Running head: EFFECTS ON LANGUAGE PROFICIENCY

The Effects of Motivation and Strategy Instruction on the

Language Proficiency of

Beginning Spanish Language Learners

Melissa S. Ferro

EDRS 811: Quantitative Methods in Educational Research

Dr. Dimiter Dimitrov

George Mason University

December 11, 2007

Introduction

The tragic events of September 11, 2001 renewed interest in foreign/world language education by a variety of public institutions. From the Department of Defense to the Department of Education, the call for new generation of global citizens had been made. One of the questions for educators, and specifically for those in the foreign/world language education, was how to provide the best possible instructional practices that would facilitate the development of communicative skills and cross-cultural understanding with beginning level foreign/world language students.

The research surrounding instructional practices in the field of foreign/world languages had gained momentum in the last several decades. In a quest to better understand the gaps in achievement among beginning level language learners many researchers have sought to identify and study factors that may affect the development of their language proficiency. In a brief review of the literature, it became evident that a dichotomy has emerged between those researchers who have focused on cognitive factors such as language learning disabilities (Dinklage, 1971; Sparks & Ganschow, 1993) and others who have studied potential psychological factors such as anxiety (Horwitz, Horwitz & Cope, 1986), motivation (Gardner, 1985; Gardner, Tremblay, & Masgoret, 1997) and choice of learning strategies (Oxford & Nyikos, 1989; Oxford, 2001).

The majority of the previous research has focused on the development of instruments to identify cognitive and psychological variables in second language learning and then to measure the interaction of these variables under certain conditions. There is, however, a paucity of research that examines the effects of variables such as motivation, anxiety, and choice of learning strategy on student performance. One recent study (Huang, 2001) conducted on Taiwanese English language learners examined the effect of learning strategy instruction on

beginning level English as a foreign language (EFL) learners. In that study, Huang used the Test of English as a Foreign Language (TOEFL), a standardized test, for measuring proficiency in listening, reading and writing in English. Other instruments included the Foreign Language Classroom Anxiety Scale (FLCAS), the Motivational Intensity Questionnaire (MIQ), and the Strategy Inventory for Language Learning (SILL). The results of that study showed that the students who did not receive learning strategy instruction as part of their regular classroom activities did not have significant gain scores on the TOEFL. However, students who did receive language strategy instruction showed significant improvement in their English language proficiency.

The results of the Huang study show that language strategy instruction is beneficial to some learners. However, there is a need for additional empirical data with beginning level foreign/world language learners. Therefore, the purpose of the present study is to determine the effects of the explicit instruction for language learning strategies on the proficiency of beginning level language learners. Based on previous research, this study also seeks to determine if the differences in gain scores can be explained by the interaction between the students' levels of motivation and whether or not they received strategy instruction.

Research Questions

This study specifically asks the following research questions:

1. Are student motivation, initial use of learning strategies, and pre-test language proficiency scores statistically significant predictors of post-test language proficiency scores?

2. Is there a difference between experimental and control groups on the post-test scores, controlling for pre-test scores?

3. Is the gain score for experimental and control groups the same across the levels of student motivation?

4. Do ethnic groups gain equally from pre-test to post-test scores or are there differences among ethnic groups on post-test scores, controlling for pre-test?

Methods

Design

This study is a quasi-experimental between-groups design. The student sample is an unbalanced design as the researcher could not deviate from the university's course registration procedures. As a result, random sampling at the student level could not be established, class sizes varied, and equal distribution of students by ethnic group was not possible. However, the researcher was able to randomly assign classes to the experimental and control groups. Prior to the collection of data, approvals were obtained from the university's Human Subjects Research Board.

Participants

The participants of this study included 696 students and 30 instructors from a large Mid-Atlantic University. Student participants had individually and independently enrolled in one of the beginning Spanish courses (level 100-200) in the spring 2007 semester using the university's online registration web site. The instructor participants were based on a convenience sample of those who taught 100-200 level Spanish courses and who voluntarily offered to be a part of the study.

