. For personalization level, the mean overall scores across the two test occasions were 8.45 (70%) for the personalized subjects and 7.93 (66%) for the nonpersonalized subjects. ANOVA revealed that the difference in mean scores between the two personalization levels was not statistically significant, $F(1, 132) = 2.37$, $MSE = 3.80$, $p = .126$. However, the difference for treatment in this analysis was reduced because the pretest and posttest scores were combined for each of the two levels of personalization owing to the repeated-measures design. The mean scores by personalization level on the pretest (i.e., prior to the treatment) were identical for the two levels: 6.54 items correct (55%) for the personalized treatment and 6.54 (55%) for the nonpersonalized treatment. In contrast, the mean posttest scores by treatment were 10.35 (86%) for the personalized treatment and 9.32 (78%) for the nonpersonalized treatment.

The differential pattern of identical scores for personalization level on the pretest, but a higher score for personalization than for nonpersonalization on the posttest, was reflected in a significant treatment by test occasion interaction, $F(1, 132) = 8.27$, $MSE = 1.09$, $p < .01$, $\eta^2 = .06$. A post hoc paired-samples t test revealed that the posttest score of 10.35 for the personalized treatment was significantly higher than the score of 9.32 for the nonpersonalized treatment,

Table 1 ☐ Mean scores by treatment, ability level, test occasion, and problem type

| | Problem Type | | | |
| | Personalized | | Nonpersonalized | |
Test Occasion and Treatment	Higher ability	Lower ability	Higher ability	Lower ability
Pretest				
Personalized				
M	4.79	1.94	4.71	1.65
SD	(0.98)	(1.28)	(1.09)	(1.28)
Nonpersonalized				
M	4.85	2.18	4.35	1.71
SD	(0.96)	(1.24)	(0.98)	(1.22)
Subtotals				
M	4.82	2.06	4.53	1.68
SD	(0.96)	(1.26)	(1.04)	(1.24)
Posttest				
Personalized				
M	5.82	4.59	5.85	4.44
SD	(0.46)	(1.84)	(0.44)	(1.74)
Nonpersonalized				
M	5.68	3.94	5.35	3.68
SD	(0.59)	(1.69)	(0.95)	(1.80)
Subtotals				
M	5.75	4.27	5.60	4.06
SD	(0.53)	(1.78)	(0.78)	(1.80)

Overall mean scores by test occasion and variables:

	Treatment			Ability			Problem Type		
Pretest	Personalized	=	6.54	Higher	=	9.35	Personalized	=	3.44
	Nonpersonalized	=	6.54	Lower	=	3.74	Nonpersonalized	=	3.10
Posttest	Personalized	=	10.35	Higher	=	11.35	Personalized	=	5.01
	Nonpersonalized	=	9.32	Lower	=	8.32	Nonpersonalized	=	4.83
Overall	Personalized	=	8.45	Higher	=	10.35	Personalized	=	4.23
	Nonpersonalized	=	7.93	Lower	=	6.03	Nonpersonalized	=	3.97

Note. Total possible score equals 12 items correct on each test for treatment, ability level, and test occasion, and six items correct for each problem type.

$t(67) = 3.01$, $p < .01$. The treatment by test occasion interaction is diagrammed in Figure 2.

Ability level. Higher-ability students outscored lower-ability students across the two tests, 10.35 (86%) to 6.03 (50%). This difference for ability level was statistically significant, $F(1, 132) = 167.49$, $MSE = 3.80$, $p < .001$, $\eta^2 = .56$.

The $2 \times 2 \times 2 \times 2$ ANOVA also yielded a significant ability level by test occasion interaction, $F(1, 132) = 52.30$, $MSE = 1.09$, $p < .001$, $\eta^2 = .28$. This interaction reflected the fact that higher-ability students improved less from pretest to posttest than did lower-ability students. Higher-ability subjects had mean scores of 9.35 (78%) on the pretest and 11.35 (95%) on the posttest, an improvement of 2.00 items correct, whereas lower-ability subjects had mean scores of 3.74 (31%) on the pretest and 8.32 (69%) on the posttest, an improvement of 4.58 items correct.

