The Use of the Johnson-Neyman Confidence Bands and Multiple Regression Models to Investigate Interaction Effects: Important Tools for Educational Researchers and Program Evaluators

John W. Fraas, Ashland University
Isadore Newman, The University of Akron

When investigating the impact of predictor variables on an outcome variable or measuring the effectiveness of an educational program, educational researchers and program evaluators cannot ignore the possible influences of interaction effects. The purpose of this paper is to present a procedure that educational researchers can follow in order to increase their understanding of the nature of the interaction effect between a dichotomous treatment variable and a continuous independent variable. This technique involves the use of three separate analytical techniques implemented in three steps. First, the interaction effect is statistically tested using a multiple regression model. Second, the interaction effect is plotted, and if the interaction effect is disordinal, the intersection point of the regression lines is calculated. Third, the Johnson-Neyman confidence limits are calculated. A list of the computer commands that can be used in conjunction with the SPSS/PC+ Statistics™ and the SPSS® for Windows™ computer software to calculate the Johnson-Neyman confidence limits is provided. In addition, this three-step analytical procedure is applied to a set of efficacy data that was collected in a study of the FOCUS instructional model in order to illustrate how it can be used by researchers and program evaluators.

Most educational researchers and program evaluators are aware of the need to investigate the possible existence of interaction effects. When an interaction effect is being examined, a researcher or an evaluator must answer two questions. First, what analytical technique can be used to test for the presence of an interaction effect? Second, what analytical technique can provide the maximum amount of information regarding the interaction effect when, in fact, it exists? Researchers and evaluators often consider the first question. The second question, however, appears to be a consideration less often. To obtain an in-depth understanding of the interaction effect, the researcher or evaluator must utilize an analytical technique that can provide such information. That is, the researcher must avoid a Type VI error (Newman, Deitchman, Burkholder, Sanders, & Ervin, 1976), which occurs when the analytical technique does not provide the appropriate or necessary information.

In this paper, we present a three-step analytical procedure for examining a linear interaction effect between a dichotomous treatment variable and a continuous independent variable. The first step in this analytical procedure, which was discussed in detail by McNeil, Newman, and Kelly (1996, pp. 127-140), requires the researcher to design models that are capable of statistically testing the interaction effect. The technique used in the second step, which was previously presented by Fraas and Newman (1977), Newman and Fraas (1979) and Pedhazur (1982, pp. 468-469), requires the researcher or program evaluator to calculate the point of intersection between the two regression lines. The third step requires that the Johnson-Neyman confidence bands be calculated. This technique has been discussed by Johnson and Neyman (1936), Rogosa (1980, 1981), Chou and Huberty (1992), and Chou and Wang (1992).

In this paper, we are stressing the importance of using these techniques together in a three-step analytical procedure. The use of this analytical procedure will provide researchers and program evaluators with the type of information that will increase their understanding of the nature of the interaction effect being examined. To illustrate the type of information that is produced by this three-step analytical procedure, we have analyzed the personal and teaching efficacy levels of teachers who were exposed to an instructional model developed by Russell (1992), which is referred to as FOCUS.

Analytical Technique Applied to Efficacy Scores

Even though Russell (1992) believed that the exposure to the FOCUS model would increase the participants' levels of personal and teaching efficacy, he was not willing to assume that those increases would be constant across the participants' pre-term efficacy levels. That is, when comparing the post-treatment personal efficacy and teaching efficacy scores of the teachers who were exposed to the FOCUS model to teachers who were not exposed to the model, the differences may not be consistent across the ranges of the pre-term efficacy scores. Thus, to understand the possible influence of the FOCUS model on the personal efficacy and teaching efficacy scores of teachers, it was essential, not only to test for the existence of pre-term efficacy scores by group interaction effects, but also to gain insight into the nature of these interaction effects, if in fact, they did exist.

Subjects

Sixty-eight teachers who were enrolled in graduate level classes offered by the Education Department of Ashland University were included in the evaluation of the FOCUS model. Ashland University is located in north-central Ohio, which contains rural, suburban, and urban school systems. The courses, which required 36 hours of instruction, were offered during a summer term. Twenty-nine of the 68 teachers were not exposed to the FOCUS model. These 29 teachers, who taught in grade levels that ranged from kindergarten to the twelfth grade, served as the Control Group. The other 39 teachers were exposed to the FOCUS model during the same academic summer term. These 39 teachers, who also taught in grade levels that ranged form kindergarten through the twelfth grade, were designated as the treatment group. This treatment group was referred to as the FOCUS Group.

