Multiple Linear Regression Viewpoints

Volume 1, Number 3

January, 1971

A Publication of the Special Interest Group on Multiple Linear Regression of the American Educational Research Association

Editor: John D. Williams, The University of North Dakota

Chairman of the SIG: Samuel R. Houston, The University of Northern Colorado

Secretary of the SIG: Carolyn Ritter, The University of Northern Colorado

Table of Contents

SPECIAL PRE-CONVENTION ISSUE

The present special issue of Multiple Linear Regression Viewpoints includes the three papers to be read at the AERA Convention in New York in February. It should be a worthwhile practice to have available and to have read the papers of the SIG before the convention. This should allow the actual paper reading session to be more informal and allow a two-way exchange of information and viewpoint, rather than the traditional one-way presentation. Also included in this issue is an article co-authored by Sam Houston, SIG President.

Members of the SIG are encouraged to submit articles or notes for publication in Viewpoints. Send your articles exactly as you wish them to appear in Viewpoints. The publication charge continues to be $1.00 a page.

JDW

CONCURRENT VALIDITY OF THE KOPPITZ SCORING SYSTEM FOR THE BENDER VISUAL MOTOR GESTALT TEST

Anne F. Goff
and
Samuel R. Houston

This study was designed to examine the Koppitz Scoring System for the Bender Visual Motor Gestalt Test (BG) and its concurrent validity utilizing a clinical sample of school age children. The study investigated correlations between assessed visual motor perception, intelligence, and academic achievement. Secondly, the study examined the efficacy of prediction in criterion by employing two systems of analysis: (a) combining two variables /Bender error score and age/ in the prediction of the criterion, intelligence or achievement, and (b) correlating *z* scores /obtained from the Bender performance/ with either intelligence or achievement. The two analysis systems were, in turn, contrasted in regard to predictive efficiency.

Method

Subjects. A clinical sample of 50 primary school children, ranging either in mental age of 5-0 to 10-5 or in classroom placement from kindergarten through fourth grade, were randomly selected for the investigation. The sample was drawn from among approximately 650 school age children residing in Williamson County, Illinois who had previously been examined in a psychoeducation clinic. All children, comprising the sample, were referred to the clinic by the respective classroom teacher because of apparent emotional disturbances, learning disabilities, and/or cultural deprivation. Standard or derived scores were obtained for intelligence and achievement and were correlated with the Bender raw score, as well as with the Bender *z* score. The sample's mean age in months was 99.4; the mean IQ was 96.9, while reading and arithmetic means were 84.0 and 87.7, respectively.

Procedure. The BG was employed to assess visual-motor perception; depending upon age and other factors pertinent to the individual case study, intelligence quotients were obtained from the Stanford-Binet (SB) or the Wechsler Intelligence Test for Children (WISC). Achievement in reading and arithmetic was measured by the Wide Range Achievement Test (WRAT). All assessments were secured during the same evaluation period for each of the 50 children and all instruments were administered by a trained school psychologist. In addition to the simple correlational analysis, the investigators used multiple linear regression, (Ward, 1962) to determine unique contribution of sets of predictor variables on a given criterion.

Results and Discussion

Intercorrelation coefficients (Table 1) are Pearson product-moment coefficients. Since the Bender performance is scored for errors, the expected correlations with this variable would be negative.

Table 1
Intercorrelations

	1	2	3	4	5
1 CA (Months)					
2 I.Q.	-.46*				
3 Read. Ach.	-.35*	.44*			
4 Arith. Ach.	-.05	.48*	.68		
5 Bender Error	-.65*	.10	.05	-.20	
6 Bender <u>z</u> Score	.24	-.40*	-.28	-.25	.51*

*Significant at .01 level

Variable 1 and varibales 2,3,4. The obtained correlations may possibly be accounted for by the nature of the clinical sample. Since the Ss were primarily evidencing difficulties in the academic setting, the coefficients indicate that as age increased the discrepancy in achievement became more pronounced. The significant correlation of age with intelligence may be justified by two possible explanations: (a) the high correlation of intelligence tests with assessed scholastic achievement and, (b) the fact that the younger Ss in this sample generally obtained higher scores on the intelligence test than did the older Ss. In the random selection of Ss, those children whose birthdates fell within the CA range of 74-88 months generally obtained the highest IQs with the poorest Bender performances. Apparently, these children who tend to score below average on the Bender were able, despite the suggested weakness in visual-motor skills, to obtain better than average intelligence quotients.

Variable 1 and 5. The significant negative correlation adds credence to the established fact that the abilities involved in the execution of the BG protocal are maturational in nature.

Variable 5 and variables 2,3,4. The obtained coefficients of intelligence and achievement were not significantly correlated with the Bender error score; while the coefficient obtained between the BG and arithmetic indicates an inverse relationship, the non-significant correlation suggests only a trend in the predicted direction. These findings, therefore, do not basically support those previously reported by Koppitz (1958a, 1958b), but do tend to more generally agree with data reported by Keogh (1965b).

Variables 6 and variables 2,3,4. The higher correlation found between variables 6 and 2, as contrasted to the error score and intelligence (5-2), may possibly be accounted for by the communality of the age component in both the $\underline{z}$ score and the IQ. Therefore, the data suggest that if the BG were to be employed as a useful screening tool for the assessment of intelligence and achievement, the $\underline{z}$ score, rather than the error score, would provide a greater degree of predictability. The achievement scores are derived scores with age as a component. Although significance was not reached in the correlations of the $\underline{z}$ score and achievement, the inverse relationships were evidenced.

REGRESSION ANALYSIS

In addition to the simple correlational analysis, the investigators sought to determine the unique contribution of proper subsets of the predictor variables to the prediction of a specified criterion variable. The contribution of a set of independent variables to prediction may be measured by the difference between two squares of multiple correlation coefficients (RSs), one obtained for a regression model in which all predictors are used, called the full model (FM), and the other obtained for a regression equation in which the proper subset of variables under consideration has been deleted; this model is called the restricted model, (RM). The RS for the RM can never be larger than the RS for the FM. The difference between the two RSs can be tested for statistical significance with the variance ratio test. The hypothesis tested states that these variables contribute nothing to the determination of the expected criterion values that is not already present in the restricted predictive system. There are several possible interpretations of the unique contribution of a variable to the prediction of a criterion. One interpretation is such that if a variable is making a unique contribution, then two Ss, who are unlike on the variable but who are exactly alike or are matched on the other predictors, will differ on the criterion.

In model 1 (Table 2),a 2-variable composite (1,5) was tested for predictability in which variable 4 served as the criterion. The investigators sought to determine the extent to which a knowledge of the age of the S (variable 1) and his error score on the Bender (variable 5) could predict the dependent variable of arithmetic achievement (variable 4). Predictability was low as about 10 percent (.0985) of the criterion variance is estimated to be attributable to the 2 variables in the predictive system. The difference between RS value for the FM and the restricted model, FM-1, yields an estimate of .0960 for the unique contribution of variable 5 which was significant beyond the .01 level. On the other hand, the difference between RS value for the FM system (.0985) and the RS value for the restricted model, FM-5, yields an estimate of .0585 for the unique contribution of variable 1 which was not significant at the .01 level.

In model 2 (Table 2),the criterion variable for the FM is reading (variable 3) with variables 1 and 5 used as predictors. About 18 percent (.1786) of the criterion variance is estimated to be attributable to the 2 predictor variables. Checking this model for significant predictability against chance, the investigators found the predictive efficiency of the model to differ significantly from chance at the .01 level even though it was weak or low from a predictive viewpoint. The estimate for the unique contribution of variable 5 is .0561 which is not significant at the .01 level. However, the unique contribution of variable 1 is estimated to be .1761 which is significant at the .01 level.

Model 3 (Table 2) used as its criterion, variable 2 which is a measure of one's intelligence. This full model was tested for predicability with variables 1 and 5 again serving as predictors. The RS for the FM is .2812 which suggests that about 28 percent of the criterion variance is estimated to be explained by the predictive pair. When checked against change, the predictive efficiency was significant beyond the .01 level. Of the three models investigated, this one had the greatest predictive accuracy. The unique contribution of variable 5 is estimated to be .0696 which is significant at the .01 level. In addition, the unique contribution of variable 1 is estimated to be .2710 which is significant beyond the .01 level.

Table 2

Proportions of Variance Attributable to Groups of Variable Believed to be Associated With Three Criteria

Variable Group	Total Contribution Proportion (RS)	Unique Contribution Probability	Multiple R
Model 1 (1,5--4)	.0985		.31
Model 1-Var. 5		.0960[a]	
Model 1-Var. 1		.0585	
Model 2 (1,5--3)	.1786		.42[b]
Model 2-Var. 5		.0561	
Model 2-Var. 1		.1761[a]	
Model 3 (1,5--2)	.2812		.53[b]
Model 3-Var. 5		.0696[a]	
Model 3-Var. 1		.2710[a]	

[a]All proportions reported as unique are significant at the .01 level for N=50. In computing F values, it was assumed that one parameter was associated with each variable in the predictive system. The degrees of freedom for the number of predictors were determined by the number of variables given an opportunity to contribute to the prediction.

[b]Significant at the .01 level.

It is interesting to note that in comparing the predictive efficiency of the three full regression models with simple correlations obtained by utilizing the Bender z score with the same three criterion variables, in all cases greater predictability existed with the three regression approaches (See Table 3). The regression models utilized age and the Bender error score as predictors of achievement in arithmetic, reading, as well as assessed intelligence. On the other hand, the correlational approach relied solely on the Bender z score as a predictor of the three criterion variables.

Table 3

Proportion of Variance Obtained From the Regression Models and the Bender z Models

Criterion Variable	Regression Model Predictive Efficiency (RS)	Bender z Model Predictive Efficiency (R
1. Arithmetic Achieve	.0985	.0625
2. Reading Achievement	.1786	.0784
3. Intelligence Quotient	.2812	.1600

Summary

The study examined correlations between assessed visual-motor perception, intelligence, and academic achievement. In addition, efficiency of prediction for criterion variables was investigated by employing two approaches of analysis: (a) regression model and (b) Bender z model. The following conculsions were formulated on the basis of the obtained data and from the comparison of the two predictive models.

