'LE LINEAR REGRESSION VIEWPOINTS
'E 14, NUMBER 1, SPRING 1986

Significant Interaction: I Got What I Needed

Keith A. McNeil and Gail Smith

Dallas Independent School District

Background

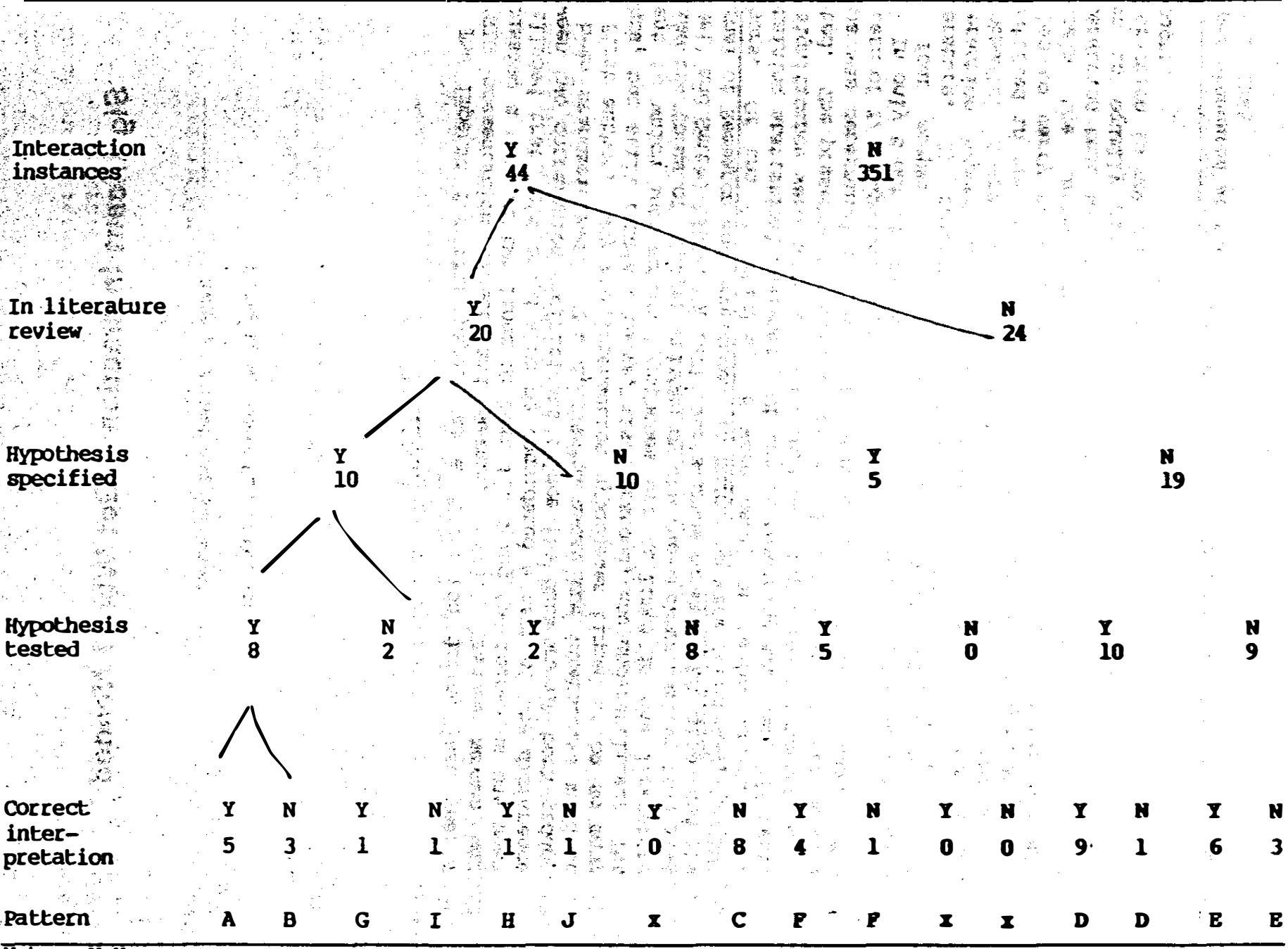
The impetus for this paper was a discussion during last year's /SIG presentation (Hoedt and Newman, 1984). Isadore Newman was cussing a test of two lines of best fit being considered as one when alluded that this could also be considered a test of the difference ween two correlation coefficients (since the data within both groups been standaridized.) The discussion awoke the interactive mind of first author. Why are interaction hypotheses hinted at on so many nts, but still remain elusive, misunderstood, and underutilized? The ent to which interaction hypotheses are utilized in the literature ame the focus of a paper written by the two authors earlier this year cNeil and Smith, 1985). A full year's issue of Urban Education and the rnal of Research and Development in Education were reviewed by the two chors. Of the 57 articles, 38 were essay or review articles not ntaining statistical analyses. Of the 19 remaining articles, 386 tests significance were computed, with only 44 interaction hypotheses being ted. The presence (Y) or absence (N) of each aspect of four crucial ps was determinwd for each of these 44 interaction instances. The ttern of Y/N responses is presented in Table 1.