Instruments

Various instruments are used to measure the level of a teacher's sense of efficacy. In this evaluation project, the Teacher Efficacy Scale, which was devised by Gibson and Dembo (1984), was used. This selection was consistent with the view expressed by Ross (1994) who stated in his extensive review of the teacher-efficacy research that:

> Future researchers should treat the [teacher efficacy] construct as a multi-dimensional entity rather than a singular trait, examining personal and general teaching efficacy separately rather than aggregating them [and they] should measure teacher efficacy with the most frequently used instruments to facilitate comparisons between studies (p. 27).

Each educator who participated in this study completed the Teacher Efficacy Scale at the beginning and end of the summer academic term. This instrument required each participant to rate each of 16 statements on a 1 (strong disagree) to 6 (strongly agree) scale. The ratings obtained from the first nine statements were summed to obtain a personal efficacy score for each teacher. A high score on these nine statements was interpreted to mean that the teacher had a high level of personal efficacy. And a low score would indicate that the teacher had a low level of personal efficacy. The other seven statements were used to measure a teacher's teaching efficacy score. The total score on these seven statements for each teacher was subtracted from 42. This procedure produced a teaching efficacy score that would be high for a teacher who had a high level of teaching efficacy. The score would be low for a teacher who had a low level of teaching efficacy.

Gibson and Dembo (1984) reported in their study that an analysis of internal consistency reliability values produced Cronbach's alpha coefficient values of .78 and .75 for the personal efficacy scores and teaching efficacy scores, respectively. In addition, Gibson and Dembo stated that a multitrait-multimethod analysis supported both convergent and discriminant validity of the instrument.

Hypotheses

Two null hypotheses were statistically tested in the efficacy study. These null hypotheses were as follows:

$1H_0$: The interaction effect between the pre-term personal efficacy scores and group membership does not account for some of the variation in the post-term personal efficacy scores.

$2H_0$: The interaction effect between the pre-term teaching efficacy scores and group membership does not account for some of the variation in the post-term teaching efficacy scores.

Each of these null hypotheses were statistically tested through the three step procedure presented in the following sections.

Step 1: Statistical Tests of the Interaction Effects

Step 1 of the three-step analytic procedure was implemented for the efficacy data by statistically testing multiple linear regression models that were designed to measure the linear interaction effects. As part of this hypothesis testing procedure, the data utilized in each model were tested for possible outlier values with tests of Cook's distance measures (Neter, Wasserman, & Kutner, 1985). Any person who had a value that would distort the regression analysis was reviewed to determine whether the data for that person should be eliminated. The test results of Cook's distance measures indicated that the data recorded for one teacher may distort the results obtained from the regression analysis of the teaching efficacy scores. After reviewing that teacher's data, the data were deleted from the regression analyses. Thus, a total of 68 teachers and 67 teachers were included in the regression analyses of the personal efficacy scores and teaching efficacy scores, repectively.

The model that was designed to test 1Ho, which dealt with the teachers' personal efficacy scores, contained three independent variables. The teachers' post-term personal efficacy scores served as the dependent variable for this model. One of the independent variables included in this model consisted of the teachers' pre-term personal efficacy scores. This variable was labeled Pre-Term PE. The second independent variable included in this model was the Group variable. This Group variable consisted of the values of zero and one. A value of one indicated that the teacher was in the FOCUS Group, and a zero value meant that the teacher was in the Control Group. The third variable included in this model was formed by multiplying the Pre-Term PE variable by the Group variable. The inclusion of this variable, which was labeled (Pre-Term PE)*(Group), allowed us to use the regression model to calculate the difference between the slopes of the Control and FOCUS groups' regression lines.

The *t*-test value of the regression coefficient for the (Pre-Term PE)*(Group) variable was used to test $1H_0$. Since this study involved two dependent variables, i.e., the personal efficacy and teaching efficacy variables, the alpha level for the t test of this regression coefficient value was set at .025, which is equal to .05 divided by 2. The chance of committing a type I error was reduced by using this alpha value (Newman & Fry, 1972).

The results obtained from the analysis of the regression model are contained in Table 1. The t test of regression coefficient for the (Pre-Term PE)*(Group) variable (t = -2.44, p = .0175) indicated that the difference between the slopes of the regression lines of the FOCUS and Control groups was statistically significant at the .025 level, that is, $1H_0$ was rejected. Thus, the differences between the post-term personal efficacy scores of the FOCUS and Control groups were not constant across the range of pre-term personal efficacy scores.

Table 1

Regression Results for the Post-Term Personal Efficacy Scores

Regression Model

Variable	Regression Coefficient	t Value	p Value
(Pre-Term PE)*(Group)	-0.538	-2.44	0.018
Pre-Term PE	0.852	5.17	<.000
Group	25.124	2.87	0.006
Constant	6.362	0.97	0.338
R2 = .370			
Adjusted R2 = .341			
N = 68			
Residual Sum of Squares = 2495.58			

Note. The values for the Group variable are zero and one for teachers in the Control and FOCUS groups, respectively.