(1) The significant negative correlation found between age and the Bender error score adds further substantiation to the fact that the ability to correctly execute the Bender protocol improves with increased age.

(2) The Bender z score correlated to a greater degree with intelligence, reading, and arithmetic achievement than did the Bender error score with the three specified variables.

(3) The obtained correlations of the Bender z score with the three criterion variables agrees with the literature in directionality and in significance with assessed intelligence.

(4) However, efficiency is enhanced by using the Bender error score and age rather than the single variable of the Bender z score to predict achievement in reading and arithmetic and assessed intelligence.

Bibliography

Keogh, B. K., The Bender Gestalt as a predictive and diagnostic test of reading performance. J. consult. Psychol., 1965, 29, (1), 83-84.

Koppitz, E. M., The Bender Gestalt Test and learning disturbances in young children. J. clin. Psychol., 1958, 14, 292-295.

Koppitz, E. M., Relationships between the Bender Gestalt Test and the Wechsler Intelligence Scale for Children. J. clin. Psychol., 1958, 14, 413-416.

Ward, J. H., Jr., Multiple linear regression models. Computer Applications in the behavioral sciences, Harold Burko (ed.). Englewood Cliffs, New Jersey: Prentice-Hall, 1962, 204-237.

CURVILINEARITY WITHIN EARLY DEVELOPMENTAL VARIABLES[1,2]

Thomas E. Jordan
University of Missouri at St. Louis

Professor Keith A. McNeil is alive and well and living in Carbondale, Illinois. In some circles there is a suspicion - amounting to a certainty - that he has been here before. Specifically, a number of people believe that he should be known as Isaac Newton McNeil, the well-known appledropper. The cloudy - if not shady - matter of exactly what this fellow Newton was up to under the apple tree has never really been settled; and yet, the matter was not entirely unproductive. Newton established that there exists a relationship between the distance travelled by a falling body and values for the elements g (gravity) and t (time). McNeil (1970) has derived the classic formula:

$$d = \frac{1}{2}gt^2$$

doing so by means of multiple linear regression (Kelly, F.J., Beggs, D.L., McNeil, K. A., Eichelberger, T., & Lyon, J., 1969). He has observed that investigators should include vectors which permit examination of complex relations, such as those illustrated in the Newtonian formula three hundred years ago.

Today, as then, the search for comprehensive models of phenomena sometimes leads to the question of non-linearity. For some time I have been bemused by some suggestions in the literature of early

---

Supported by the National Program for Early Childhood Education (CEMREL), and the Bureau for Education of the Handicapped, USOE, Contract No. OEG-0-70-1204(607).

Paper presented to the American Educational Research Association

child development. The classic papers from Scotland by Drillien (1964), indicate that low birth weight relates to subsequent growth in a manner quite different from that observed when birth weight is normal. Still another relationship is suggested by Babson's (1969) work on overweight babies. Let me add to the consideration one more disparate observation. Extended regression models of early development yield very low accounts of criterion variance. A phenomenon of that sort is rather like Galileo's limited explanation of falling bodies (McNeil, 1970). It may be that the phenomenon is simply not explicable. On the other hand, it may be that complex relationships obtain, and that more elaborate explanations are called for (Jordan & Spaner, 1970).

PROBLEM

This paper is an account of an attempt to raise the predictability of developmental criteria in the first three years of life. The data are drawn from my prospective longitudinal study of one thousand newborns in St. Louis City and County (Jordan, 1971). The 1966 cohort is now four years old, and it is quite representative of the St. Louis metropolitan area's population by SES level and race. Tables I, II and III show the characteristics of the subjects used.

I N S E R T   T A B L E S   I,   I I,   A N D   I I I   A B O U T   H E R E

METHOD

A regression model was generated based on the generally accepted contribution to development of selected variables. The predictors selected were, sex, race, SES level, Apgar score, birth order, birth weight, and birth length. Apgar scores (Apgar and James, 1962) are ratings of physical condition one or five minutes after delivery. SES level was McGuire & White's (1955) index of income, education, and occupation. These basic vectors were supplemented by additional vectors representing squared values for the continuous variable. Table IV shows the full model.

The criteria for the analysis were height and weight (Physical domain) at twelve, twenty four and thirty six months. At those points in time the following measures (cognitive domain) were gathered, Ad Hoc Scale development Score (Jordan, 1967), Preschool Attainment Record - selected subtests (Doll, 1966), and the Peabody Picture Vocabulary Scale (Dunn, 1965).

I N S E R T   T A B L E   I V   A B O U T   H E R E

The regression model in Table IV was applied to nine criteria, three at each birthday. Restricted models were also applied; each of them deleted a vector in the presence of the other vectors. Tables V - XIV list the results grouped for each criterion.

RESULTS

A glance at Table V shows that the phenomenon of low R^2 values for regression models of early development persists. Fluctuations in R^2's seen for height and weight probably represent slight differences in the subjects. They also are quite typical of what many other analyses of data from the 1966 cohort have shown. The full model of the

---

I N S E R T   T A B L E   V   A B O U T   H E R E

---

36 month PPVT scores is quite substantial, as such things go.

We may now turn to the focal point of the analysis, the results of applying the restricted models which delete squared and cubed representations of the continuous vectors. For ease of presentation the results are grouped by age and criteria.

INSERT TABLES VI, VII, AND VIII ABOUT HERE

---

1. Twelve Months. A. Height. Only one vector, the continuous vector for sex, significantly reduced the R^2 value below that of the full model ($F = 12.03$, $p = .0006$). B. Weight. Again, sex reduced the R^2 of the model when deleted ($F = 14.42$, $p = .0001$). C. Ad Hoc Scale. The same result was obtained; sex difference were revealed ($F = 4.15$, $p = .04$).

INSERT TABLES IX, X, AND XI ABOUT HERE

---

2. Twenty Four Months. A Height. Restricted model 2, testing the race vector, was significant ($F = 6.91$, $p = .009$). B. Weight. The sex vector was influential in prediction of the criterion ($F = 7.86$, $p = .005$). C. PAR Total. The same phenomenon was produced when the sex vector was deleted ($F = 7.62$, $p = .006$).

INSERT TABLES XII, XIII, AND XIV ABOUT HERE

---

3. Thirty Six Months. A. Height. The sex vector was significant in

predicting the criterion ($F = 6.39$, $p = .01$). B. Weight. The same effect was evident for weight ($F = 3.96$, $p = .04$), although to a lesser extent. The race vector was significant ($F = 11.58$, $p = .0001$). C. PPVT. The race vector significantly affected prediction of scores ($F = 6.95$, $p = .008$).

DISCUSSION

The results, stated baldly, are to the effect that no squared and cubed vectors contributed to a significant degree in the process of accounting for nine criterion scores. However, there are some aspects of the matter which may be elucidated beyond that undisputed fact. They constitute the remainder of this paper, and deal with its major purpose.

Negative F-Values. Inspection of Tables VI - XIV indicates that a number of negative F-values were generated by applying the Bottenberg and Ward (1963) multiple linear regression program to the data. This tells us that the restricted model in a comparison contains more useable data, i.e. provides a better picture of the phenomena under consideration. The negative F-value should not be interpreted since it is not really a meaningful statement about the R^2 values of the full and restricted model.

Squared and cubed vectors. An inspection of extended R^2 values - within admittedly statistically indifferent restricted models (vis à vis full models, but not model zero) - can illustrate some interesting things about non-linearity. Taking the R^2 values in Tables VI - XIV to four decimal places permits some illustrations - if not conclusions - in order to explore multiple linear regression and non-linear relationships.

The critical elements in the regression models applied to nine criteria were five continuous vectors plus generated vectors representing their squared and cubed values. Usually, additional information increases the predictive power of regression models. In this investigation squared and cubed values occasionally reduced the percentage of criterion variance accounted for. That is, a full regression model incorporating squared and cubed vectors occasionally has a lower R^2 value than models without a squared or cubed vector. Restricted model 5(b) does not contain the $Apgar^2$ vector. Its R^2 value of .2423 is greater than that of the full model which includes $Apgar^2$ ($R^2 = .2417$). Model 18(c) deletes birth $order^3$ from the full model of 12 month development. The result is a higher R^2 (.1055) than the full model (.1035). The same phenomenon is produced by birth $order^3$ for 24 month height. Deletion of birth $order^3$ raises the R^2 value to .1653 from .1636 for the full model. The effect is repeated for 24 month PAR [model 18(f)].

There exists the pattern in which no degree of manipulation of the the quantified form of the independent variable makes a difference. At 36 months the full model of height 1(a) has an R^2 value of .1205. Birth $height^2$ and cubed fail to account for a portion of the height variance. In other words, manipulating a trivial variable does not alter its lack of significance.

Ideally, one would hope to see the use of curvilinearity through squared and then cubed vectors increase the amount of variance accounted for. Birth $weight^1$ is deleted in model 13(h) which accounts for .1694% of 36 month weight variance. Birth $weight^2$ is deleted in model 14(h) which accounts for less of the variance, $R^2 = .1686$; Birth weight^3 [model 15(h)] is more significant (R^2 = .1660). A degree of significance attached to squared vectors but not to cubed vectors is seen in models 6(i) R^2 = .2645: 7(i) SES^2 deleted = .2642: 8(i) SES^3 deleted = .2646.

Significance of cubed vectors, but not squared vectors, may also be illustrated. Birth weight^3, model 15(g), has an R^2 value of .1204, a value which is different from the identical values (.1205) for birth weight^1, and birth weight^2. In some instances vectors seem to have a depressing effect, raising R^2's when deleted. See models 16(c), 17(c), and 18(c). Deletion of birth order^3 raises the R^2 to .1055.

In some instances squared and cubed values of scores may have the same effect. Models 8(a) and 9(a) delete SES^2 and SES^3 in prediction of 12 month height. In both cases an R^2 value of .2238 was obtained. The same effect for the same criterion is seen in models 17(a) and 18(a).

The optimal pattern one would hope for is a steady increment in the value of data as it is squared and cubed, that is, as vectors are erected to encourage non-linear representation of the data. Restricted models of 36 month weight drop in value as the more elaborate vectors of birth weight, birth weight^2, and birth weight^3 are sequentially deleted while the others are retained. Regression model 13(h) deleting birth weight^1 has an R^2 value of .1694. Deletion of the squared vector in model 14(h) reduces the R^2 progressively (.1686), model 15(h), deleting the cubed vector is still lower (R^2 = .1660).