Test occasion. The mean scores for test occasion were 6.54 (55%) for the pretest, and 9.84 (82%) for the posttest, a mean pretest-to-posttest increase of 3.30 items correct. This difference was statistically significant, $F(1, 132) = 338.85$, $MSE = $

Figure 2 ☐ Treatment by test occasion
interaction
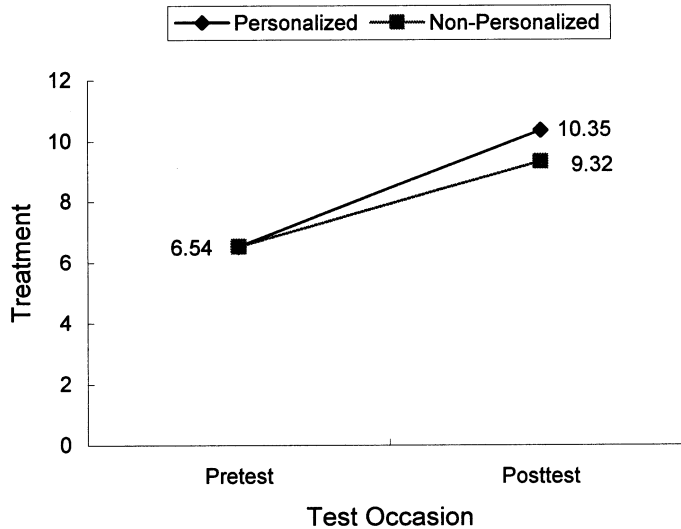
1.09, $p < .001$, $\eta^2 = .72$. The two interactions that involved test occasion, treatment by test occasion and ability by test occasion, were reported above.

Problem type. The overall mean scores for problem type were 4.23 (71%) for the six personalized problems and 3.97 (66%) for the six nonpersonalized problems, a statistically significant difference, $F(1, 132) = 17.63$, $MSE = .51$, $p < .001$, $\eta^2 = .12$. The ANOVA also yielded a significant two-way interaction for treatment by problem type, $F(1, 132) = 4.66$, $MSE = .51$, $p < .05$, $\eta^2 = .03$. This interaction reflected the fact that there was a rather small difference in the scores of subjects in the personalized treatment on the personalized and nonpersonalized problems (4.29, or 72%, on personalized items and 4.16, or 69%, on nonpersonalized items), and a larger difference in the scores of nonpersonalized subjects on these items (4.16, or 69%, on personalized items and 3.77, or 63%, on nonpersonalized items). Post hoc paired-samples *t* tests of the scores contributing to this interaction revealed that the mean score of 3.77 for nonpersonalized treatment on nonpersonalized items was significantly lower at the $p < .01$ level than the means of each of the other three groups.

Both personalized and nonpersonalized problems were included in the pretest in order to determine whether students would perform better on personalized word problems than on nonpersonalized problems prior to the instruction. The significant main effect for problem type across test occasions indicates that subjects did perform better on personalized items than on nonpersonalized items both before and after they received instruction. Paired-sample *t* tests of the difference between scores on personalized and nonpersonalized items on the pretest and again on the posttest confirmed that significant achievement differences favoring personalized items occurred on both occasions.

Practice

Performance on the enroute practice items was analyzed by treatment and ability level to determine whether differences occurred in subject achievement on the eight practice problems. The overall mean score across all subjects on the eight practice items was 6.62 (83%), very similar to the overall posttest score of 82%. The mean scores on eight practice items were 6.84 (86%) for the personalized treatment and 6.40 (80%) for the nonpersonalized treatment, a nonsignificant difference for treatment. The practice

mean scores by ability were 7.50 (94%) for higher-ability subjects and 5.74 (72%) for lower-ability subjects, a statistically significant difference, $F(1, 132) = 32.26$, $MSE = 3.28$, $p < .001$, $\eta^2 = .20$.

Time in Program

The overall mean number of minutes across all subjects was 32.83. The mean times by treatment were 32.06 min for personalization and 33.60 min for nonpersonalization, a nonsignificant difference, $F(1, 132) = 2.81$, $MSE = 28.81$, $p = .096$. Mean times for ability level were 30.65 min for higher-ability subjects and 35.02 min for lower-ability subjects, a statistically significant difference, $F(1, 132) = 22.52$, $MSE = 28.81$, $p < .001$, $\eta^2 = .15$. The additional time spent on the experimenter-led review section was approximately 20 min for each class.