Table 2
Regression Results for the Post-Term Teaching Efficacy Scores

Regression Model

Variable	Regression Coefficient	t Test Value	p Value
(Pre-Term TE)*(Group)	0.703	2.742	0.008
Pre-Term TE	0.153	0.79	0.433
Group	-14.569	-2.339	0.023
Constant	19.8	4.331	<.000
R2 = .347			
Adjusted R2 = .316			
N = 67			
Residual Sum of Squares = 1334.318			

Note. The values for the Group variable are zero and one for teachers in the Control and FOCUS groups, respectively.

The teaching efficacy scores served as the dependent variable in the regression model that was used to test 2Ho. Similar to the previous regression model, this model included three independent variables. One of these independent variables was composed of the teachers' pre-term teaching efficacy scores. This variable was labeled Pre-Term TE. A second independent variable included in the model was the Group variable. The third independent variable included in the model was generated by multiplying the Pre-Term TE variable by the Group variable. This variable, which was labeled (Pre-Term TE)*(Group), was used to estimate the difference between the slopes of the regression lines for the Control and FOCUS groups.

The values generated by the analysis of the regression model used to test $2H_0$ are listed in Table 2. The t test of the regression coefficient for the (Pre-Term TE)*(Group) variable (t = 2.742, p = .008) indicated that the interaction effect was statistically significant at the .025 level. Thus, the differences between the post-treatment teaching efficacy scores of the FOCUS and Control groups were not constant across the range of pre-term teaching efficacy scores.

Step 2: Calculation of the Point of Intersection

The second step of the three-step analytical procedure was implemented by, first, graphing each of the interaction effects. If a given the interaction effect is disordinal, the point of intersection between the two regression lines would be calculated. If the interaction effect is ordinal, that is, the regression lines do not intersect in the relevant range, the researcher would proceed to Step 3.

The interaction effect between the Pre-Term PE variable and the Group variable is diagramed in Figure 1. Since the interaction effect was disordinal, the point at which the two regression lines intersected was calculated as follows:

1. The value of zero was substituted for the Group variable in the regression equation contained in Table 1 to obtain the regression line for the Control Group.

Y = 6.362 - .538*(Pre-Term PE)*(Group) + .852*(Pre-Term PE) + 25.124*(Group)

Y = 6.362 - .538*(Pre-Term PE)*(0) + .852*(Pre-Term PE) + 25.124*(0)

Y = 6.362 + .852*(Pre-Term PE)

2. The value of one was substituted for the Group variable in the regression equation contained in Table 1 to obtain the regression line for the FOCUS Group.

Y = 6.362 - .538*(Pre-Term PE)*(Group) + .852*(Pre-Term PE) + 25.124*(Group)

Y = 6.362 - .538*(Pre-Term PE)*(1) + .852*(Pre-Term PE) + 25.124*(1)

Y = 31.486 + .314*(Pre-Term PE)

3. The two regression lines were set equal to each other and the researcher solved the equation for Pre-Term PE.

6.362 + .852*(Pre-Term PE) = 31.486 + .314*(Pre-Term PE)

.538*(Pre-Term PE) = 25.124

Pre-Term PE = 46.7

As indicated by the results of this calculation and the graph of the disordinal interaction effect contained in Figure 1, the post-term personal efficacy scores of the teachers in the FOCUS Group were higher than the post-term personal efficacy scores of the teachers in the Control Group when their pre-term personal efficacy scores were less than 47. The post-term personal efficacy scores of the teachers in the Control Group, however, were higher than the post-term personal efficacy scores of the teachers in the FOCUS Group when their pre-term personal efficacy scores were greater than or equal to 47.

The interaction effect between the Pre-Term TE variable and the Group variable, which is diagramed in Figure 2, was also disordinal. Using the values produced by the regression analysis contained in Table 2, the point at which the two regression lines for the post-term teaching efficacy scores intersected was calculated in the same manner as was the intersection point for the personal efficacy scores. The calculations were as follows:

1. The value of zero was substituted for the Group variable in the regression equation contained in Table 2 to obtain the regression line for the Control Group.

Y = 19.800 + .703*(Pre-Term TE)*(Group) + .153*(Pre-Term TE) - 14.569*(Group)

Y = 19.800 + .703*(Pre-Term TE)*(0) + .153*(Pre-Term TE) - 14.569*(0)

Y = 19.800 + .153*(Pre-Term TE)

2. The value of one was substituted for the Group variable in the regression equation contained in Table 2 to obtain the regression line for the FOCUS Group.