Student Sample. The total student sample (N = 696) included several subgroups. First, classes were randomly assigned to experimental and control groups. The experimental group consisted of 20 instructors and 457 students. The control group consisted of 10 instructors and

239 students. The average class size was 22.8 and 23.9 for the experimental and control groups respectively. Second, there were four subgroups based on ethnicity. The ethnic subgroups consisted of 239 Caucasians, 136 African Americans, 158 Hispanics, and 163 Asians. Initial analysis of frequencies revealed that data on gender was incomplete. Therefore, subgroups by gender could not be established.

Data Collection Instruments

There were three primary instruments used to collect data in this study. Demographic information on each participant was collected by using a short questionnaire that was attached to each of the following instruments.

CLEP. To measure language proficiency, the researcher used the College Level Examination Program (CLEP) for Spanish. The CLEP for Spanish is used to evaluate the proficiency level of students whose native language is not Spanish. It is the test used by many post-secondary institutions to determine if a student's proficiency score in reading, listening, and writing in a foreign/world language qualifies for college credit. In this study, it was used to assess the participants' Spanish language proficiency at both the beginning and end of the semester. The College Board reports that CLEP examinations are studied for criterion-related evidence of validity by comparing multiple data sources of student knowledge on the subject matter. It is also tested for internal reliability using Kuder-Richardson 20 (KR-20).

SILL. The Strategy Inventory for Language Learning (SILL) was used to measure the initial frequency of strategy use by the learners prior to the intervention. The SILL consists of 50 items that ask the participants to rate their responses based on a five-point scale from "almost never" to "almost always." Oxford and Nyikos (1989) report that the SILL has an internal

reliability with Cronbach's alpha = .956. This reliability coefficient is based on one study with 1,200 participants and a second study with 750 participants.

MIQ. The Motivational Intensity Questionnaire (MIQ) was used to determine the students' level of motivation in learning the subject matter. The MIQ contains 10 multiple choice questions that are then rated on three point scale of "low", "average", and "high." Previous research conducted on the MIQ reports that it has an internal reliability of Cronbach alpha = .78 (Hsiao, 1997).

Data Collection Procedures

Instructors and their classes were randomly assigned to either the control or experimental group. The intervention in this study was the use of explicit language learning strategy instruction with the student participants in the experimental group. Instructors who taught classes in the experimental group received training during a one-day workshop at the beginning of the semester. The researcher provided each experimental group instructor with specific materials that were used for the instruction of language learning strategies. These strategies focused on the development of reading and writing skills as well as the retention of level appropriate vocabulary. Student participants in the experimental group received approximately 20 minutes of explicit language learning strategy instruction per week. Instructors in the control group taught the same lessons and followed the same syllabus as the experimental group with the exception that they did not provide explicit language learning strategy instruction. The researcher frequently visited classrooms to ensure that procedures were adhered to by the instructors in both groups.

At the beginning of the semester, the researcher administered the CLEP, the MIQ, and the SILL to all participants in both groups. The results were recorded and stored using the

Statistical Package for Social Sciences (SPSS) software. Specifically, the researcher recorded the initial CLEP for Spanish scores as pre-test scores. During the 14-week semester, the participants in the experimental group continued to receive explicit instruction on language learning strategies. At the end of the semester, the researcher re-administered the CLEP for Spanish exam to both groups. These scores were then added to the SPSS files as post-test scores.

Statistical Data Analysis

The purpose of the first research question is to determine to what extent the students' motivation (MIQ), pre-test CLEP scores (PT), and frequency of learning style strategies (SILL) are good predictors of post-test scores (PTS) for future samples of beginning language learners. To address this question, multiple regression analysis on the three predictor variables was conducted. It is important to remember that the correlation of predictor variables does not necessarily lead to conclusions of causation (Dimitrov, 2008). One should not assume that any changes in pre-test scores, MIQ scores, or SILL scores would cause a change in post-test CLEP scores. The goal then was to determine the best-fit regression line that would allow for the use of this sample's data to predict future post-test CLEP scores for similar samples.