Student Attitudes

The mean attitude scores by treatment and ability level for subject responses to the eight statements on the four-point Likert-type attitude survey administered after completion of the instructional program are shown in Table 2. Responses were scored as 4 for the *most positive* response to 1 for the *most negative* response.

The overall mean score across the eight student attitude survey items was 3.42, a favorable rating indicating agreement with positive statements about the instructional program. The three highest-rated statements on the survey were "I learned a lot from this program" ($M = 3.71$), "It is important to know how to solve two-step math problems" ($M = 3.68$), and "I would like to do more math word problems like the ones in the program" ($M = 3.55$). The lowest-rated statement was "I am able to do two-step math problems well" ($M = 3.12$).

The data in Table 2 were analyzed using a 2 (Treatment) × 2 (Ability) × 8 (Survey Items) MANOVA to test for significant differences. A significant overall effect across the eight items was obtained for treatment, ($M = 3.52$ for personalization and 3.31 for nonpersonalization), $F(8, 127) = 10.83$, $MSE = .13$, $p < .001$, $\eta^2 = .08$, and for ability, ($M = 3.54$ for higher-ability stu-

dents and 3.29 for lower-ability students), $F(8, 127) = 16.00$, $MSE = .13$, $p < .001$, $\eta^2 = .11$. The treatment by ability level interaction across the eight items was not statistically significant.

Univariate analyses on the eight survey items by personalization level revealed significantly more positive attitudes on four of the items for subjects in the personalized treatment than for those in the nonpersonalized treatment. Students in the personalized treatment had significantly more favorable scores at the $p < .001$ level on the items: "This program was interesting" ($M = 3.62$ for personalization and 3.24 for nonpersonalization, $\eta^2 = .09$), "This program was easy" ($M = 3.50$ for personalization and 2.85 for nonpersonalization, $\eta^2 = .19$), and "This program had many familiar persons, places, and things" ($M = 3.62$ for personalization and 3.13 for nonpersonalization, $\eta^2 = .11$), and at the .05 level on the item "I would like to do more math word problems like the ones in the program" ($M = 3.66$ for personalization and 3.44 for nonpersonalization, $\eta^2 = .04$).

Univariate analyses for ability level revealed significant attitude differences favoring higher-ability students on five of the items. Higher-ability students responded significantly more positively than did lower-ability subjects at the .001 level to the statements: "I learned a lot from this program" ($M = 3.91$ for higher ability and 3.51 for lower ability, $\eta^2 = .14$), and "I am able to do two-step math problems well" ($M = 3.35$ for higher ability and 2.88 for lower ability, $\eta^2 = .11$), and at the .05 level to the statements: "This program was easy" ($M = 3.32$ for higher ability and 3.03 for lower ability, $\eta^2 = .04$), "I liked this program" ($M = 3.44$ for higher ability and 3.12 for lower ability, $\eta^2 = .05$), and "I would like to do more math word problems like the ones in the program" ($M = 3.68$ for higher ability and 3.43 for lower ability, $\eta^2 = .05$).

The frequency of constructed responses on the attitude survey to the two open-ended questions about what students liked most and what they liked least was also tabulated. Student responses indicated that what they liked most was that the program was interesting, a response given by 42 of the 136 students (31%). The second most common response to what students liked most was the use of the names of car-

Table 2 ☐ Student attitude scores by treatment and ability level

| | Treatment | | | Ability | | |
Item	Personal-ized	Non-personal-ized	F	Higher	Lower	F
1. It is important to know how to solve two-step math problems.	3.65	3.72	.57	3.72	3.65	.57
2. This program was interesting.	3.62	3.24	13.24**	3.40	3.46	.29
3. This program was easy.	3.50	2.85	31.00**	3.32	3.03	5.41*
4. This program had many familiar persons, places, and things.	3.62	3.13	15.81**	3.49	3.26	2.99
5. I learned a lot from this program.	3.69	3.74	.22	3.91	3.51	20.85**
6. I am able to do two-step math problems well.	3.16	3.07	.52	3.35	2.88	16.65**
7. I liked this program.	3.28	3.28	.00	3.44	3.12	6.29*
8. I would like to do more math word problems like the ones in the program.	3.66	3.44	4.81*	3.68	3.43	6.26*
Overall means	3.52	3.31	10.83**	3.54	3.29	16.00**

*$p < .05$. **$p < .01$.

toon characters, persons, and things with which they are familiar, a response indicated by 38 students (28%). When asked what they liked least about the program, 48 students (35%) indicated nothing and 33 students (24%) responded that the lesson was too difficult.