Y = 19.800 + .703*(Pre-Term TE)*(Group) + .153*(Pre-Term TE) - 14.569*(Group)

Y = 19.800 + .703*(Pre-Term TE)*(1) + .153*(Pre-Term TE) - 14.569*(1)

Y = 5.231 + .856*(Pre-Term TE)

3. The two regression lines were set equal to each other and the researcher solved the equation for Pre-Term TE.

19.800 + .153*(Pre-Term TE) = 5.231 + .856*(Pre-Term TE)

.703*(Pre-Term TE) = 14.569

Pre-Term TE = 20.7

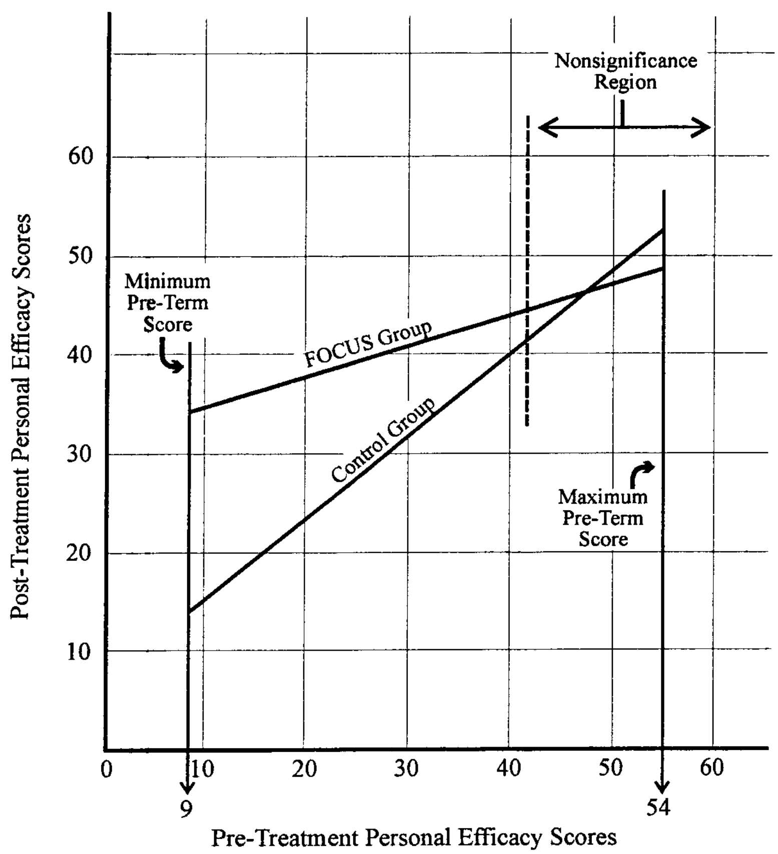

Figure 1. Pre-Term Personal Efficacy Scores by Group Interaction.