I N S E R T   T A BLL E   X V   A B O U T   H E R E

Consideration of Table XV, the matrix of correlations, is helpful in understanding the phenomena of this investigation. SES date (McGuire

& White, 1955) in the original form relate well to 24 month height (r = -.25). The relationship rises for the squared vector (r = -.27), and a little more for the cubed vector (r = -.28). This is the sort of increase in association one would hope for, although at a less trivial rate of increment.

The opposite effect, a decreasing association between the predictor and the criterion, is seen between weight at birth and at age twelve months. Weight at birth and a year later relate well, r = .34. A lower relationship exists when birth weight is squared, r = .33, and it drops again, when birth weight is present as a cubed value (r = .31).

In the case of birth order it is possible to illustrate a slightly different effect of squared and cubed vectors on correlations. Birth order prime relates to 12 month height insignificantly and negatively, r = -.08. There is a slight rise for birth order^2 (r = -.09), and a subsequent decline for birth order^3 (r = -.08). The reverse of this, a rise for the squared predictor and a decline for the cubed, is illustrated in the correlations for birth order and 12 month development. Finally, there is the situation in which no manipulation of the data into squared and cubed forms has any effect. Birth height in its original, squared and cubed forms shows unchanged correlations with several criteria in Table XV, 12 month weight, 24 month height, 24 month weight, 36 month height, and 36 month weight.

The preceding illustrations may now be used to generate some remarks about non-linearity.

1. The data of this report failed to reveal any instances of significant non-linearity within data from the first three years of life in several domains.

2. 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.

3. Squared and cubed vectors may lead to proportionately better accounts of criterion variance. However, it does not follow that higher order exponents will progressively help. There is probably a point at which no great advantage continues to accrue. The principle of diminishing return for greater effort probably applies.

4. Multiple linear regression permits a quantitatively satisfactory view of developmental data.

SUMMARY

Squared and cubed vectors were introduced into eighteen regression models each applied to nine criteria. Data came from study of several hundred children in the first three years of life. Departure from linearity did not provide better accounts of the relationship between five predictors and development at 12, 24, and 36 months of age. Illustrations of various patterns of $vectors^2$ and $vectors^3$ were presented from the data.

BIBLIOGRAPHY

1. Apgar, V., & James, L.S., "Further Observations on the Newborn Scoring System," Amer. J. Dis. Child., 1962, 104, 419-428.

2. Babson, S.G., Henderson, N.B., & Clark, W.M., "The Preschool Intelligence of Oversized Newborns," paper presented to the American Psychological Association, 1969.

3. Bottenberg, R., & Ward, J.H., Applied Multiple Linear Regression, U.S. Dept. of Commerce, Washington, D.C., 1963.

4. Drillien, C.M., The Growth and Development of the Prematurely Born Infant, E. & S. Livingstone, London, 1964.

5. Doll, E.A., The Preschool Attainment Record, American Guidance Service, 1966.

6. Dunn, L.M., The Peabody Picture Vocabulary Scale, American Guidance Service, 1965.

7. Jordan, T.E., EDAP Technical Note #3.1: Extension of the Ad Hoc Development Scale, CEMREL, St. Louis, Mo., 1968.

8. Jordan, T.E., "Early Developmental Adversity and the First Two Years of Life," Multivar. Behav. Res. Mon., 1971, 1.

9. Jordan, T.E., & Spaner, S.D., "Biological and Ecological Influences on Development at Twelve Months of Age," Hum. Devpm., 1970, 13, 178-187.

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

11. McGuire, C.M., & White, G., "The Measurement of Social Status," Res. Paper Hum. Devpm., #3, University of Texas, 1955.

12. McNeil, K., "Meeting the Goals of Research with Multiple Linear Regression," Multivar. Behav. Res., 1970, 5, 375-386.

TABLE I

RANGES, MEANS AND STANDARD DEVIATIONS OF PREDICTORS, AND CRITERION VARIABLES AT TWELVE MONTHS

(N = 217)

Predictor Variable	Race (%W)	Sex (%M)	Apgar	SES	B.Height (in.)	B.Weight (lb.)	B.Order
Range			2-10	16-84	16-23	3.43-12.00	1-11
Mean	75	58	8.78	51.12	19.95	7.31	2.82
Standard Deviation			1.26	16.55	1.35	1.22	2.20

Criterion Variable	Weight (lb.)	Height (in.)	Ad Hoc Score
Range	15.98-30.01	24-36	9-19
Mean	22.31	29.64	14.93
Standard Deviation	2.57	1.59	2.04

TABLE II

RANGES, MEANS, AND STANDARD DEVIATIONS OF PREDICTORS, AND CRITERION VARIABLES AT TWENTY FOUR MONTHS

(N = 277)

Predictor Variable	Race (%W)	Sex (%M)	Apgar	SES	B.Height (in.)	B.Weight (lb.)	B.Order
Range			2-10	16-84	16-23	3.43-12.00	1-11
Mean	64	54	8.90	54.06	19.75	7.22	2.89
Standard Deviation			1.22	16.37	1.35	1.79	2.24

Criterion Variable	Weight (lb.)	Height (in.)	PAR Total
Range	20-43	25-39	14-40
Mean	28.05	33.50	26.07
Standard Deviation	3.63	2.65	4.79

TABLE III

RANGES, MEANS AND STANDARD DEVIATIONS OF PREDICTORS, AND CRITERION VARIABLES AT THIRTY SIX MONTHS

(N = 321)

Variable	Race (%W)	Sex (%M)	Apgar	SES	B.Height (in.)	B.Weight (lb.)	B.Order	Weight (lb.)	Height (in.)	PPVT
Range			2-10	16-84	16-23	2.43-12.00	1-13	22.50-46.00	25.00-55.50	0-54
Mean	58	52	9.00	55.31	19.72	7.29	2.84	31.66	37.78	24.09
Standard Deviation			1.20	16.45	1.35	1.18	1.05	3.82	1.99	11.50

TABLE IV

FULL REGRESSION MODEL FOR CRITERIA (a) - (i)

$$Y_{a-i} = a_o u + \text{race} + \text{sex} + \text{Apgar} + \text{Apgar}^2 + \text{Apgar}^3 + \text{SES} + \text{SES}^2 + \text{SES}^3$$
$$+ \text{birth ht.} + \text{birth ht.}^2 + \text{birth ht.}^3 + \text{birth wt.} + \text{birth wt.}^2$$
$$+ \text{birth wt.}^3 + \text{birth order} + \text{birth order}^2 + \text{birth order}^3 + e$$

TABLE V

R^2 AND SIGNIFICANCE OF THE DIFFERENCE FROM ZERO OF FULL

REGRESSION MODELS FOR NINE CRITERIA

Model		Criterion	R^2	F	P
1 (a)	(12 Mos.)	Height	.2238	3.60	.00001
1 (b)		Weight	.2417	3.98	<.00001
1 (c)		Devpm. Score	.1035	1.44	.12
1 (d)	(24 Mos.)	Height	.1636	3.17	.00005
1 (e)		Weight	.1155	2.12	.007
1 (f)		PAR Total	.0927	1.66	.05
1 (g)	(36 Mos.)	Height	.1205	2.60	<.00001
1 (h)		Weight	.1700	3.89	<.00001
1 (i)		PPVT	.2647	6.84	<.00001

TABLE VI

COMPARISON OF REGRESSION MODELS FOR CRITERION: TWELVE MONTH HEIGHT

Models Compared***	$R^{2\#}$	P*	F	P
Full Model and Model 2 (a)	.2238	<.00001	.00	1.00
Full Model and Model 3 (a)	.1771	<.0001	12.03	.0006
Full Model and Model 4 (a)	.2244	<.00001	-.15	1.00**
Full Model and Model 5 (a)	.2238	<.00001	.00	1.00
Full Model and Model 6 (a)	.2238	<.00001	.00	1.00
Full Model and Model 7 (a)	.2228	.00001	.24	.61
Full Model and Model 8 (a)	.2238	<.00001	.00	1.00
Full Model and Model 9 (a)	.2238	<.00001	.001	.97
Full Model and Model 10 (a)	.2241	<.00001	-.07	1.00**
Full Model and Model 11 (a)	.2238	<.00001	.00	1.00
Full Model and Model 12 (a)	.2234	<.00001	.10	.74
Full Model and Model 13 (a)	.2214	.00001	.60	.43
Full Model and Model 14 (a)	.2239	<.00001	-.02	1.00**
Full Model and Model 15 (a)	.2272	<.00001	-.87	1.00**
Full Model and Model 16 (a)	.2238	<.00001	.006	.93
Full Model and Model 17 (a)	.2228	.00001	.26	.60
Full Model and Model 18 (a)	.2228	.00001	.26	.60

* = Significance of the difference from R^2 = .0 of the restricted Model R^2
** = Negative R-ratio yields an uninterpretable probability statement
*** = R^2 Full Model = .2238
# = Restricted Model

TABLE VII

COMPARISON OF REGRESSION MODELS FOR CRITERION: TWELVE MONTH WEIGHT

Models Compared***	R^2 #	P*	F	P
Full Model and Model 2 (b)	.2422	<.00001	-.10	1.00**
Full Model and Model 3 (b)	.1870	.0001	14.42	.0001
Full Model and Model 4 (b)	.2418	<.00001	-.02	1.00**
Full Model and Model 5 (b)	.2423	<.00001	-.15	1.00**
Full Model and Model 6 (b)	.2413	<.00001	.13	.71
Full Model and Model 7 (b)	.2418	<.00001	.00	1.00
Full Model and Model 8 (b)	.2417	<.00001	.00	1.00
Full Model and Model 9 (b)	.2413	<.00001	.11	.73
Full Model and Model 10 (b)	.2372	<.00001	1.19	.27
Full Model and Model 11 (b)	.2417	<.00001	.00	1.00
Full Model and Model 12 (b)	.2369	<.00001	1.25	.25
Full Model and Model 13 (b)	.2415	<.00001	.06	.79
Full Model and Model 14 (b)	.2417	<.00001	.00	1.00
Full Model and Model 15 (b)	.2416	<.00001	.02	.86
Full Model and Model 16 (b)	.2433	<.00001	-.40	1.00**
Full Model and Model 17 (b)	.2417	<.00001	.00	1.00
Full Model and Model 18 (b)	.2413	<.00001	.11	.73