Teacher Attitudes

The overall mean teacher rating on the six items on the teacher survey was 3.71, a favorable rating indicating strong agreement with positive statements about the instructional program. All four respondents agreed very strongly ($M =$ 4.00) with three statements: "Personalization was a good teaching strategy", "I liked the personalized version of the program", and "I would enjoy teaching a personalized lesson to my students occasionally."

On the one open-ended question on the survey, "Please make any comments or suggestions that you would like to make about this instructional program," two teachers indicated that making math word problems more personalized increases student motivation and interest. One teacher reported that the students would understand the personalized problems better because they could relate the information in the problems to their real-life situations and "shorten the distance" of their thinking patterns on the problems.

DISCUSSION

The primary research question in this study addressed the issue of whether or not personalization of instruction increases the achievement of Taiwanese students on mathematics word problems. The treatment by test occasion interaction revealed that the personalized treatment did, in fact, result in significantly higher pretest-to-posttest gains than the nonpersonalized treatment. This positive finding for personalization of instruction is consistent with the results obtained in several studies of personalized mathematics instruction in the United States (Anand & Ross, 1987; Lopez & Sullivan, 1991, 1992). It differs from the overall result for personalization in the earlier study by Ku and Sullivan (2000) in Taiwan, in which a ceiling effect limited the

potential gain of higher-ability fifth graders.

The strategy of conducting this study with fourth-grade students, instead of fifth graders as in the earlier study, had the desired result of reducing the strong ceiling effect observed previously (Ku & Sullivan, 2000). The overall pretest score of 55% in this study was considerably lower than the pretest score of 70% in the earlier research. The posttest score of 82% in the current study also was below the posttest score of 87% with fifth graders. The lower pretest score left more room for maximizing the potential differential effectiveness of the two personalization treatments, and surely was a factor that contributed to the significant achievement difference for personalization in this study after such a difference had not been obtained previously.

Despite the successful effort to reduce the ceiling effect with higher-ability students in this study, the significant ability level by test occasion interaction revealed that this effect was not completely eliminated. Higher-ability subjects averaged 95% (11.35 of 12 correct) on the posttest. This represented only 17% gain over their pretest score of 78%, but nevertheless approached the maximum possible score on the test. In contrast, lower-ability subjects were able to make a much higher gain, due in part to their much lower pretest score of only 31%.

The significant overall improvement from 55% on the pretest to 82% on the posttest indicated that the instructional program itself was generally effective across all treatment groups. This increase in scores occurred over an average time period of only about 53 min of instruction and review. In addition, the fact that students averaged 83% on the practice items and 82% on the posttest indicates that they retained their learning quite well from the instructional phase of the study to the final testing phase.

The finding that students scored higher on personalized test items than on nonpersonalized items, even prior to instruction (i.e., on the pretest), has potential implications for mathematics assessment. Certainly this finding merits further investigation in other settings and with larger samples to determine its consistency and generality. However, it suggests that students generally may score higher on math problems that include more familiar or personalized settings than on problems with unfamiliar or nonpersonalized settings. If that is the case, test developers who write standardized and criterion-referenced mathematics tests may want to consider the appropriateness of using settings for their problems that are generally familiar or popular with the projected test population.

A potential concern about relying more heavily on familiar problem settings in mathematics instruction and assessment is that it may reduce transfer of learning to general or less familiar settings. The present results for nonpersonalized test items do not support this concern. Prior to instruction (i.e., on the pretest), subjects who ultimately were in the personalized treatment and those who ultimately were in the nonpersonalized treatment had similar scores (53% and 51% respectively) on the nonpersonalized test items. Following instruction (on the posttest), subjects who had received personalized instruction scored considerably higher (86% to 75%) on nonpersonalized test items than did those who had received nonpersonalized instruction. This higher posttest score for subjects in the personalized treatment on nonpersonalized test items came under conditions that could be considered to require greater transfer of learning for them than for their nonpersonalized counterparts, who practiced only nonpersonalized items during their instruction.