In only 5 out of the 44 instances (Pattern A) did the author follow four steps: 1) identify the interaction hypothesis in the terature, 2) specify the interaction hypothesis, 3) test the teraction hypothesis, and 4) correctly interpret the interaction pothesis. There were 8 instances of Pattern C, wherein the author lentified in the review of literature juicy interaction hypotheses, but iled to carry through. Pattern D represents the computer society, erein the canned computer program automatically provides the teraction test so the author feels obligated to interpret the results. at is equally disturbing is the last two bins, Pattern E. Here teraction is not discussed until the interpretation stage -- food for hought.

---

aper presented at the American Educational Research Association, Chicago, pril, 1985

(From McNeil and Smith, 1985)



Note: Y=Yes
N=No

Review of Multiple Linear Regression Viewpoints for Applied Interaction Studies

Interaction hypotheses can easily be tested within the Multiple Linear Regression (MLR) approach, and there has been a history of MLR being taught alongside "complex behavioral science models incorporating interaction and non-linear variables (Kelly, Beggs, and McNeil, 1969; Fraser, 1979; Bottenberg and Ward, 1963). It was therefore predicted that a higher percentage of interaction hypotheses would appear in Multiple Linear Regression Viewpoints, (the journal of SIG/MLR) than in the two journals previously reviewed.

> When the predictors were used to predict the criterion "for the experimental and control groups separately, apparent differences were found in the two regression equations. It was these differences that led to the present consideration of the interaction of the predictors with experimental condition (Group) as a way of exploring the differences statistically." (Dinero, 1976)

So begins one of the few research studies which tests an interaction hypothesis in a meaningful way. All issues from 1975 through 1980 were reviewed. Only nine applied studies were found, with 49 of the 506 tests of significance involving interaction. Of the five studies which did consider an interaction hypothesis, two studies fit Pattern B (10 interaction instances), one Pattern F (28 interaction instances), and two Pattern D (11 interaction instances). In no case did the researcher include all four of what we consider to be crucial steps. Additionally, the percentage of interaction hypotheses is lower in Viewpoints than in the two applied education journals discussed earlier. This finding is particularly disconcerting because much has been written in Viewpoints about interaction and how easily one can test it within the MLR framework. The following (selected) review is intended to once again reinforce these interaction notions. Fraser (1979) provides a comprehensive approach to research with MLR. Researchers who haven't "interacted" within the last five years ought to reread the article.