* = Significance of the difference from R^2 = .0 of the restricted Model R^2
** = Negative F-ratio yields an uninterpretable probability statement
*** = R^2 Full Model = .2417
\# = Restricted Model

TABLE VIII

COMPARISON OF REGRESSION MODELS FOR CRITERION: TWELVE MONTH DEVELOPMENT

Models Compared***	$R^{2\#}$	P*	F	P
Full Model and Model 2 (c)	.1049	.12	-.31	1.00**
Full Model and Model 3 (c)	.0849	.24	4.15	.04
Full Model and Model 4 (c)	.1071	.07	-.80	1.00**
Full Model and Model 5 (c)	.1035	.09	.00	1.00
Full Model and Model 6 (c)	.1055	.08	-.44	1.00**
Full Model and Model 7 (c)	.1035	.09	.00	1.00
Full Model and Model 8 (c)	.1035	.09	.01	.90
Full Model and Model 9 (c)	.1051	.08	-.35	1.00**
Full Model and Model 10 (c)	.1034	.09	.02	.88
Full Model and Model 11 (c)	.1035	.09	.00	1.00
Full Model and Model 12 (c)	.1017	.10	.41	.52
Full Model and Model 13 (c)	.1011	.10	.54	.46
Full Model and Model 14 (c)	.1038	.08	-.06	1.00**
Full Model and Model 15 (c)	.1055	.08	-.43	1.00**
Full Model and Model 16 (c)	.0960	.13	1.68	.19
Full Model and Model 17 (c)	.1022	.09	.30	.58
Full Model and Model 18 (c)	.1055	.08	-.44	1.00**

* = Significance of the difference from R^2 = .0 of the restricted model R^2
** = Negative F-ratio yields an uninterpretable probability statement
*** = R^2 Full Model = .1035
\# = Restricted Model

TABLE IX

COMPARISON OF REGRESSION MODELS FOR CRITERION: TWENTY FOUR MONTH HEIGHT

Models Compared***		R^2 #	P*	F	P
Full Model and Model 2	(d)	.1413	.0003	6.91	.009
Full Model and Model 3	(d)	.1623	.00003	.07	.77
Full Model and Model 4	(d)	.1634	.00003	.05	.82
Full Model and Model 5	(d)	.1636	.00003	.00	1.00
Full Model and Model 6	(d)	.1634	.00003	.04	.84
Full Model and Model 7	(d)	.1636	.00003	-.02	1.00**
Full Model and Model 8	(d)	.1636	.00003	.00	1.00
Full Model and Model 9	(d)	.1643	.00003	-.24	1.00
Full Model and Model 10	(d)	.1630	.00003	.16	.68
Full Model and Model 11	(d)	.1636	.00003	.00	1.00
Full Model and Model 12	(d)	.1635	.00003	.02	.86
Full Model and Model 13	(d)	.1638	.00003	-.06	1.00**
Full Model and Model 14	(d)	.1636	.00003	.00	1.00
Full Model and Model 15	(d)	.1634	.00003	.04	.83
Full Model and Model 16	(d)	.1623	.00003	.37	.54
Full Model and Model 17	(d)	.1635	.00003	.01	.90
Full Model and Model 18	(d)	.1653	.00002	-.55	1.00**

* = Significance of the difference from R^2 = .0 of the restricted model R^2
** = Negative F-ratio yields an uninterpretable probability statement
*** = R^2 Full Model = .1636
# = Restricted Model

TABLE X

COMPARISON OF REGRESSION MODELS FOR CRITERION: TWENTY FOUR MONTH WEIGHT

Models Compared***	$R^{2\#}$	P*	F	P
Full Model and Model 2 (e)	.1166	.004	-.03	1.00**
Full Model and Model 3 (e)	.0888	.05	7.86	.005
Full Model and Model 4 (e)	.1162	.004	-.20	1.00**
Full Model and Model 5 (e)	.1155	.004	.00	1.00
Full Model and Model 6 (e)	.1144	.005	.33	.56
Full Model and Model 7 (e)	.1188	.003	-.09	1.00**
Full Model and Model 8 (e)	.1155	.004	.00	1.00
Full Model and Model 9 (e)	.1190	.003	-1.03	1.00**
Full Model and Model 10 (e)	.1213	.002	-1.69	1.00**
Full Model and Model 11 (e)	.1155	.004	.00	1.00
Full Model and Model 12 (e)	.1162	.004	-.20	1.00**
Full Model and Model 13 (e)	.1138	.005	.51	.47
Full Model and Model 14 (e)	.1156	.004	-.03	1.00**
Full Model and Model 15 (e)	.1155	.004	.00	1.00
Full Model and Model 16 (e)	.1150	.005	.16	.68
Full Model and Model 17 (e)	.1117	.007	1.13	.28
Full Model and Model 18 (e)	.1084	.009	2.07	.15

* = Significance of the difference from R^2 = .0 of the restricted model R^2
** = Nagative F-ratio yields an uninterpretable probability statement
*** = R^2 Full Model = .1155
# = Restricted Model

TABLE XI

COMPARISON OF REGRESSION MODELS FOR CRITERION: TWENTY FOUR MONTH PAR TOTAL

Models Compared***	$R^{2\#}$	P*	F	P
Full Model and Model 2 (f)	.0822	.08	3.00	.08
Full Model and Model 3 (f)	.0660	.24	7.62	.006
Full Model and Model 4 (f)	.0935	.03	-.24	1.00**
Full Model and Model 5 (f)	.0927	.03	.00	1.00
Full Model and Model 6 (f)	.0933	.03	-.17	1.00**
Full Model and Model 7 (f)	.0932	.03	-.16	1.00**
Full Model and Model 8 (f)	.0927	.03	.00	1.00
Full Model and Model 9 (f)	.0930	.03	-.10	1.00**
Full Model and Model 10 (f)	.0916	.04	.30	.58
Full Model and Model 11 (f)	.0927	.03	.00	1.00
Full Model and Model 12 (f)	.0917	.04	.26	.60
Full Model and Model 13 (f)	.0927	.03	-.01	1.00**
Full Model and Model 14 (f)	.0927	.03	.00	1.00
Full Model and Model 15 (f)	.0928	.03	-.05	1.00**
Full Model and Model 16 (f)	.0936	.03	-.26	1.00**
Full Model and Model 17 (f)	.0927	.03	.00	1.00
Full Model and Model 18 (f)	.0937	.03	-.31	1.00**

* = Significance of the difference from $R^2 = .0$ of the restricted model R^2
** = Nagative F-ratio yields an uninterpretable probability statement
*** = R^2 Full Model = .0927
\# = Restricted Model

TABLE XII

COMPARISON OF REGRESSION MODELS FOR CRITERION: THIRTY SIX MONTH HEIGHT

Models Compared***	R^2#	P	F	P
Full Model and Model 2 (g)	.1201	.0004	.13	.71
Full Model and Model 3 (g)	.1020	.004	6.39	.01
Full Model and Model 4 (g)	.1213	.0004	-.30	1.00**
Full Model and Model 5 (g)	.1205	.0004	.00	1.00
Full Model and Model 6 (g)	.1192	.0005	.42	.51
Full Model and Model 7 (g)	.1205	.0004	.00	1.00
Full Model and Model 8 (g)	.1205	.0004	.00	1.00
Full Model and Model 9 (g)	.1208	.0004	-.11	1.00**
Full Model and Model 10 (g)	.1204	.0004	.02	.86
Full Model and Model 11 (g)	.1204	.0004	.03	.85
Full Model and Model 12 (g)	.1204	.0004	.03	.85
Full Model and Model 13 (g)	.1205	.0004	.00	1.00
Full Model and Model 14 (g)	.1205	.0004	.00	1.00
Full Model and Model 15 (g)	.1204	.0004	.01	.89
Full Model and Model 16 (g)	.1198	.0005	.22	.63
Full Model and Model 17 (g)	.1205	.0004	.00	1.00
Full Model and Model 18 (g)	.1206	.0004	-.03	1.00 **

* = Significance of the difference from R^2 = .0 of the restricted model R^2
** = Negative F-ratio yields an uninterpretable probability statement
*** = R^2 Full Model = .1205
\# = Restricted Model

TABLE XIII

COMPARISON OF REGRESSION MODELS FOR CRITERION: THIRTY SIX MONTH WEIGHT

Models Compared***		$R^{2\#}$	P*	F	P
Full Model and Model 2	(h)	.1383	.00005	11.58	.0001
Full Model and Model 3	(h)	.1592	<.00001	3.96	.04
Full Model and Model 4	(h)	.1685	<.00001	.54	.46
Full Model and Model 5	(h)	.1700	<.00001	.00	1.00
Full Model and Model 6	(h)	.1687	<.00001	.48	.48
Full Model and Model 7	(h)	.1708	<.00001	-.29	1.00**
Full Model and Model 8	(h)	.1700	<.00001	.00	1.00
Full Model and Model 9	(h)	.1724	<.00001	-.88	1.00**
Full Model and Model 10	(h)	.1687	<.00001	.47	.49
Full Model and Model 11	(h)	.1700	<.00001	.00	1.00
Full Model and Model 12	(h)	.1691	<.00001	.32	.56
Full Model and Model 13	(h)	.1694	<.00001	.22	.63
Full Model and Model 14	(h)	.1686	<.00001	.52	.47
Full Model and Model 15	(h)	.1660	<.00001	1.46	.22
Full Model and Model 16	(h)	.1688	<.00001	.43	.50
Full Model and Model 17	(h)	.1690	<.00001	.34	.55
Full Model and Model 18	(h)	.1696	<.00001	.14	.67

* = Significance of the difference from R^2 = .0 of the restricted model R^2
** = Negative F-ratio yields an uninterpretable probability statement
*** = R^2 Full Model = .1700
\# = Restricted Model