The attitude data clearly indicated student preference for personalized instruction, a result consistent with the findings in previous studies (Ku & Sullivan, 2000; Lopez & Sullivan, 1992; Ross & Anand, 1987). The strongest differences, all at the .001 level, were on items stating that the program was interesting; the program had many familiar persons, places, and things; and the program was easy. The first two of these statements reflect the intended nature of a personalized program and the third is consistent with the explanation that personalization may make learning easier by reducing subjects' cognitive load. Personalized subjects also agreed significantly more strongly ($p < .05$) that they would like to do more math word problems like the ones in the program, a statement that sug-

gests greater continuing motivation or willingness to return to task on their part. In general, the significant differences on these particular items support the claimed motivational and learning advantages of personalization.

Several significant differences between ability levels on the attitude survey also reflected differences that one might expect between higher-ability and lower-ability students. Higher-ability students agreed more strongly that they learned a lot from the program ($p < .001$), that they were able to do two-step math problems well ($p < .001$), and that the program was easy ($p < .05$). These items appear to be the ones in the survey that were most closely associated with ability level. In contrast, higher-ability and lower-ability subjects did not differ significantly in their responses to items that logically seem to be less associated with ability.

Teacher attitudes were favorable toward the instructional program and were especially positive toward the personalized version. Supporting the positive attitude survey data was the fact that, at the end of the study, all four teachers asked the experimenter to give them copies of both versions of the instructional program and mentioned that they would like to try a similar personalization strategy in their regular classroom teaching. In addition, one teacher provided the experimenter with a copy of the newest Taiwanese education reform document, titled "The Basic Theory and Practice for Nine-Year Schooling with Coherent Curricula." She indicated that the personalization strategy would fit very well with the new educational reform policy in Taiwan, which states that elementary school teachers, instead of teaching from textbooks, will have the opportunity to design and implement their own lesson plans with fewer students per class.

This study was not designed as a test of theoretical explanations for the effects of personalization. However, certain data from the teacher and student surveys tend to support the increased-interest and reduced-cognitive-load explanations described in the introductory section of this article. In open-ended comments, two teachers reported that personalizing math problems increases student motivation and interest. Student responses also indicated that they thought the personalized program was significantly more interesting than the nonpersonalized one and that the greater interest was what they liked most about the program. The teacher comment that students could better relate the information in the personalized problems to their real life and thus shorten the distance of their thinking patterns on the problems is consistent with the reduced-cognitive-load explanation. This explanation could also help to account for the significantly stronger agreement by subjects under personalization with the statement that the problem was easy.

The present results were obtained using group personalization in a low-technology environment that is typical of the schools in Taiwan and in many other countries. A basic use of personalization, or interest-based instruction, in such settings is for teachers to make a conscious effort to learn their students' interests and to incorporate them regularly into their instruction. Teachers can supplement their existing knowledge of student interests by using an interest survey, as was done in the present study, or by holding occasional discussions with their students about current topics and events that may be popular with them.

Technology-rich environments, of course, offer greater potential for the design of instruction based on personal information provided by the target audience. Students and trainees in high-technology settings can enter their relevant personal information (e.g., interests, department or type of job, preferred learning structure) into a general database or one created for a particular instructional or training unit. This information can be incorporated directly into computer-based instruction, computer-based training, or on-line learning to personalize the instruction for each individual or for groups based on factors such as their common interests, work responsibilities, or instructional preferences. The potential of such technology-based applications has long been recognized, but little research has been reported on their effects. Future research that includes technology-based personalization efforts can help us understand how we can best use learner interests and other learner characteristics to design more effective instruction and training. □

Heng-Yu Ku is an Assistant Professor in the Educational Technology Program, College of Education, at the University of Northern Colorado, Greeley. This research was conducted while he was a doctoral student at Arizona State University.

Howard J. Sullivan is a Professor in the Educational Technology Program, Division of Psychology in Education, at Arizona State University, Tempe.

REFERENCES

Anand, P.G., & Ross, S.M. (1987). Using computer-assisted instruction to personalize arithmetic materials for elementary school children. *Journal of Educational Psychology, 79*(1), 72–78.

Bransford, J.D., Sherwood, R.D., Hasselbring, T.S., Kinzer, C.K., & Williams, S.M. (1992). Anchored instruction: Why we need it and how technology can help. In D. Nix & R. Spiro (Eds.), *Cognition, education, and multimedia* (pp. 115–141). Hillsdale, NJ: Erlbaum.