The second question regarding the differences between experimental and control groups on post-test scores while controlling for pre-test scores required the use of analysis of covariance (ANCOVA). This statistical test was specifically selected because it allows the researcher to compare adjusted group means while controlling for the effects of a confounding variable. In this particular question involving the comparison of the post-test adjusted means of the experimental and control groups, it is important to acknowledge that some variance in the post-test scores may have been due to pre-existing differences between the two groups (Balkin & Erford, 2008). To nullify any pre-existing differences among the groups, the researcher assigned the pre-test scores

as the covariate. By controlling for pre-test scores, the use of ANCOVA analysis allowed for the comparison of adjusted means among the post-test CLEP scores for the control group and the experimental group.

Along the same line of reasoning for the second research question, was the use of ANCOVA for the fourth research question. This question seeks to examine the differences in post-test scores among ethnic groups while controlling for pre-test differences. Here, the covariate was once again the pre-test CLEP scores, which was used to equalize any variance in post-test adjusted means that may be due to pre-existing differences among the ethnic groups. If the ANCOVA results had shown statistically significant differences in adjusted post-test mean scores among the four ethnic groups, then additional post hoc tests on adjusted means would have been conducted to show where these differences exist.

The third research question addresses whether or not there are differences in gain scores between the experimental and control groups across motivation levels. For this question, differences in the dependent variable, gain scores, were examined by using two independent variables; control-experimental groups, and motivation levels. Using factorial analysis of variance, or two-way ANOVA, the main effect of each independent variable was determined as well as the interaction between the two variables. Because the results of the two-way ANOVA show statistically significant differences in gain scores between the two groups across motivation levels, the Tukey post hoc tests on unadjusted means was necessary. As part of this post hoc analysis, a split-data t-test was used to help determine if the difference in gain scores between the control and experimental groups for each motivation level was significant.

Results

The descriptive statistics show that the mean SILL score for all student participants is 64.39 with a standard deviation of 22.36. The mean pre-test CLEP score for all student participants is 43.39 with a standard deviation of 15.62. The mean post-test score for all student participants is 63.35 with a standard deviation of 19.95. A split-data analysis of the mean and standard deviation shows that mean pre-test and the mean post-test CLEP scores for the control group are lower than the mean pre-test and the mean post-test scores for the total sample. Also, as one would expect, the mean pre-test and the mean post-test scores for the experimental group are higher than the mean pre-test and mean post-test scores for the total sample. Additional information on the split data descriptive statistics is provided in Table 1.

The frequency analysis shows that data are missing in the classification of student participants by gender. Therefore, the researcher could not perform any inferential statistical analysis that examined differences between males and females. However, the frequencies of both motivational levels and ethnic groups are based on the total sample. As expected, the majority (46.9 percent) of student participants had "average" motivational levels. Also, the majority (34.9 percent) of student participants were Caucasian. Additional information on the sub groups for motivation and ethnicity can be found in Table 2.

Prior to conducting the multiple regression analysis, the data set was cleaned to eliminate any outliers and any influential data points. Leverage h was used to determine the presence of outliers for variables on the x-axis. The leverage h cut score was computed, $\hat{h} = .017$. It was then compared to the values on the table for residual statistics. The leverage h minimum value of .000 and maximum value of .019 showed that there is at least one outlier on the x-axis, $h > .017$. The data set was scanned and one outlier was removed. Studentized deleted residual (SDR) was performed to identify the presence of outliers on the y-axis. The SDR minimum value of -4.91

and the maximum value of 1.315 showed that there is at least one outlier on the y-axis, $|SDR| >$ 3.00. The data set was scanned and 12 outliers were removed. Lastly, Cook's Distance (CD) was conducted to determine if any influential data points were present. The CD minimum value of .000 and maximum value of .100 show that there are no influential data points, $CD > 1.00$. The total sample size after cleaning the data was reduced to 683 student participants with degrees of freedom = 682.