TABLE XIV

COMPARISON OF REGRESSION MODELS FOR CRITERION: THIRTY SIX MONTH PPVT

Models Compared***		$R^{2\#}$	P*	F	P
Full Model and Model 2	(i)	.2479	<.00001	6.95	.008
Full Model and Model 3	(i)	.2647	<.00001	.00	1.00
Full Model and Model 4	(i)	.2644	<.00001	.15	.69
Full Model and Model 5	(i)	.2647	<.00001	.00	1.00
Full Model and Model 6	(i)	.2638	<.00001	.35	.55
Full Model and Model 7	(i)	.2645	<.00001	.08	.76
Full Model and Model 8	(i)	.2642	<.00001	.20	.65
Full Model and Model 9	(i)	.2646	< .0001	.03	.85
Full Model and Model 10	(i)	.2647	<.00001	.00	1.00
Full Model and Model 11	(i)	.2647	<.00001	.00	1.00
Full Model and Model 12	(i)	.2648	<.00001	.00	1.00
Full Model and Model 13	(i)	.2648	<.00001	-.03	1.00**
Full Model and Model 14	(i)	.2647	<.00001	.00	1.00
Full Model and Model 15	(i)	.2651	<.00001	-.17	1.00**
Full Model and Model 16	(i)	.2624	<.00001	.97	.32
Full Model and Model 17	(i)	.2645	<.00001	.10	.74
Full Model and Model 18	(i)	.2649	<.00001	.00	1.00

* = Significance of the difference from R^2 = .0 of the restricted model R^2

** = Negative F-ratio yields an uninterpretable probability statement

*** = R^2 Full Model = .2647

\# = Restricted Model

1 R^2

TABLE XV

CORRELATION MATRIX

	12 Month Height[a]	12 Month Weight[a]	12 Month Devpm.[a]	24 Month Height[b]	24 Month Weight[b]	24 Month PAR Total[b]	36 Month Height[c]	36 Month Weight[c]	36 Month PPVT[c]
Race (N)	-.06	-.12*	.05	.34**	-.04	.00	-.06	.03	-.41**
Sex (M)	-.29**	.29**	-.15*	.00	.22**	-.15**	.17	.19**	.00
Apgar	.10	-.05	.08	-.09	-.03	.08	.06	-.02	-.18**
Apgar^2	.09	-.06	.10	-.11	-.04	.07	.07	-.02	-.20**
Apgar^3	.08	-.06	.11	-.12*	-.05	.06	.07	-.02	-.21**
SES	-.04	.01	.01	-.25**	-.03	-.08	-.11*	-.05	-.44**
SES^2	-.05	-.03	.02	-.27**	-.05	-.10	-.12*	-.06	-.45**
SES^3	-.06	-.06	.02	-.28**	-.06	-.11	-.13*	-.08	-.45**
B. Height	.35**	.26**	-.01	.25**	.20**	.04	.22**	.26**	.11*
B. Height^2	.34**	.26**	-.09	.25**	.20**	.03	.22**	.26**	.11*
B. Height^3	.34**	.26**	-.10	.25**	.20**	.03	.22**	.26**	.12*
B. Weight	.32**	.34**	.01	.13*	.21**	-.02	.19**	.29**	.10
B. Weight^2	.29**	.33**	.00	.13*	.20**	-.03	.19**	.29**	.09
B. Weight^3	.26**	.31**	-.01	.12*	.20**	-.04	.18**	.29**	.08
B. Order	-.08	.03	.02	-.08	-.08	-.15**	.01	.00	-.17**
B. Order^2	-.09	.03	.01	-.10	-.09	-.15**	.00	-.01	-.17**
B. Order^3	-.08	.03	.02	-.10	-.10	-.13*	-.01	-.02	-.16**

a,b,c = df 215, 275, 319

*p = <.05

**p = <.01

REGRESSION MODELS IN EDUCATIONAL RESEARCH

Cheryl L. Reed, John F. Feldhusen, and Adrian P. Van Mondfrans

Purdue University

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. Some predictor variables have also been shown to be curvilinearily related to a criterion, such as anxiety to achievement (Fein, 1963), but very rarely has this type relationship been included as a possibility within the regression model. The purpose of this paper is to propose the use of and state some research support for the use of higher order regression models within educational research.

Moderator variables as suggested by Ghiselli (1956) and Saunders (1956) lend somewhat limited support for the use of more complex models. Moderators improve prediction by acknowledging possible interactive effects of the moderator variable with other variables in the regression analysis. Fellows (1967) compared a regression equation with a moderator and a predictor to a regression equation with a moderator, predictor, and a cross-product of the two. The cross-product produced the moderating effect and resulted in more efficient prediction. Rock (1969) suggested that when moderators do exist, a more complex regression model may give satisfactory predictor-criterion fit without subdividing the sample. The authors also feel that the more complex regression model allows more meaningful interpretation by pinpointing the predictor-moderator interactions which improve prediction in the moderated system.

Several comparitive studies of predictive efficiency of first-order and second-order regression models have been made. Rock (1965) compared a regression equation containing linear and quadratic terms to one containing cross-product terms. He found the interaction term regression to be superior to the quadratic form in predictive efficiency. Whiteside (1964), in a study predicting high school grade point average, found that both interaction and quadratic terms were significant and reliable predictors.

Linear prediction models have been assumed almost exclusively (Lavin, 1965) in the prediction of success in nursing school. However, this may not actually be the case. Fein (1963) reported finding a curvilinear relationship between achievement and anxiety at a school of nursing. Personality variables shown to be predictive of achievement have been highly school specific (Thurston, Brunclik and Feldhusen, 1970). This may indicate that interaction terms involving the school in which the student is enrolled could be important predictors in studies involving more than one school.

The authors, however, have compared regression models containing both interaction and quadratic terms to those containing only first-order terms in the prediction of academic achievement in nursing school.

The results presented here are part of an ongoing research project on the prediction of student achievement within nursing education programs. Students from five schools of nursing entering from 1964 to 1968 were involved in the study. Scores on a number of cognitive affective and biographical variables were secured for these students. The variables included:

1. Scholastic Aptitude Test (SAT) both verbal and mathematics subscores.
2. Creativity Self Rating Scale (C-R) for fluency, flexibility, and overall subscore (Feldhusen, Denny, and Condon, 1965).
3. Parents occupations, weighted on the seven point scale according to the Index of Social Position (Hollingshead and Redlich, 1958).
4. Number of years of education of each parent.
5. Taylor Manifest Anxiety Scale score (Taylor, 1953).
6. Sarason Test Anxiety Scale score (Sarason, 1952).
7. Percentile rank in high school graduating class.
8. Years of student's education prior to entering nursing school.
9. Age in months upon entering nursing school.
10. Year of entry into the nursing program.

The school attended by the student was also included in the analysis as a categorical variable. Using these scores as variables, first and second semester grades were predicted. First semester grades were included in the battery when second semester grades were predicted.

Several models were compared for predictive efficiency. The first model, given below, is the first-order linear model assumed traditionally in regression studies.

Model 1: $Y_1 = b_o + b_1X_1 + b_2X_2 + \ldots + b_{19}X_{19} + E_1$

Where: Y_1 is the criterion.

b_o to b_{19} are weighting coefficients selected to minimize the sum of squared components in E_1.

E_1 is the difference between the predicted and actual Y_1 value.

X_1 is the subject's score on variable one.

X_2 is the subject's score on variable two.

Etc.

The second regression model, given below, includes interaction or cross-product terms. This is an extension of the moderator approach discussed earlier.

Model 2: $Y_1 = b_o + b_1X_1 + b_2X_2 + \ldots + b_{19}X_{19} + b_{20}X_1X_2 + b_{21}X_1X_3 + \ldots + b_{209}X_{18}X_{19} + E_2$

Where: Y_1, X_1 through X_{19}, and the b_i are defined as in model 1.

X_1X_2 is the cross-product of the value corresponding to X_1 and the value corresponding to X_2.

X_1X_3 is the cross-product of the value corresponding to X_1 and the value corresponding to X_3.

Etc.

E_2 is the difference between the predicted and actual Y value.

The third model allows for a second degree curve and is given below:

Model 3: $Y_1 = b_0 + b_1X_1 + \ldots + b_{19}X_{19} + b_{20}X_1^2 + b_{21}X_2^2 + \ldots + b_{38}X_{19}^2 + E_3$

Where: All symbols are defined as in model one, except

X_1^2 is the squared value of the corresponding value of X_1.

X_2^2 is the squared value of the corresponding value of X_2.

Etc.

E_3 is the difference between the predicted and actual Y_1 value.

Finally, the fourth model combines the others into a full second-order regression model. That is, first order, quadratic, and interaction terms were included in the model given below:

Model 4: $Y_1 = b_0 + b_1X_1 + \ldots b_{19}X_{19} + b_{20}X_1^2 + \ldots + b_{38}X_{19}^2 + b_{39}X_1X_2 + \ldots + b_{229}X_{18}X_{19} + E_4$.

Where: All the symbols are defined as in previous models.

Results of the analyses using the four models are summarized in tables one through four. In order to assess the usefulness of the higher order models, each was compared to model one using the F statistic given in Kelly, Beggs, and McNeil (1969, p. 86). In all cases models two through four did show significantly higher multiple correlations (at the .05 level or beyond) than the linear model. Models two through four were not statistically compared to each other.

Cross-validation on the regression equations resulted in the largest amount of shrinkage when models two and four were used to predict semester one grade point average. Since models one and three performed well under the cross validation, it may be that one or more of the interaction terms is not a stable nor a reliable predictor. This may be an example of a problem with the use of higher order polynomials discussed by Kelly et. al. (1969, p. 1960). Two variables which were only moderately reliable may have been multiplied together, resulting in a geometric increase in unreliability. Also it is a possibility that one or more of the interaction terms is not applicable to the cross validation sample.

Results of this study do not support the findings of Rock (1965) that the interaction term regression was superior to the quadratic form in predictive efficiency. The most efficient regression model will depend upon: 1. how the variables and criterion are related, 2. The reliability of the predictor variables, and 3. the research question asked.

The studies reviewed in this paper seem to indicate that complex regression models are in some cases more efficient predictors of complex behavior than the most frequently assumed first order model. When quadratic and interaction terms are significant, however, interpretation is made more difficult (Darlington, 1968). Still, an attempt at interpretation seems somewhat better than ignoring the problem or assuming it does not exist. The shortest distance between two points may be a straight line, but the obstacles between the points often deter us from this line of travel.