Brown , J.S., Collins, A.S., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42.

Cheng, S.-S., Chang, W.-H., Chiang, W.-H., & Guo, C.-J. (1998). *Development of a professional development program for science and mathematics teachers—An action research.* Paper presented at the annual meeting of the National Association for Research in Science Learning, San Diego, CA.

Choi, J.-I., & Hannafin, M. (1995). Situated cognition and learning environments: Roles, structures, and implications for design. *Educational Technology Research and Development, 43*(2), 53–69.

Cordova, D.I., & Lepper, M.R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88*(4), 715–730.

d'Ailly, H.H., Simpson, J., & MacKinnon, G.E. (1997). Where should "you" go in a math compare problem? *Journal of Educational Psychology, 89,* 562–567.

Davis-Dorsey, J., Ross, S.M., & Morrison, G.R. (1991). The role of rewording and context personalization in the solving of mathematical word problems. *Journal of Educational Psychology, 83*(1), 61–68.

Dwyer, H.J. (1996). Effect of personalization on reading comprehension. (Doctoral dissertation, Arizona State University, 1996). *Dissertation Abstracts International, 59,* Z3810.

Enright, B.E., & Choate, J.S. (1993). Mathematical problem solving: The goal of mathematics. In J.S. Choate (Eds.), *Successful mainstreaming: Proven ways to detect and correct special needs* (pp. 283). MA: Allyn and Bacon.

Hart, J.M. (1996). The effect of personalized word problems. *Teaching Children Mathematics, 2*(8), 504–505.

Herndon, J.N. (1987). Learner interests, achievement, and continuing motivation in instruction. *Journal of Instructional Development, 10*(3), 11–14.

Ku, H.-Y., & Sullivan, H.J. (2000). Personalization of mathematics word problems in Taiwan. *Educational Technology Research and Development, 48*(3), 49–59.

Lopez, C.L., & Sullivan, H.J. (1991). Effects of personalized math instruction for Hispanic students. *Contemporary Educational Psychology, 16*(1), 95–100.

Lopez, C.L., & Sullivan, H.J. (1992). Effect of personalization of instructional context on the achievement and attitudes of Hispanic students. *Educational Technology Research and Development, 40*(4), 5–13.

Mayer, R.E. (1982). Memory for algebra story problems. *Journal of Educational Psychology, 74*(2), 199–216.

Mayer, R.E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science, 26*(2), 49–63.

Miller, D.C., & Kulhavy, R.W. (1991). Personalizing sentences and text. *Contemporary Educational Psychology, 16*(3), 287–292.

Ministry of Education. (1998). *Blue books for education reform: Coherent curricula from grade 1 through grade 9.* The Ministry of Education, Taipei, Taiwan, R.O.C.

National Assessment of Educational Progress. (1992a). *NAEP 1992 mathematics report card for the nation and the states.* Washington, DC: National Center for Education Statistics. Report No. 23-ST02.

National Assessment of Educational Progress. (1992b). *NAEP 1992 school effects on educational achievement in mathematics and science.* Washington, DC: National Center for Education Statistics, Research and Development Report.

Rosen, D.R. (1984). *Students' schemata for algebra word problems.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Ross, S.M. (1983). Increasing the meaningfulness of quantitative material by adapting context to student background. *Journal of Educational Psychology, 75*(4), 519–529.

Ross, S.M., & Anand, P.G. (1987). A computer-based strategy for personalizing verbal problems in teaching mathematics. *Educational Communication and Technology Journal, 35*(3), 151–162.

Ross, S.M., McCormick, D., & Krisak, N. (1986). Adapting the thematic context of mathematical problems to student interests: Individualized versus group-based strategies. *Journal of Educational Research, 79*(4), 245–252.

Ross, S.M., McCormick, D., Krisak, N., & Anand, P.G. (1985). Personalizing context in teaching mathematical concepts: Teacher-managed and computer-assisted models. *Educational Communication and Technology Journal, 33*(3), 169–178.

U.S. Department of Commerce. (1999). *Statistical abstract of the United States: The National Data Bank.* Economics and Statistics Administration and U.S. Census Bureau.