To address the first research question regarding the prediction of post-test scores based on predictor variables, the results of the multiple regression analysis show that there is a high correlation among the actual scores for motivation (MIQ = 1), initial strategy use (SILL = 2), and pre-test proficiency scores (PT = 3), $R_{Y.123} = .979$. Additionally, 96 percent of the variability in post-test proficiency scores can be explained jointly by motivation, initial strategy use and pre-test scores. Finally, at the $\alpha = .05$ level, motivation, initial strategy use, and pre-test proficiency scores are statistically significant predictors of post-test proficiency scores, $F(3, 678) = 5247.48$, $p < .001$.

The regression line for the prediction of post-test scores (PTS) based on pre-test CLEP scores (PT), initial student use of learning strategies (SILL), and motivation level (MIQ) is $PTS = 1.23(PT) - 0.022(SILL) + 0.43(MIQ) + 10.49$. The interpretation of the slope indicates that an increase in pre-test (PT) by 1 point, holding all other variables constant, will increase post-test scores by 1.23 points. Also, an increase in initial student use of learning strategies (SILL) by 1 point will decrease post-test scores by 0.022 points. And, an increase in student motivation (MIQ) by one point will increase post-test scores by 0.43 points.

To determine if each variable has its own unique contribution, the results at the $\alpha = .05$ level show that pre-test proficiency is a statistically significant regression coefficient, $p < .001$;

strategy use is a statistically significant regression coefficient, p = .002; and motivation is not a statistically significant regression coefficient, p =.191. The relative importance of each predictor variable was compared using the absolute values of beta standardized coefficients. The beta results show that pre-test scores are the strongest predictor of post-test scores, β =.964. The beta results also show that strategy use, β =.025 and motivation β =.016 are weak predictors compared to pre-test scores.

Lastly, because the correlations among the predictors was high, $R_{Y.123}$ = .979, there may have been an issue of multicollinearity (Dimitrov, 2008). To address multicollinearity, the researcher examined the collinearity statistics for tolerance and variable inflation factors (VIF) for each predictor variable. The results show that there should not be serious concern for multicollinearity (e.g. all *VIFs* < 3). Additionally, the tolerance, or the degree of non-redundancy between one predictor variable and the other predictors, is higher for pre-test scores (Tolerance = .417, *VIF* = 2.40), and motivation (Tolerance = .423, *VIF* = 2.36) than for strategy use (Tolerance = .958, *VIF* = 1.04). Also, an analysis of a restricted model for predicting post-test CLEP scores using only pre-test scores and motivation was conducted to see if the change in R^2 between the restricted and full models was significant. The results show that the change in R^2 from the full to restricted model is not statistically significant at the α = .05 level, $F(1, 678)$ = 9.67, p =.002. Therefore, the use of the more parsimonious model is not recommended.

To address the second research question as to whether or not there a difference between experimental and control groups on the post-test scores, controlling for pre-test scores, two assumptions for using ANCOVA must be met. First, the results of the Pearson correlation test shows that pre-test is a good covariate for post-test, r =.979. However, at the α = .05 level ,the results for the interaction between experimental and control groups and pre-test scores showed

that the assumption of equal slopes is not met, $p < .001$. Therefore, the ANCOVA test could not be performed.

Similar results followed when addressing the fourth research question as to whether or not there is a difference between ethnic groups on the post-test scores, controlling for pre-test scores. Once again, the results of the Pearson correlation test shows that pre-test is a good covariate for post-test, r = .979. However, at the $\alpha = .05$ level, the results for the interaction between ethnic groups and pre-test scores showed that the assumption of equal slopes is not met, $p < .001$. Therefore, the ANCOVA test could not be performed and there are no further inferential statistics to support any findings for this research question.