Table I

SUMMARY OF RESULTS USING MODEL 1

Semester One Grade Point Average N=495

Variable	
1. Age in months	Multiple correlation = .55
2. SAT math	
3. SAT verbal	Standard error of estimate = .68
4. Previous education	
5. High School Rank	$R_{Cross-validation}$ = .52

Semester Two Grade Point Average N=439

Variable	
1. Year of entry	Multiple correlation = .79
2. Age in months	
3. High School Rank	Standard error of estimate = .49
4. School 3	
5. Grades semester one	$R_{Cross-validation}$ = .74

Table II

SUMMARY OF RESULTS USING MODEL 2

Semester One Grade Point Average N=495

Variable	
1. Age in months	Multiple correlation = .59
2. SAT math	
3. SAT verbal	Standard error of estimate = .66
4. Previous education	
5. High School Rank	$R_{Cross-validation}$ = .49
6. Year of entry X C-R Fluency	
7. Age X High School Rank	$F_{model\ 1\ vs\ model\ 2}$ = 6.02
8. SAT math X previous education	
9. SAT math X High School Rank	
10. Previous education X High school rank	

Table II cont'd.

Semester Two Grade Point Average N=439

Variable	
1. Year of Entry	Multiple correlation = .81
2. Age in months	
3. SAT verbal	Standard error of estimate = .48
4. High School Rank	
5. School 3	$R_{Cross\text{-}validation}$ = .74
6. Grades Semester one	
7. Year X Sarason Anxiety	$F_{model\ 1\ vs\ model\ 2}$ = 7.82
8. SAT verbal X Grades Semester One	
9. School 3 X High School Rank	

Table III

SUMMARY OF RESULTS USING MODEL 3

Semester One Grade Point Average N=495

Variable	
1. Age in months	Multiple correlation = .58
2. SAT math	
3. SAT verbal	Standard error of estimate = .67
4. Previous education	
5. High school rank	$R_{Cross\text{-}validation}$ = .55
6. Age^2	
7. Previous $education^2$	$F_{model\ 1\ vs\ model\ 3}$ = 5.99
8. High school $rank^2$	

Semester Two Grade Point Average N=439

Variable	
1. Year of entry	Multiple correlation = .80
2. Age in months	
3. High school rank	Standard error of estimate = .49
4. School 3	
5. Grades semester one	$R_{Cross\text{-}validation}$ = .76
6. High school $rank^2$	
7. Grades semester one^2	$F_{model\ 1\ vs\ model\ 3}$ = 6.24

Table IV

SUMMARY OF RESULTS USING MODEL 4

Semester One Grade Point Average N=495

Variable	
1. Age in months	Multiple correlation = .60
2. SAT math	
3. SAT verbal	Standard error of estimate = .66
4. Previous education	
5. High school rank	$R_{\text{Cross-validation}}$ = .51
6. Year X C-R fluency	
7. SAT math X Previous education	$F_{\text{model 1 vs model 4}}$ = 6.27
8. SAT math X High school rank	
9. Age X High school rank	
10. Age^2	
11. $Previous\ education^2$	

Semester Two Grade Point Average N=439

Variable	
1. Year of entry	Multiple correlation = .80
2. Age in months	
3. High school rank	Standard error of estimate = .48
4. School 3	
5. Grades semester one	$R_{\text{Cross-validation}}$ = .76
6. School 3 X High school rank	
7. $Grades\ semester\ one^2$	$F_{\text{model 1 vs model 4}}$ = 6.02
8. $High\ school\ rank^2$	

Whiteside, R. Dimensions of teacher evaluation of academic achievement. Unpublished doctoral dissertation, The University of Texas, Austin, Texas, 1964.

References

Darlington, R. B. Multiple regression in psychological research and practice. Psychological Bulletin, 1968, 69, 161-182.

Fein, L. G. Evidence of a curvilinear relationship between IPAT anxiety and achievement at nursing school. Journal of Clinical Psychology, 1963, 19, 374-376.

Feldhusen, J., Denny, T., and Condon, C. F. Anxiety, divergent thinking, and achievement. Journal of Educational Psychology, 1965, 56, 40-45.

Fellows, T. T., Jr. A comparison of multiple regression equations and moderated multiple regression equations in predicting scholastic success. Unpublished doctoral dissertation, University of Utah, Salt Lake City, Utah, 1967.

Ghiselli, E. E. Differentiation of individuals in terms of their predictability. Journal of Applied Psychology, 1956, 40, 374-377.

Hollingshead, A. B. and Redlich, F. C. Social Class and Mental Health, John Wiley & Sons, 1958.

Kelly, F. J., Beggs, D. L., and McNeil, K. A. Multiple Regression Approach, Carbondale, Illinois: Southern Illinois University Press, 1969.

Lavin, D. E. The Prediction of Academic Performance. New York: Russell Sage Foundation, 1965.

Reed, Cheryl L. The prediction of attrition in an associate degree nursing program using cognitive and noncognitive predictor variables. Unpublished master's thesis, Purdue University, 1970.

Rock, D. A. Improving the prediction of academic achievement by population moderators. Unpublished doctoral dissertation, Purdue University, 1965.

Rock, D. A., Evans, F. R., and Klein, S. P. Predicting multiple criteria of creative achievements with moderator variables. Journal of Educational Measurement, 1969, 6(4), 229-235.

Sarason, G. B., Mandler, G. Some correlates of test anxiety. Journal of Abnormal and Social Psychology, 1952, 47, 810-817.

Taylor, Janet A. A personality scale of manifest anxiety. Journal of Abnormal and Social Psychology, 1953, 16(2), 285-290.

Thurston, J. R., Brunclik, Helen L., and Feldhusen, J. F. Comprehensive Manual for Use with Luther Hospital Sentence Completions; Nursing Sentence Completions; Nursing Attitudes Inventory, Forms I and II; Empathy Inventory, Wisconsin: Nursing Research Associates, 1970.

Directional Hypotheses With the Multiple Linear Regression Approach

Keith A. McNeil and Donald L. Beggs
Southern Illinois University

Abstract

Two well known directional tests of significance are presented within the multiple linear regression framework. Adjustments on the computed probability level are indicated. The case for a directional interaction research hypothesis is defended. Conservative adjustments on the computed probability level are offered and a more precise computation is requested of statisticians. Emphasis is placed more on the research question being asked than on blind adherence to conventional formulae.

Introduction

The generalized F ratio within the context of multiple linear regression is known to be applicable to a large number of research questions. There is a class of questions, though, which requires an adjustment in the probability level which is reported by canned computer programs. This reported probability level is for an equally divided "two-tailed" test of significance, but often the researcher has justified a "one-tailed" test of significance. Indeed, whenever the research hypothesis contains directionality, then the required test of significance is "one-tailed." A good deal of the research hypotheses that appear in the literature develop a valid rationale for directionality but very few of them proceed to fully take advantage of their stated alpha level. One only needs to look at, for example, Volume 11 of the Journal of Personality and Social Psychology. Numerous articles in this issue propose directional hypotheses and proceed to use a non-directional test. Indeed, Levinger and Schneider (1969) indicate that the results for one hypothseis was significant in the direction opposite to that hypothesized. In reliability and validity research, the research hypothesis of necessity must be directional. It is seldon that a researcher gets excited about a negative reliability coefficient. Likewise, the researcher hypothesizes the sign of the correlational value indicating validity. A negative correlation would only be expected when two scales are measuring the same phenomenon, but one scale has been reversed. (In this case we would still have all of the critical region in one tail of the sampling distribution.)

There are at least three situations that might require a "one-tailed" test of significance: (1) a research hypothesis suggesting one treatment resulting in a higher mean than another treatment; (2) a research hypothesis specifying either a positive correlation between two variables or a negative correlation between two variables; and (3) a research hypothesis specifying a directional interaction. The first two situations are well documented in the statistical literature, but the last is not mentioned.

Case 1: Directional mean difference research hypothesis

We must be careful in interpreting the probability associated with directional hypotheses because the full and restricted regression models are the same with a one-tailed test as with a two-tailed test. A non-directional research hypothesis would take the form: There is a difference in the mean effect of treatments T_1 and T_2. A directional research hypothesis would take the form: Treatment T_1 results in a larger mean effect than does treatment T_2. The full model in both cases would be:

Model 1: $Y_1 = a_oU + a_1T_1 + a_2T_2 + E_1$; the full model where:

Y_1 = the criterion vector.

U = the unit vector.

T_1 = a 1 if the Y_1 score comes from a person in treatment 1, 0 otherwise.

T_2 = a 1 if the Y_1 score comes from a person in treatment 2, 0 otherwise.

a_o, a_1, a_2 are weighting coefficients which will produce the smallest sum of squared components in the E_1 vector.

E_1 is the error in prediction, or $(Y_1 - \tilde{Y}_1)$, using the weighting coefficients and the predictor variables in the full model.

For each of the above research hypotheses, the statistical hypothesis is: There is no difference in the (population) treatment means. The statistical hypothesis inplies the restriction: $a_1 = a_2$. Forcing this restriction on the full model, we arrive at:

Model 2: $Y_1 = a_o U + E_2$; the restricted model.

All symbols are as defined before, with E_2 being the error in prediction using the weighting coefficients and predictor variables in the restricted model.

The two models can of course be compared with the F test, and the associated probability value will be reported by most canned programs. The probability value is the probability of this large a discrepancy or one larger occuring under the restriction that the two population means are equal. The first two rows in Table 1 indicate the state of affairs when the research hypothesis is non-directional. The reported probability value is for a non-directional test of significance and thus no correction is necessary.

If we are concerned about differences in a given direction, then we must look at the sample means to see if the difference between the means is in the hypothesized direction. If the means are in the direction hypothesized, the third example in Table 1, then we must halve the reported probability level, for it indicates to the researcher how often he would expect this large a discrepancy in both directions. If the means are not in the hypothesized direction (the last example in Table 1), then we surely do not want to hold as tenable the research hypothesis. The correct probability level in this case is $(1 - \frac{PROB}{2})$, where PROB is the reported probability value. Since PROB can never be larger than 1, the smallest actual probability level can never be less than .50, i.e. can never lead to holding as tenable the research hypothesis.