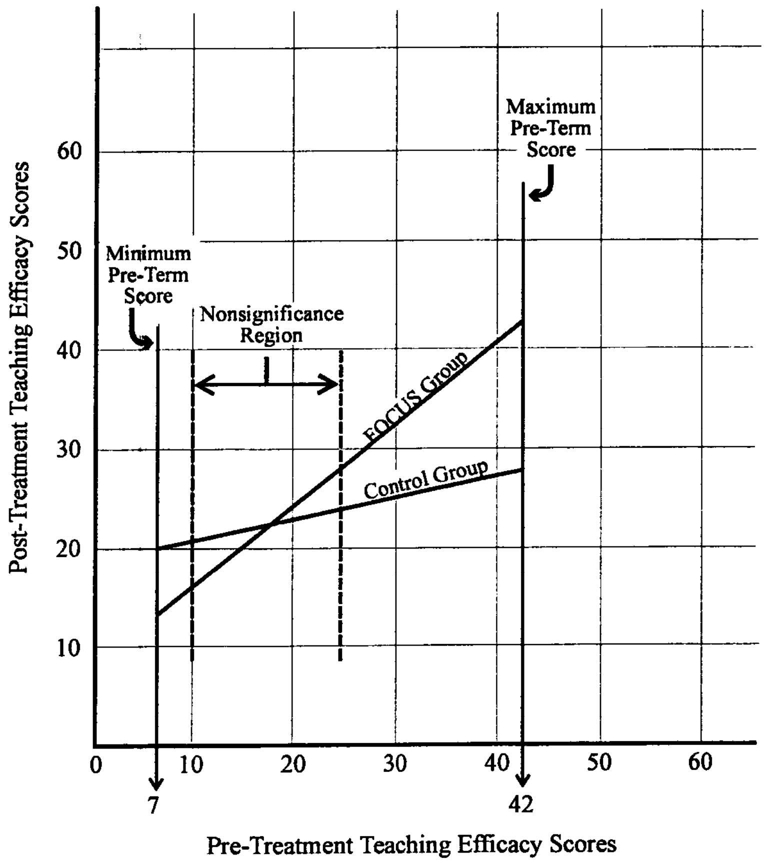

Figure 2. Preterm Teaching Efficacy Scores by Group Interaction

Table 3
Percentage of Teachers with Pre-Term Efficacy Scores Located In Various Regions Above and Below the Points of Intersection Between the Two Pairs of Regression Lines

Post-Term Personal Efficacy Scores	Post-Term Teaching Efficacy Scores FOCUS > Control	FOCUS < Control
FOCUS > Control	72%	19%
FOCUS < Control	6%	3%

The post-term teaching efficacy scores of the teachers in the Control Group were greater than the post-term teaching efficacy scores of the teachers in the FOCUS Group when their pre-term teaching efficacy scores were below 21. In addition, the post-term teaching efficacy scores of the teachers in the FOCUS Group were greater than the post-term teaching efficacy scores of the teachers in the Control Group when their pre-term teaching efficacy scores were greater than or equal to 21.

After the intersection point is calculated in a study that investigates an interaction effect between a continuous independent variable and a treatment variable, it is important to note the percentage of the study's participants who have scores above and below the intersection point. For the efficacy data of the 67 teachers who were included in both analyses, the percentages are listed in Table 3. As indicated in Table 3, 72% of the teachers had pre-term efficacy scores that corresponded to points on the regression lines where the teachers had higher post-term personal efficacy scores and higher post-term teaching efficacy scores when exposed to the FOCUS model. Only 3% of the teachers had pre-term efficacy scores that corresponded to points on the regression lines where the teachers had lower post-term personal efficacy scores and lower post-term teaching efficacy scores when exposed to the FOCUS model. Nineteen percent of the teachers had pre-term efficacy scores that corresponded to points on the regression lines where the teachers had higher post-term personal efficacy scores and lower post-term teaching efficacy scores when exposed to the FOCUS model. And 6% of the teachers had pre-term efficacy scores that corresponded to points on the regression lines where the teachers had lower post-term personal efficacy scores and higher post-term teaching efficacy scores when exposed to the FOCUS model.

With respect to these percentages, It is important to realize that the differences between the post-term efficacy scores of the FOCUS and Control groups may be statistically significant only for certain ranges of the pre-term efficacy scores. Thus, before conclusions are drawn with respect to who benefits and who does not benefit from being exposed to the FOCUS model, it is essential to determine the ranges of pre-term efficacy scores in which the differences between the post-term efficacy of the teachers in the FOCUS Group and the teachers in the Control Group are statistically significant. Step 3 of this three-step analytical procedure is designed to determine these statistically significant ranges.

Step 3: Calculation of the Johnson-Neyman Confidence Bands

The third step of the three-step analytical procedure requires that the Johnson-Neyman confidence limits be calculated for each statistically significant interaction effect. It should be noted that some researchers have argued that the Johnson-Neyman regions of significance are non simultaneous ones (Potthoff, 1964 and Rogosa, 1980, 1981). Based on empirical results by Chou and Huberty (1992) and Chou and Wang (1992), it appears that the Johnson-Neyman technique can be used to make simultaneous inferences provided that the slope homogeneity assumption is statistically tested and rejected. Since $1H_0$ and $2H_0$ were rejected, it was appropriate to calculate Johnson-Neyman (1936) confidence bands for the nonsignificance regions for the efficacy scores.

The program that was used to calculate the Johnson-Neyman confidence bands, which can be used in conjunction with the SPSS/PC+ Statistics™ software (SPSS Inc., 1990) and the SPSS® Base 7.0 for Windows™ (SPSS Inc., 1996), is listed in the Appendix. The program, which calculates the Johnson-Neyman significance bands as suggested by Pedhazur (1982, pp. 169-171), requires that 12 values be provided. A description of the required values, as well as their labels, are as follows:

1. The symbol <u>ss1</u> represents the pre-term sum of squares value for the Control Group.
2. The symbol <u>ss2</u> represents the pre-term sum of squares value for the FOCUS Group.
3. The symbol <u>n1</u> represents the sample size of the Control Group.
4. The symbol <u>n2</u> represents the sample size of the FOCUS Group.
5. The symbol <u>sumresid</u> represents the residual sum of squares value of the regression model.
6. The symbol <u>mean1</u> represents the mean of the pre-term scores of the Control Group.
7. The symbol <u>mean2</u> represents the mean of the pre-term scores of the FOCUS Group.
8. The symbol <u>slope1</u> represents the slope of the regression line for the Control Group.
9. The symbol <u>slope2</u> represents the slope of the regression line for the FOCUS Group.
10. The symbol <u>int1</u> represents the intercept point of the regression line for the Control Group.
11. The symbol <u>int2</u> represents the intercept point of the regression line for the FOCUS Group.
12. The symbol *fcrit* represents the critical F value with 1 and N - 4 degrees of freedom.