The third research question examined whether or not the gain score for experimental and control groups was the same across all motivation levels. The results of the two-way ANOVA test with dependent variable of gain scores at the $\alpha = .05$ level show that there is a statistically significant main effect for motivation, $F(2, 676) = 198.56$, $p < .001$. Also, there is a statistically significant main effect for control-experimental groups, $F(1, 676) = 88.98$, $p < .001$. And, lastly, there is a statistically significant interaction between motivation and control-experimental groups, $F(2, 676) = 32.04$, $p < .001$. Full results are found in Table 3. Because statistically significant differences were found, post hoc tests were conducted.

The Tukey post hoc analysis on motivation at the $\alpha = .05$ level shows that here is a statistically significant difference between low and average levels of motivation, $p < .001$. Specifically, the 95% confidence interval shows that the gain scores for students with low motivation are higher than the gain scores for students with average motivation by at least 6.59 but not more than 8.24 points. Also, there is a statistically significant difference between low and high levels of motivation, $p < .001$. Specifically, the 95% confidence interval shows that the gain

scores for students with low motivation are higher than the gain scores for students with high motivation by at least 7.21 but not more than 9.10 points. Lastly, there is no statistically significant difference in gain scores between average and high levels of motivation, $p = .106$.

The graphic representation in Figure 1 provides visual evidence that the experimental group had higher gain scores than the control group for students with low and average motivation. From Figure 1, the question as to how significant the differences between groups for each individual motivation level was addressed by conducting a split-data t-test at the $\alpha = .05$ level. First, to test homogeneity of variance, the Levene's test shows that equal variances are assumed for low motivation, but not for average or high motivation. The results of the split-data t-test can be found in Table 4. The results show that there are statistically significant differences in the gain scores between the control and experimental groups for each individual level of motivation. Specifically, there are statistically significant differences between the control and experimental groups for students with low motivation, $t(186) = -8.79$, $p < .001$, average motivation, $t(139.61) = -9.98$, $p < .001$ and for high motivation, $t(40.96) = 2.31$, $p = .026$.

Discussion

The findings from the multiple regression analysis initially indicated that motivation, strategy use, and pre-test scores were strong predictors of post-test scores. However, a review of beta results showed that pre-test scores are the strongest predictor of post-test scores with strategy use and motivation as weaker predictor variables. This was not an anticipated result as the researcher had hoped to show that strategy use and motivation were also high predictors of post-test CLEP scores. With regards to these findings, one might ask if beginning level language instructors should include language learning strategy instruction in their daily lessons. The

question as to whether or not strategy instruction would improve student target language proficiency in reading, writing, listening, and vocabulary retention needed further investigation.

To address this question, an analysis of the gain scores did show that the differences between pre-test and post-test CLEP scores are higher for the experimental group—specifically for students with low or average levels of motivation. Therefore, there is statistically significant evidence that the explicit instruction of language learning strategies does make a difference on the target language proficiency of some beginning level language learners. Additional findings showed that students with high motivation in the experimental group actually scored lower on the post-test CLEP than students with high motivation in the control group. These results were not surprising as the researcher believes that students with high motivation would be the least affected by explicit learning strategy instruction. Two possible explanations are that students with high motivation may not have been affected by the intervention because they may have already been using these strategies. Also, it is very possible that the students with high motivation in the experimental group may have had a negative reaction to the use of class time for teaching strategies that they already knew. Considering the results of the split-data t-test, it appears that explicit language learning strategy instruction should be used especially for beginning level language learners with low or average motivation.

Limitations

There are several limitations that should be considered when drawing conclusions from the results of this study. Particular attention should be noted in the results of the two-way ANOVA and split-data t-test as some of the model assumptions were not met. First, the assumption of independence is questionable due to the unbalanced design. Second, the results of the Levene's test show that some of the variances among motivation levels are not equal. Third,

although normal distribution of scores on the CLEP, SILL and MIQ is assumed, this assumption is based on previous use of these instruments in similar studies.