Pedagogically, one might want to illustrate the F distribution as in Figure 1. The top half of the F distribution can be thought of as the F ratios resulting when Treatment 2 has a higher mean than Treatment 1. The bottom half then represents those F ratios resulting when Treatment 1 has a higher mean than Treatment 2. It should be quite clear from Figure 1 that if one's alpha level is .05 the appropriate lower limit for a non-directional test is F = 4.20, whereas if the research hypothesis involves directionality, then F = 2.89 is the appropriate lower limit (this being the tabled F value for alpha = 2 x .05, or for an alpha of .10; degrees of freedom equal 1 and 28).

Case 2: Directional correlational research hypothesis

The argument for this case is similar to the previous argument, the only difference is that here we have a continuous predictor variable rather than a dichotomous predictor variable. Often in correlational research, the research hypothesis is something like: There is a non-zero relationship between X_1 and Y_2. The statistical hypothesis in this case would be: There is a zero relationship between X_1 and Y_2. The full and restricted models would be:

Model 3: $Y_2 = a_0U + a_1X_1 + E_3$; the full model where:

Y_2 = the criterion vector.

U = the unit vector.

X_1 = the continuous predictor vector.

a_0 and a_1 are weighting coefficients which will provide the sum of squared component in the E_3 vector.

E_3 is the error in prediction ($Y_2 - \widetilde{Y_2}$) using the weighting coefficients and predictor variables in the full model.

The restriction: $a_1 = 0$ results in

Model 4: $Y_2 = a_0U + E_4$; the restricted model where all symbols are as above, and where E_4 is the error in prediction $(Y_2 - \tilde{Y}_2)$ using only the overall mean (a_0).

One's research hypothesis might involve a directional relationship such as: There is a positive correlation between X_1 and Y_2. The full and restricted models would be the same, but again one would have to inspect the sign of the weighting coefficient to make sure the non-zero correlation is in the hypothesized direction. The same kinds of corrections in the probability level are called for in this case as in the case for directional differences, and examples are depicted in Table 2. Indeed, we would expect this to be the case because the test for the difference between two means is algebraically equivalent to the test of significance for the point biserial correlation, a special case of the Pearson Product Moment Correlation (Kelly, Beggs, McNeil, Eichelberger, and Lyon, 1969).

Case 3: Directional interaction research hypothesis

This third case has probably not been utilized in the literature because it has not been described in the standard statistical texts. We are not aware of any applied examples of this case, although many research hypotheses in the literature actually call for such an analysis. When a two-tailed interaction analysis is run on a directional interaction hypothesis rather than the legitimate one-tailed analysis, the researcher is reporting a probability level which is not indicative of the actual pobability. As in the previous cases, if the results are in the hypothesized direction the actual probability value is less than that which the researcher reports. We are not aware of the actual correction, as will be indicated shortly.

An example from the literature may help clarify the problem. Gentile (1968) hypothesized: "the lower the sociocultural level of the student, the more he should benefit from the definition treatment (as compared to the no-definition treatment)." Figure 2 illustrates the kind of interaction indicated by the research hypothesis. Figure 3 illustrates the other half of the situations wherein an interaction can occur. These kinds of interaction in Figure 3 are evidently not of interest to Gentile. Therefore, the reported probability level should be at least halved if the results are in the hypothesized direction.

We say at least halved because there are other kinds of interactions similar to that depicted in Figure 1 which would not reflect the research hypothesis. Figure 4 contains one such situation wherein the definition treatment is inferior to the no-definition treatment. Again one would not want to hold as tenable the research hypothesis with this set of data.

As in the first two cases, the full and restricted models for the directional and non-dirctional interaction questions are exactly the same (See Table 3). The sociocultural levels can be treated as categorical variables or as continuous, and we prefer the latter. (The discussion would become more involved if we didn't do it this way.)

The full model which allows interaction to occur would be:

Model 5: $Y_3 = a_0U + a_1T_1 + a_2T_2 + b_1X_1 + b_2X_2 + E_5$

Where:

Y_3 = the criterion vector.

U = the unit vector.

T_1 = 1 if the subject received the definition treatment, otherwise 0.

T_2 = 1 if the subject received the no-definition treatment, 0 otherwise.

X_1 = sociocultral level of the subject if he received the definition treatment, 0 otherwise.

X_2 = sociocultural level of the subject if he received the no-definition treatment, 0 otherwise.

a_0, a_1, a_2, b_1, b_2 are weighting coefficients which will produce the smallest sum of squared components in the E_5 vector.

E_5 = the error in prediction, $(Y_3-\tilde{Y}_3)$, using the wighting coefficients and predictor variables in the fuul model.

In this example b_1 and b_2 are the slopes of the straight lines of best fit for the two treatments. The hypothesis of no interaction in the population stipulates that the population slopes are equal $(B_1 = B_2)$. Since the sample slopes are the best estimators of the population slopes, the restriction which does not allow interaction to occur is: $b_1=b_2$. This restriction placed on the full model results in the following restricted model:

Model 6: $Y_3 = a_0U + a_1T_1 + a_2T_2 + b_3X_3 + E_6$

All symbols are as defined above, and where X_3 is the sociocultural level of the subject, no matter which treatment he received. E_6 is the error in prediction, $(Y_3 - \tilde{Y}_3)$, using the weighting coefficients and predictor variables in the restricted model. Again, the full and restricted models can be compared via the generalized F ratio.

If one has a non-directional interaction question and the F is significant then the results can simply be plotted and the reported probability level reported.

If one has a directional interaction question and the F is significant, then the results must be plotted to see if the interaction occurs in the direction hypothesized. If the results are opposite to that hypothesized, we surely would not want to hold as tenable the research hypothesis. If the interaction is in the direction hypothesized, then the exact probability is at least one-half the reported probability.

We feel that the above adjustment is not an exact adjustment, but at this time we are not able to describe the exact probability. We would encourage researchers to consider this question and in the future try to develop the exact probability. Certainly though, the interaction plot must reflect the research hypothesis before the researcher can reject the statistical hypothesis and hold as tenable the research hypothesis.

What we question is the probability statement associated with the interaction test of statistical significance when the researcher has stated a directional interaction research question. The reader should be reminded that the statistical hypothesis when testing either interaction or directional interaction is: There is no interaction, or, the lines are parallel. There are many ways of obtaining interaction and only a small subset of these is of interest to the researcher who is interested in a directional interaction question.

These thoughts seem to be important because many decisions are based on statistical grounds whech are being used incorrectly. Many research hypotheses involve a directional hypothesis. The researcher is hurting himself when he uses a two-tailed test rather than a one-tailed test. If his results are in the hypothesized direction, the statistic may not fall in the critical region of the two-tailed test, whereas it might have fallen in the critical region of the one-tailed test. (Please remember to also report the amount of variance being accounted for in either case, as that index will probably communicate more than will the probability value.) What is even more disheartening is to see a researcher develop a beautiful directional hypothesis and then report that his data indicate significance in the "opposite direction." He has used a two-tailed test of significance for the directional hypothesis and has found that the statistic falls in the critical region. A little thought would indicate that the researcher cannot hold as tenable his directional hypothesis under these conditions.

He should report what he found and urge future researchers to develop directional hypotheses to correspond with his data; it is ironical to report something as being significant which was completely opposite to that which was expected. In essence, the rationale behind the directional hypothesis may be incorrect, but that cannot be determined on the initial data.

Table 1

Several hypothetical examples for the differences between two groups
(Full model takes the form of Model 1 and restricted model the form of Model 2.)

Research Hypothesis	Sample Index	Statistical Hypothesis	Restriction	Sample (a_0+a_1)	Means (a_0+a_2)	Outputted Probability	Correction Needed	Actual Probability
$\mu_1 \neq \mu_2$	$a_1 \neq a_2$	$\mu_1 = \mu_2$	$a_1 = a_2$	20	15	.07	no correction	.07
$\mu_1 \neq \mu_2$	$a_1 \neq a_2$	$\mu_1 = \mu_2$	$a_1 = a_2$	15	20	.07	no correction	.07
$\mu_1 > \mu_2$	$a_1 > a_2$	$\mu_1 = \mu_2$	$a_1 = a_2$	20	15	.07	$\frac{PROB}{2}$	.035
$\mu_1 > \mu_2$	$a_1 > a_2$	$\mu_1 = \mu_2$	$a_1 = a_2$	15	20	.07	$1 - \frac{PROB}{2}$	.965

Table 2

Several hypothetical examples for correlational hypotheses
(Full Model takes the form of Model 3 and restricted Model the form of Model 4)

Research Hypothesis	Sample Index	Statistical Hypothesis	Restriction	Sample Correlation	Outputted Probability	Correction Needed	Actual Probability
$\rho \neq 0$	$a_1 \neq 0$	rho $\rho = 0$	$a_1 = 0$	.36	.07	no correction	.07
$\rho \neq 0$	$a_1 \neq 0$	$\rho = 0$	$a_1 = 0$	- .36	.07	no correction	.07
$\rho > 0$	$a_1 > 0$	$\rho = 0$	$a_1 = 0$	.36	.07	$\frac{PROB}{2}$	.035
$\rho > 0$	$a_1 > 0$	$\rho = 0$	$a_1 = 0$	-.36	.07	$1 - \frac{PROB}{2}$	.965
$\rho < 0$	$a_1 < 0$	$\rho = 0$	$a_1 = 0$	.36	.07	$1 - \frac{PROB}{2}$	.965
$\rho < 0$	$a_1 < 0$	$\rho = 0$	$a_1 = 0$	-.36	.07	$\frac{PROB}{2}$	.035

Table 3

Several hypothetical examples
for interaction hypotheses
(Full model takes the form of Model 5 and
restricted model the form of Model 6.)