The sum of squares values, the sample sizes, and the mean values can be obtained from the printout generated by the DESCRIPTIVE subprogram of the SPSS/PC+ STATISTICS™ software (SPSS Inc., 1990) or the SUMMARIZE subprogram of the SPSS® Base 7.0 for Windows™ software (SPSS Inc., 1996), with each of the two groups being analyzed separately. The residual sum of squares value, the slope values, and the intercept-point values can be obtained from the printouts generated by the REGRESSION subprogram of either the SPSS/PC+ STATISTICS™ software or the SPSS® Base 7.0 for Windows™ software. The critical F value can be obtained from an F-Distribution Table.

The data line of the program listed in the Appendix, which utilized the freefield format, contains the data used to generate the Johnson-Neyman confidence limits for the personal efficacy scores. The data line used for the analysis of the teaching efficacy scores was as follows: 567.30 745.82 29 38 1334.32 23.24 24.71 .15 .86 19.80 5.23 4.00. Note that the numerator degrees of freedom (df_n) and the denominator degrees of freedom (df_d) values were 1 and 64 (68-4), respectively, for the analysis of the post-term personal efficacy scores. For the analysis of the post-term teaching efficacy scores, the values for df_n and the df_d were 1 and 63 (67-4), respectively. In addition, the confidence level was set at .95 for each set of limits.

The upper limit for the 95% confidence bands for the personal efficacy scores was 81.8, which was above the maximum score of 54 points on the personal efficacy section of the Teacher Efficacy Scale. The lower limit was 40.7. Based on these limits, which are included in Figure 1, it can be concluded that the post-term personal efficacy scores for the teachers in the FOCUS and Control groups were not statistically significantly different when their scores were greater than or equal to 41. The post-term personal efficacy scores of the teachers in the Focus Group were statistically significantly higher than the corresponding scores of the teachers in the Control Group, however, when their pre-term scores were less than 41.

The lower limit of the 95% Johnson-Neyman confidence limits for the regression lines diagramed in Figure 2 was equal to 9.97, which was less than three points above the minimum score of 7 that a teacher could receive on the teaching efficacy section of the Teacher Efficacy Scale. It should be noted, however, that none of the teachers included in this analysis had a pre-term teaching efficacy score below 13. Thus, none of the teachers included in this study had a score below the lower limit of the nonsignificance region. The upper limit of the nonsignificance region of the Johnson-Neyman 95% confidence limits for the pre-term teaching efficacy scores was 23.8. Thus, the post-term teaching efficacy scores of the teachers in the FOCUS and Control groups were not statistically significantly different when their pre-term teaching efficacy scores were less than 24. The post-term teaching efficacy scores of the teachers in the FOCUS Group, however, were statistically significantly higher than the post-term teaching efficacy scores of the teachers in the Control Group when their pre-term teaching efficacy scores were equal to or greater than 24.

To understand the implications of the nonsignificant regions as well as the significant regions for the two sets of regression lines, it is important to note the location of the teachers' pre-term efficacy scores along the two sets of regression lines. As indicated by the percentages contained in Table 4, 31% of the teachers who were included in both regression analyses had pre-term efficacy scores that corresponded to points on the regression lines where the post-term efficacy scores of the teachers in the FOCUS Group were statistically significantly higher than the scores of the teachers in the Control Group on both efficacy scales. In addition, 42% of the teachers had pre-term efficacy scores that corresponded to points on the regression lines where the post-term efficacy scores of the teachers in the FOCUS Group were statistically significantly higher than the scores of the teachers in the Control Group on one of the two efficacy scales. The remaining 27% of the teachers had pre-term efficacy scores that corresponded to points on the regression lines where the post-term efficacy scores of the two groups were not statistically significantly different on either efficacy scale.