Review of Multiple Linear Regression Viewpoints for Interaction Comments

Why so few researchers test interaction questions remains a puzzle. All canned ANOVA computer programs routinely provide a test for interaction. All stat texts discuss the concept, most in a negative light though. (The Kelly, Beggs, and McNeil (1969) text had the audacity to place curvilinear interaction on the text's cover.) Of most relevance to the members of SIG/MLR is the paucity of good applied interaction studies outside our journal. This is particularly disconcerting given the extensive discussion by numerous authors in Viewpoints. Upon rereading the early volumes of Viewpoints, we were astounded at the frequency and quality of interaction discussions. Desiring the work of these early "interactive pioneers" to not remain shelved, we will quote liberally.

Construction of interaction variables

An interaction variable is reflected in MLR as a product of two variables. If both variables are dichotomous then traditional ANOVA designs are being reflected. If one of the variables is dichotomous and the other continuous, then a difference between groups is being considered (evaluating "the question of homogeneous slopes [Jennings, 1972) or the difference between two correlation coefficients [Hoedt and Newman, 1984]). It has been shown that in the test for homogeneity of regression slopes, both methods of calculating analysis of covariance - traditional ANCOVA and MLR -- are exactly the same" (Newman and Fry 1972). (See also Jennings, 1972 and Williams, Naresh, and Peebles, 1972.)

If both of the variables are continuous then "continuous interaction" (McNeil and McNeil, 1975) or moderator variables are being investigated. Moderator variables "lend somewhat limited support for the use of more complex models. Moderators improve preciction by acknowledging possible interactive effects of the moderator variable with other variables in the regression anlaysis." (Reed, Feldhusen, and Van Modfrans, 1971)

If the variables are actually the same variable, then a higher order effect (curvilinearity) is being implemented. This extension of interaction into curvilinearity was first brought to the senior author's atttention by Jack Byrne during his Doctoral prelims. Dinero (1977) later makes the connection: "Now that one has decided to use interaction terms in his prediction model, he has to decide which ones to include. The predictors raised to the first power, these variables squared or cubed or any of their cross-products may be used."

Dinero (1977) also reiterates the ease and value of conceptualizing research within the MLR approach. "Once a researcher understands how to generate interactions, more avenues of investigation are open. The regression model brings with its flexibility a set of decisions many researchers in the past have either ignored or been unaware of."

Interpretation of interaction

Many researchers avoid interactions because of interpretation problems. Here is what Viewpoints authors have to say about the interpretation issue.

> "A significant interaction hampers the interpretation of main effects, but the positive view is that a significant F test of interaction tells us how to appropriately limit our generalization" (Spaner, 1977).
>
> "A final word of warning is that second and higher order interactions must be interpreted with great care, if meaningless or erroneous conclusions are not to be drawn from research data." (Brebner, 1972)
>
> "In general, significant three-way interaction is seen to reflect different two-way interactions: if the ABC interaction is significantly different from zero, then either AB varies across C, AC varies across B, or BC varies across A. In any case, these differences would be manifest by significant cross-products of the standardized predictors." (Dinero, 1977)

"Indeed the value of need for interaction tests has been grossly underemphasized in MLR studies. I suspect that this phenomenon arises out of a misunderstanding, perhaps even fear, of a significant interaction finding." (Spaner, 1977)

"Indeed the value of need for interaction tests has been grossly underemphasized in MLR studies. I suspect that this phenomenon arises out of a misunderstanding, perhaps even fear, of a significant interaction finding." (Spaner, 1977)

McNeil and Beggs (1971) accepted the reality of interaction and challenged researchers to think about directional interactions -- thus fully utilizing the power of their statistical test. No directional hypotheses have appeared in our review of Viewpoints.

Nonlinear predictors

"Since many of the simplest functional relationships in the physical sciences have been found to be non-linear or interactive, we find it interesting that few non-linear relationships have been established in the behavioral sciences, especially since most behavioral scientists would maintain that human behavior is no less complicated than physical behavior." (McNeil, Evans, and McNeil, 1979)

There are "two reasons for including non-linear terms - either the expected functional relationship is non-linear, or the way the construct has been originally measured needs to be modified." (McNeil, 1976)

"A more important situation occurs when there is theoretical or empirical justification for the inclusion of such a variable." (McNeil and Spaner, 1971)

Interpretation problems with non-linear terms have been addressed.