Research Hypothesis	Sample Index	Statistical Hypothesis	Restriction	Sample Values b_1	Sample Values b_2	Outputted Probability	Correction Needed	Actual Probability
$\beta_1 \neq \beta_2$	$b_1 \neq b_2$	$\beta_1 = \beta_2$	$b_1 = b_2$	.4	.2	.07	no correction	.07
$\beta_1 \neq \beta_2$	$b_1 \neq b_2$	$\beta_1 = \beta_2$	$b_1 = b_2$	.2	.4	.07	no correction	.07
$\beta_1 > \beta_2$	$b_1 > b_2$	$\beta_1 = \beta_2$	$b_1 = b_2$	.4	.2	.07	$\frac{PROB}{2}$	.035
$\beta_1 > \beta_2$	$b_1 > b_2$	$\beta_1 = \beta_2$	$b_1 = b_2$	.2	.4	.07	$1 - \frac{PROB}{2}$	.965

Conservative estimate

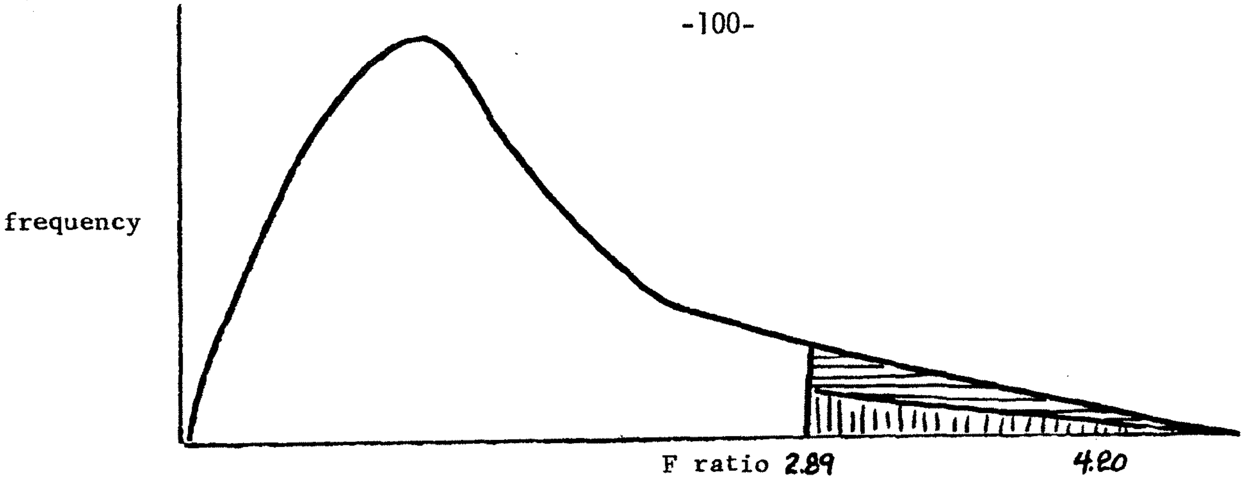

Figure 1

Exemplary F destribution ($df_1 = 1$, $df_2 = 28$) indicating F ratios resulting under the statistical hypothesis of equal population means. The area depicted by vertical lines represents those F ratios resulting when, say, Treatment 1 has a higher sample mean than Treatment 2. The area depicted by the horizontal lines represents those F ratios resulting when research hypothesis is directional, then the researcher must use the tabled F value for (2 x alpha). This process is analogous to adjusting the reported probability values as indicated in Tables 1, 2, and 3.

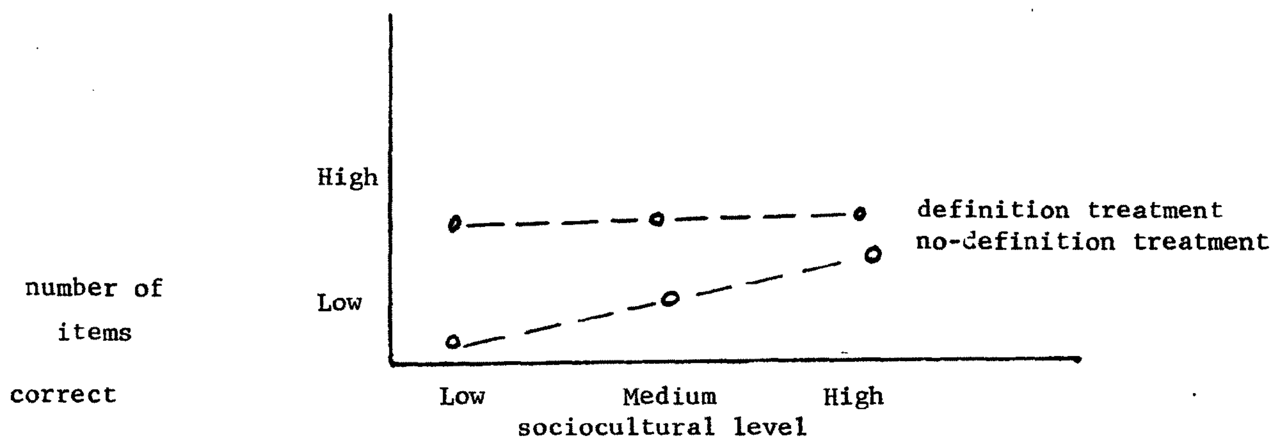

Figure 2

Schematic diagram representing directional interaction hypothesis of Gentile (1968).

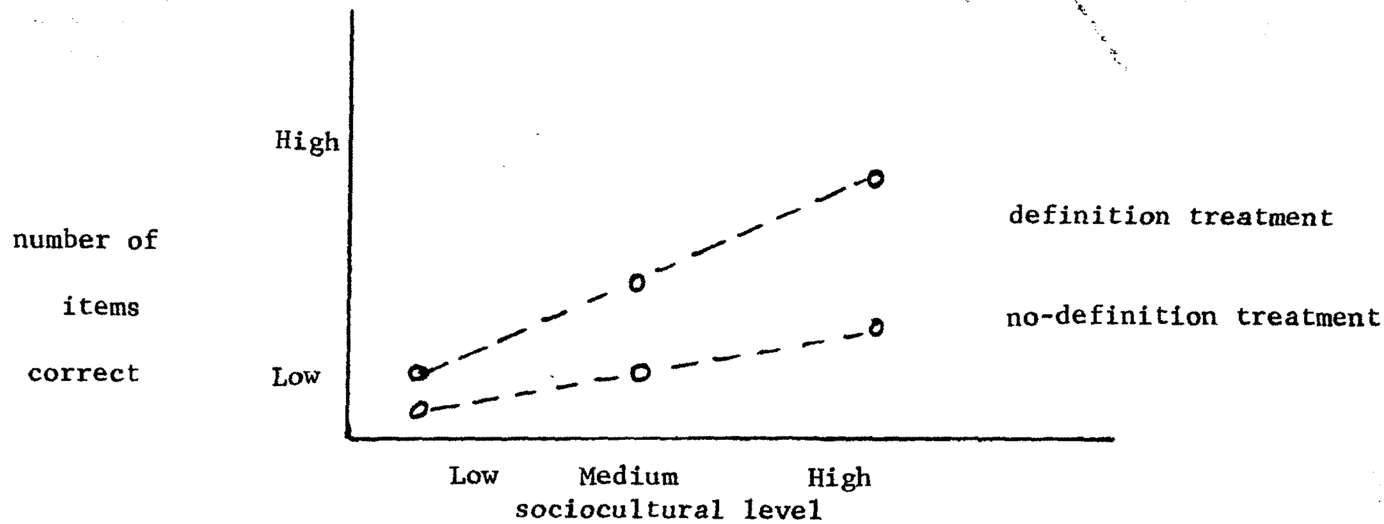

Figure 3

Schematic diagram representing other interactions which could occur but were of no interest to Gentile (1968).

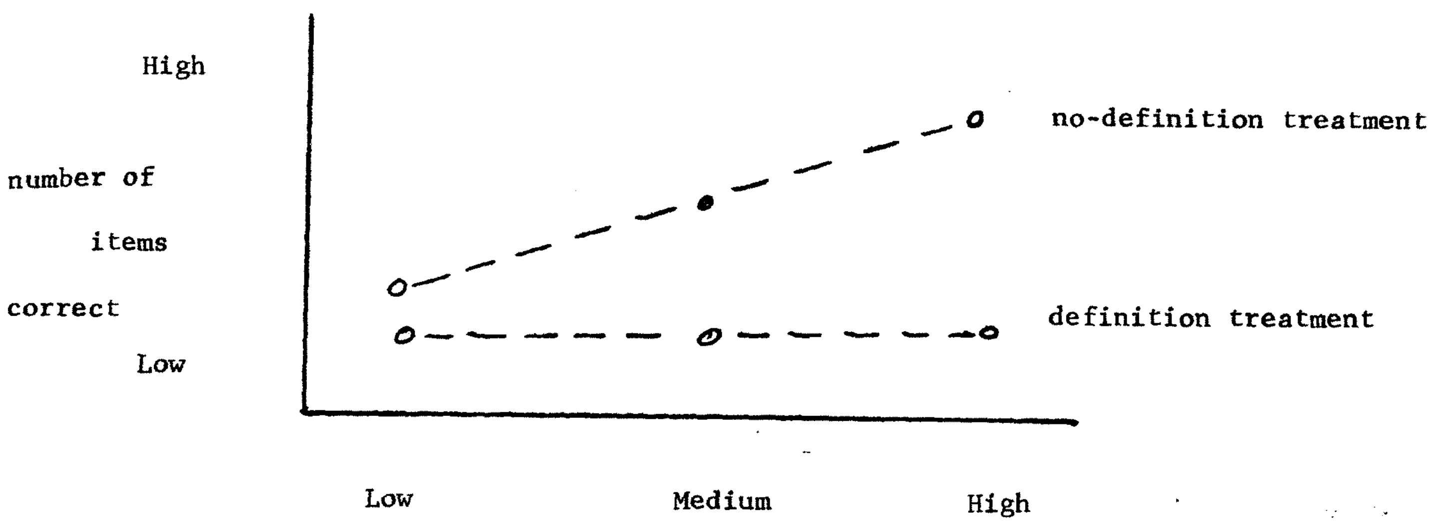

Figure 4

Schematic diagram representing lines similar to Figure 2 but with the definition treatment consistently inferior to the no-definition treatment.

REFERENCES

Gentile, J.R., Sociocultural Level and Knowledge of Definitions in the Solution of Analogy Items, American Educational Research Journal, Vol.5,4, 626-638,1968.

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

Levinger, G. and Schneider, D. J., Test of the "Risk Is a Value" Hypothesis, Journal of Personality and Social Psychology, Vol.II, 2, 165-169, 1969.