Table 4
Percentage of Teachers with Pre-Term Efficacy Scores Located in the Various Significant and Nonsignificant Regions

Post-Term Personal Efficacy Scores	Post-Term Teaching Efficacy Scores: FOCUS > Control	FOCUS = Control	FOCUS < Control
FOCUS > Control	31%	21%	0%
FOCUS = Control	21%	27%	0%
FOCUS < Control	0%	0%	0%

Thus, a total of 73% had pre-term efficacy scores that were located at points on the regression lines where the post-term efficacy scores of the teachers in the FOCUS Group were statistically significantly higher than the post-term efficacy scores of the teachers in the Control Group on at least one of the two efficacy scales. None of the teachers (0%) had pre-term efficacy scores that were located at points on the regression lines where the post-term efficacy scores of the teachers in the Control Group were statistically significantly higher than the post-term efficacy scores of the teachers in the FOCUS Group on either of the two efficacy scales.

Implications Based on the Results of the Three-Step Analytical Procedure.

It is important to understand what each step in this three-step analytical procedure reveals about the linear interaction effects. The results of Step1 indicate that both interaction effects were statistically significant. A more in-depth understanding of these interaction effects, however, is obtained by reviewing the information generated by Steps 2 and 3 of this three-step analytical procedure.

The graphs containing the interaction effects and the points of intersection between the regression lines for the personal efficacy scores and the teaching efficacy scores, which were completed in Step 2, revealed that both interaction effects were disordinal and the regression lines for the personal efficacy scores and the teaching efficacy scores intersected at 46.7 and 20.7, respectively. These graphs and the intersection points appear to suggest that, with respect to their post-term efficacy scores, certain teachers would benefit from being exposed to the FOCUS model, while exposure to the FOCUS model would be detrimental to other teachers. In addition, these points of intersection could possibly be used to identify which teachers would and would not benefit from exposure to the FOCUS model. Before such a conclusion is reached, however, it is important to realize that the differences between the post-term efficacy scores of the teachers in the FOCUS and Control groups, who have pre-term scores near the intersection points, could simply be due to noise or random variation. That is, the post-term scores of the students in the two groups are statistically significantly different only for pre-term scores that are located some distance above and below the intersection points. Thus, before one should draw a conclusion with respect to the nature of these interaction effects, it is essential to review the information provided by the Johnson-Neyman confidence limits calculated in Step 3.

The significance region between the two regression lines that were designed to analyze the post-term personal efficacy scores included only the pre-term personal efficacy scores that were less than 41. In addition, the significance region between the two regression lines that were designed to analyze the post-term teaching efficacy scores included only the pre-term teaching efficacy scores that were greater than or equal to 24. Thus, as indicated by the interaction effects contained in Figures 1 and 2, whenever the post-term efficacy scores of the two groups were statistically sigficantly differerent, the post-term efficacy scores of the Focus Group exceeded the post-term efficacy scores of the Control Group.

Thus, a majority of teachers (73%) had pre-term efficacy scores that placed them in ranges along the regression lines that indicated that the post-term efficacy scores of the teachers in the Focus Group, on at least one of the efficacy scales, were statistically significantly higher than the post-term efficacy scores of the teachers in the Control Group. It is important to also note that in spite of the fact that the interaction effects were disordinal, the reverse statement is not true. That is, none of the teachers had pre-term efficacy scores in the ranges along the regression lines that indicated that the post-term efficacy scores of the Focus Group were statistically significantly lower than the post-term efficacy scores of the Control Group on either of the two efficacy scales. The remaining 27% of the teachers had pre-term efficacy scores in the ranges along the regression lines that indicated that the post-term efficacy scores of the FOCUS and Control groups were not statistically significantly different on either of the two efficacy scales.

Based on this information, one would not use the intersection points between the regression lines to determine who would and who would not benefit from being exposed to the FOCUS model. Rather, it would be more appropriate, keeping in mind research design limitations, to suggest that, based on pre-term efficacy levels, exposing the teachers to the FOCUS model would be beneficial to the majority of teachers and it would not be detrimental to any one group of teachers. Educational researchers and program evaluators would reach this conclusion only by using this three-step analytical procedure.

Summary

It is important for educational researchers and program evaluators to increase their understanding of the interaction effects that may be present in their data. We believe that a more in-depth understanding of a linear interaction effect between a continuous independent variable and a dichotomous treatment variable can be obtained if the educational researcher or program evaluator follows the three-step analytical procedure that was presented in this paper.

Two points should be noted regarding this three-step analytical procedure. First, the use of a multiple regression model to statistically test the interaction effect, which is undertaken in Step 1, is an essential analytical procedure to consider when investigating the difference between the scores of two groups. This test of the homogeneity of the slopes of the regression lines allows the researcher to not only to determining if the interaction effect is statistically significant, but it also permits simultaneous inferences to be made from the Johnson-Neyman confidence bands, which are calculated in the third step of this analytical procedure.