"When quadratic and interaction terms are significant, however, interpretation is made more difficult. Still, an attempt at interpretation seems somewhat better than ignoring the problem or assuming it does not exist." (Reed, Feldhusen, and Van Modfrans, 1971)

"The range of manipulations available in order to test forms of curvilinearity is endless. However, contrived departure from linearity in regression models will not make trivial predictors into important ones." (Jordan, 1971)

Nonlinear criterion

There are two instances that come to mind when a nonlinear criterion would be used. One instance is when the functional relationship is indeed nonlinear (McNeil, Evans, and McNeil, 1979). The Pythagorean Theorem is one such example. Any criterion that is a ratio of one variable to another is another example. A second instance when a nonlinear criterion would be used is when the measure of the construct does not map the construct, and some rescaling of the measure is necessary (McNeil, et al 1979).

Potential problems

When continuous variables are multiplied to reflect the interaction term several potential problems must be avoided. One potential problem is that the product is dependent on the means and variances of the original scores. Thus, researchers might want to standardize the variables before obtaining the product (Dinero, 1977). McNeil and McNeil (1975) also discussed the scaling effect on the resultant R^2. The product of two continuous predictor variables may not accurately reflect the interaction. The precictor variables must be rescaled such that the product term does match the expectations of the criterion.

Miscellaneous techniques

The search for interaction in the hypothesis generating mode has been well stated by Dinero (1977).

> "Given the problem of shrinkage, any regression anlaysis should be run in two phases, the first to estimate and the second to corroborate. This being the case, it may be just as wise to explore with the data of the first phase, to the extent of plotting the scatter diagrams, and use this information to select the interaction term to be used in the second phase. This type of exploration would seem to be almost a necessity in educational and phychological studies where there is little such comparative data available, where interaction has been something more to be avoided than awaited, and where complex aptitude-treatment interactions could bring exciting new interpretations to old data."

A computer program has been written to assist in finding the interactions which account for the most variance.

> "The primary value of AID-4 to the task scientist is its ability to identify the maximum amount of variance in the criterion which can be accounted for by the predictors available; it relieves the task scientist of the trial-and-error task of attempting to identify the various relevant combinations of linear and non-linear interaction terms presently required by the multiple linear regression technique. The splitting process of AID-4, being based upon maximizing the between sums-of-squares and minimizing the within sums-of-squares, automatically takes all present interaction into account, indiciting the maximum variance predictable in the cirterion from the predictors." (Koplyay, 1972)

Finally, the dectection of interaction is one of the major advantages of the "regression model" in evaluating compensatory education programs (McNeil and Findlay, 1980).

Discussion

The purpose for providing all the quotes in the previous sections was to document the interaction efforts made by authors in Viewpoints. The fact that the majority of these references are over 10 years old reflects more our concern for being aware of, and implementing existing methodology, rather than our lack of concern for improving existing methodology.

Given that this methodology exists for studying interaction questions, why don't more researchers look at interaction? We don't have the answer, but we have some thoughts, and we will present them grouped by the four major hypothesis testing steps.

With respect to literature review, most authors do not review interaction results, and when they do, they review them poorly. Furthermore, part of the publish or perish mentality is to invent new predictor variables, rather than try to increase the amount of variance accounted for. Finally, most researchers do not understand that different results from two studies implies an underlying interaction variable

In this world of posthoc orthogonal contrast coding and alpha protection levels few researchers realize that an interaction hypothesis can be specified all by itself, if no other question is of interest. But most of the statistics texts insist on a step-by-step procedure, looking at interaction in particular ways. What ever happened to the notion of the research question guiding the statistical tool?