Since the assumptions of independence and homogeneity of variances were not completely met, the results of the t-test and two-way ANOVA test are compromised as there is a higher likelihood for making a type I error (Balkin & Erford, 2008). However, the random assignment of classes to the control and experimental groups does eliminate some of the risk of stating there was a significant difference in gain scores when there really was not. And since homogeneity of variance was met for low motivation level, the results of the two-way ANOVA and the split-sample t-test do provide some statistical evidence in addressing the third research question on the differences among gain scores between groups across motivation levels.

Recommendations for Future Research

Based on the findings and discussion, there is a call for additional research that would replicate and expand this study. The current data from the Department of Education show that there is an achievement gap among minority students in the field of world language education. To help close this gap, future studies should specifically investigate the effect of learning strategy instruction on the language proficiency of minority students. Also, one cannot ignore the findings in this study regarding students with high levels of motivation. The question as to whether or not language strategy instruction negatively impacts these students remains unanswered. It would not be prudent or advisable to implement a change in instruction that provides benefits to some learners at the expense of others. Therefore, there is a need for additional research that focuses on this particular population of beginning level language learners.

References

Balkin, R., & Erford, B. T. (2008). Univariate inferential statistics. In Bradley T. Erford (Ed.), *Research and evaluation in counseling* (pp. 405-430). Boston: Houghton Miffin.

Dimitrov, D. (2008). Correlation and regression. In Bradley T. Erford (Ed.), *Research and evaluation in counseling* (pp. 405-430). Boston: Houghton Miffin.

Dinklage, K. T. (1971). The inability to learn a foreign language. In G. Blaine & C.C. McArthur (Eds.), *Emotional problems of the student* (pp. 185-206). New York: Appleton Century-Crofts.

Gardner, R.C. (1985). *Social psychology and second language learning: The role of attitudes and motivation.* London: Edward Arnold.

Gardner, R.C., Tremblay, P.F., & Masgoret, A.M. (1997). Towards a full model for second language learning: An empirical investigation. *Modern Language Journal, 81 (3)*, 334-362.

Horwitz, E.K., Horwitz, M.B., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal, 70 (2),* 125-132.

Huang, S.G. (2001). *Effects of language learning strategy training on English learning.* (ERIC Document Reproduction Service No. ED461287).

Oxford, R. (2001). Language learning styles and strategies. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (3rd ed.). Boston: Heinle & Heinle.

Oxford, R. & Nyikos, M. (1989). Variables affecting choice of language learning strategies by university students. *Modern Language Journal, 73 (3),* 291-300.

Sparks, R.L. & Ganschow, L. (1993). The impact of native language learning problems on

foreign language learning: Case study illustrations of the linguistic coding deficit

hypothesis. *Modern Language Journal*, 77 (1), 58-74.

Table 1

Mean and Standard Deviation for Pre-test CLEP and Post-test CLEP by Group

Group	Pre-test		Post-test	
	Mean	SD	Mean	SD
Control	31.11	10.52	47.13	16.59
Experimental	49.97	13.84	72.04	15.72
Total Sample	43.39	15.62	63.35	19.95

Table 2

Frequency by Ethnic Group and Motivation Level

Sub group	Frequency	Frequency Percentage
Students (N = 682)		
Motivation		
Low	188	27.6
Average	320	46.9
High	174	25.5
Students (N = 682)		
Ethnic Group		
African American	134	19.6
Hispanic	153	22.4
Asian	157	23.0
Caucasian	238	34.9

Table 3

Two-way Analysis of Variance for Gain Scores

Source	df	MS	F	p
Between- Subjects				
Control-Experimental Group (G)	1	1298.64	88.98	.000
Motivation (M)	2	2897.96	198.56	.000
G x M	2	467.71	32.04	.000
S within-group error	676	14.60		

*p < .05

Table 4

Split-data T-test on Gain Scores by Motivation Level

Motivation Level	Levene's Test	*t*	*df*	*p*	*Mean Difference*
Low	.096	-8.79	186	.000	-6.89
Average	.000	-9.98	139.61	.000	-4.54
High	.015	2.31	40.96	.026	1.09

*p < .05

Figure 1

Interaction between Control and Experimental Groups and Motivation Levels on Gain Scores