Second, the calculation of the intersection point between the two regression lines in Step 2 could posssibly provide a researcher or program evaluator with information that could be used to identify groups of people who would benefit from being exposed to the treatment being investigated. It is important to realize, however, that the difference between the post-term scores of the students in the two groups who have pre-term scores that are located near this intersection point could be simply due to noise or random variation. That is, the post-term scores of the students in the two groups are statistically significantly different only for pre-term scores that are located some distance above and below that intersection point. The calculation the Johnson-Neyman confidence limits in Step 3 allows the researcher or program evaluator to determine the pre-term scores at which the post-term scores of the two groups are statistically significantly different. This information may lead the researchers or program evaluators to modify conclusions that were based solely on information provided by the analytical techniques contained in the first two steps of this process.

As was demonstrated by the analyses of the personal efficacy and teaching efficacy scores that were presented in this paper, following the three-step analytical procedure can provide essential information not only regarding whether an interaction effect does, in fact, exist but also with respect to the nature of the interaction effect. Such information can be invaluable to educational researchers and program evaluators.

References

Chou, T., & Huberty, C. J. (April, 1992). *The robustness of the Johnson-Neyman Technique.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Chou, T., & Wang, L. (April, 1992). *Making simultaneous inferences using Johnson- Neyman technique.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Fraas, J. W., & Newman, I. (April, 1977). *Malpractice of the interpretation of statistical analysis.* Paper presented at the annual meeting of The Ohio Academy of Science, Columbus, OH.

Gibson, S., & Dembo, M.H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76,* 569-682.

Johnson, P.O., & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. In J. Neyman and E.S. Pearson (Eds.), *Statistical Research Memoirs,* 1936, 1, 57-93.

McNeil, K., Newman, I., & Kelly F. J. (1996). *Testing research hypotheses with the general linear model.* Carbondale, IL: Southern Illinois University Press.

Neter, J., Wasserman, W., & Kutner, M.H. (1985). *Applied linear statistical models* (2nd ed.). Homewood, IL: Irwin.

Newman, I., Deitchman, R., Burkholder, J., Sanders, R., & Ervin, L. (1976). Type VI error: Inconsistency between the statistical procedure and the research question. *Multiple Linear Regression Viewpoints,* 6 (4), 1-19.

Newman, I., & Fraas, J. W. (1979). Some applied research concerns using multiple linear regression [Monograph]. *Multiple Linear Regression Viewpoints,* 9, (4).

Newman, I., & Fry, W. (1972). A response to 'A note on multiple comparisons and a comment on shrinkage'. *Multiple Linear Regression Viewpoints, 2,* (1), 36-39.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research: Explanation and prediction.* (2nd ed.). Fort Worth, TX: Harcourt Brace Jovanovich.

Rogosa, D. (1980). Comparing nonparallel regression lines. *Psychological Bulletin, 88,* 307-321.

Rogosa, D. (1981). On the relationship between the Johnson-Neyman region of significance and statistical tests of parallel within-group regressions. *Educational and Psychological Measurement, 41*, 73-84.

Ross, J.A. (1994, June). *Beliefs that make a difference: The origins and impacts of teacher efficacy*. Paper presented at the meeting of the Canadian Association for Curriculum Studies, Calgary, Canada.

Russell, G. (1992). *FOCUS: An explanation of the human behavioral system.* Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, IL.

SPSS, Inc. (1990). SPSS/PC+ Statistics™ (Version 4.0) [Computer software]. Chicago: SPSS Inc.

SPSS, Inc. (1996). SPSS® Base 7.0 for Windows™ [Computer software]. Chicago: SPSS Inc.

Appendix

Computer Program for the Calculation of the Johnson-Neyman Confidence Limits.

```
Data list free/
  ss1  ss2  n1  n2  sumresid  mean1  mean2  slope1  slope2  int1
int2  fcrit
Begin data.
1434.21  1821.59  29  39  2495.58  39.31  38.50  .85  .31  6.36
31.49  3.99
End data.
Compute term1 = (fcrit/(n1+n2-4))*sumresid.
Compute terma = term1*(-1).
Compute a = ((terma)*((1/ss1)+(1/ss2)))+(slope1-slope2)**2.
Compute  b = (term1*((mean1/ss1)+(mean2/ss2)))+((int1-
int2)*(slope1-slope2)).
Compute c = (terma)*(((n1+n2)/(n1*n2))+((mean1**2)/ss1)+
                     ((mean2**2)/ss2))+((int1-int2)**2).
Compute RegionU = ((b*(-1))+(sqrt((b**2)-(a*c))))/a.
Compute RegionL =  ((b*(-1))-(sqrt((b**2)-(a*c))))/a.
List RegionU RegionL.
```