With respect to the actual testing of the hypothesis, we have three major concerns. First, canned ANOVA programs generally don't allow for testing specific interaction questions. Second, canned MLR programs encourage the inclusion of linear terms first. (Stepwise linear programs, though of value for some purposes, totally ignore the testing of a specific hypothesis.) Third, most statistics texts still present the interaction question as being valuable only for meeting assumptions -- to reject so that main effects can be tested.

The fourth step in hypothesis testing, interpretation, also causes some problems for those considering interaction questions. Unfortunately most of our quoted Viewpoints authors acknowledge that interpreting an interaction result can be difficult. But if interaction is significant, then that is reflecting reality -- and shouldn't it be more valuable to make a "difficult" interpretation of reality as it is, than to make some "easier" statement about some constrained aspect of reality. Perhaps researchers need to become more familiar with significant interaction.

Summary

Fortunately, for us, the summary of our paper was published in Viewpoints over 12 years ago.

> "Perhaps one of the most overused assumptions within multivariate studies in educational research is that only simple linear relationships exist among the variables. Although interactive effects have been acknowledged within analysis of variance studies, the logical extension to regression analysis has rarely been actualized (Reed, Feldhusen, and Van Modfrans, 1971).
>
> "Too often, even plausible interactions are ignored and all subjects are lumped together and, hence, treated as similar. Our conceptual theories have long ago turned to distinct groupings, and it is about time that our statistical procedures reflect this empirical possibility." (Newman, Lewis, and McNeil, 1973).

Unfortuanately these comments seem to still be appropriate today. Hopefully tomorrow they will not be appropriate.

Epilogue

An examination of why interaction studies are not conducted in one specific area may shed some light on possible solutions. The two authors have been involved with educational program evaluations for several years. As such, we function as the program evaluator, providing evaluation information to the program manager.

In order to study an interaction question, the evaluator first needs to understand interaction concepts and be able to calculate interaction effects. Second, the evaluator must be able to translate these concepts into terms that the program manager can understand. Third, the interaction question must become of interest to the program manager, a person who often wants to use only the simplest of statements.

Collection of interaction information

Program managers usually want all students to be provided the best possible educatinal opportunity. This notion is usually envisioned in the same treatment for all. Denying treatments or parts of treatments is often not desired, and obtaining additional information from students is sometimes difficult if not impossible.

Verbal outcome

The program manager has a vested outcome in the program. Often the program has been devised by the manager and therefore the manager "knows" that the best program has been devised. Providing the same program to all students probably costs less, is easier administratively, and is usually more defensible to outside interests. The program manager is hard put to take the neutral stance towards the program that evaluators easily take.

Implications if interaction is significant

First, the program evaluator must clearly communicate to the program manager the implications of a significant interaction. Then the program manager must incorporate this finding into next year's program, a task which requires additional administrative attention.

When programs are constructed around significant interactions much additional administrative work is required. Program descriptions and guidelines must clearly reflect such interactions. Alternative programs must be delineated and procedures must be identified to get the right students (and probably the right teachers) into those programs. Different teaching materials may be required for the various programs, as well as different staff development. Classroom monitoring and program evaluations will continually need to incorporate those interactive variables. Consequently, additional administrative effort and commitment is required. Significant interactions imply that the KISS (Keep It Simple Stupid) principle is no longer applicable.

Roadblocks to replacing significant interactions

Everyone, including program managers, knows that results need to be replicated. The extent to which replicated results can be generalized to different settings and different students is usually an interesting question. But in the educational arena programs are often changed due to factors unrelated to evaluation results: a) new local, state, or Federal mandates, b) change in program manager, c) availability of personnel to plan and implement the program, and d) availability of funds.

Some possible next steps for SIG/MLR members

Now that we've a) established that adequate methodology exists to investigate interactive questions, b) documented that few interactive questions are being investigated, and c) specified some of the roadblocks to studying interactions in our field, we would like to propose some remediation.

First, we should all strive in our own daily endeavors to consider interaction hypotheses. We understand the methodology and can provide exemplary behavior to other researchers.

Second, we could infuse other SIGs and the various AERA Divisions. We challenge each of you to become involved in another SIG, to spread the interaction hypothesis.

Third, many of you participate in other national or regional educational meetings where more program managers are in attendance. These program people need to know that interaction questions can be tested — for behind every good program manager is an interaction hypothesis.

References

Bottenberg, R. A. and Ward, J. H. Applied Multiple Linear Regression. Lackland Air Force Base, Texas: Aerospace Medical Division, AD 413128, 1963.

Brebner, M. A. Conditions for no second-order interaction in multiple linear regression models for three factor anlysis of variance. Viewpoints, 1972, 3(1), 46-57.

Dinero, T. E. An empirical example of the use of interaction terms in the multiple regression model. Multiple Linear Regression Viewpoints, 1977, 7(2), 75-100.

Fraser, B. J. A multiple regression model for research on teacher effects. Multiple Linear Regression Viewpoints, 1979, 9(3), 37-52.

Hoedt, K. and Newman, I., Testing the hypothesis of a difference between P_1, and P_2 using independent and dependent samples. Paper presented at the meeting of the American Educational Research Association, New Orleans, March, 1984.

Jennings, E. Linear models underlying the anlysis of covariance, residual gain scores and raw gain scores. Viewpoints, 1972, 3(1), 17-24.

Jordan, T. E. Curvilinearity within early developmental variables. Viewpoints, 1971, 1(1), 53-77.

Kelly, F. J., Beggs, D. L., McNeil, K. A., Eichelberger, T. and Lyon, J. Research Design in the Behavioral Sciences: Multiple Regression Approach. Carbondale: Southern Illinois University Press, 1969.

Koplyay, J. B. Automatic interaction detector AID-4. Viewpoints, 1972, 3(1), 25-38.

McNeil, K. Position statement on the roles and relationships between stepwise regression and hypothesis testing regression. Multiple Linear Regression Viewpoints, 1976, 6(4), 46-49.

McNeil, K. A, & Beggs, D. L. Directional hypotheses with the multiple linear regression approach. Viewpoints, 1971, 1(1), 89-102.

McNeil, K., Evans, J., & McNeil, J. Nonlinear transformation of the criterion. Multiple Linear Regression Viewpoints, 1979, 9(5), 1-9.

McNeil, K., & Findlay, E. Evaluating Title I early childhood programs: Problems, the applicability of Model C, and several evaluation plans. Multiple Linear Regression Viewpoints, 1980, 10(4), 41-50.

McNeil, K., & McNeil, J. Some thoughts on continuous interaction. Viewpoints, 1975, 5(3), 41-46.

McNeil, K. A. and Smith, G. Educationally significant interaction. Paper presented at the meeting of the Southwest Educational Research Association, Austin, Texas, January, 1985.

McNeil, K. A. and Spaner, S. D. Brief report: Highly correlated predictor variables in multiple regression models. Multivariate Behavioral Research, 1971, 6, 117-125.

Newman, I., Lewis, E. L., & McNeil, K. A. Multiple linear regression models which more closely reflect Bayesian concerns. Viewpoints, 1972, 3(1), 71-77.

Newman, I., & Fry, J. Proof that the degrees of freedom for the traditional method of calculating analysis of covariance and the multiple regression method are exactly the same. Viewpoints, 1972, 3(1), 42-45.

Reed, C. L., Feldhusen, J. F., & Van Mondfrans, A. P. Regression models in educational research. Viewpoints, 1971, 1(1), 78-88.

Spaner, S. D. What inferences are allowable with a significant F in regression analysis? Multiple Linear Regression Viewpoints, 1977, 7(2), 62-74.

Williams, J. D., Maresh, R. T., & Peebles, J. D. A comparison of raw gain scores, residual gain scores, and analysis of covariance with two modes of teaching reading. Viewpoints, 1972, 3(1), 2-16.