Developmental Psychology
1973, Vol. 8, No. 2, 172–179

Interpreting Regression toward the Mean in Developmental Research

LITA FURBY[1]
Yale University

The fundamental nature of regression toward the mean is frequently misunderstood by developmental researchers. While errors of measurement are commonly assumed to be the sole source of regression effects, the latter also are obtained with errorless measures. The conditions under which regression phenomena can appear are first clearly defined. Next, an explanation of regression effects is presented which applies both when variables contain errors of measurement and when they are errorless. The analysis focuses on cause and effect relationships of psychologically meaningful variables. Finally, the implications for interpreting regression effects in developmental research are illustrated with several empirical examples.

Regression toward the mean is ubiquitous in developmental psychological research. However, as Campbell and Erlebacher (1970) have pointed out, it is "often referred to but less often understood [p. 192]." The purpose of this article is to help eliminate the ignorance on this topic, particularly for the great majority of developmental researchers who are not full-time methodologists, but who nevertheless repeatedly encounter regression effects in their own data and for whom sound interpretation of results demands an understanding of this paradoxical phenomenon. The importance of regression effects has recently been highlighted in discussions of several particularly controversial issues in developmental psychology: social class and racial differences in intelligence (Eysenck, 1971; Jensen, 1968); changes in traits or abilities over time (Baltes, Nesselroade, Schaie, & Labouvie, 1972); and evaluations of compensatory education based on comparisons of pretest and posttest measures (Campbell & Erlebacher, 1970).

The existing literature on regression effects is valid, but it is inadequate for the needs of developmental psychology. Regression toward the mean is most often treated indirectly or tangentially in discussions of measuring change over time and/or covariance analyses (Campbell & Erlebacher, 1970; Lord, 1956, 1958, 1960; McNemar, 1958; O'Connor, 1970; Porter, 1967; Thorndike, 1942). The few fairly straightforward discussions of regression effects (Campbell & Stanley, 1963; Lord, 1963; Rulon, 1941) have been limited to descriptions of regression, accompanied by warnings about how such effects may lead to fallacious data interpretations. Any explanatory analyses of the causes of regression have dealt solely with regression due to error of measurement. The more fundamental nature of regression (obtained even with errorless measures) remains a mystery to most developmental psychologists (and others as well). An explanation of the nature of regression phenomena is lacking—an explanation in terms of cause and effect relationships which make psychological sense.

For didactic purposes, the present analysis attempts to keep mathematical discussion to a minimum. However, letters of the alphabet are often used to represent psychological variables in order to efficiently specify the statistical assumptions and to indicate the generality of the phenomena under study.

[1] Susan Harter made significant contributions both to theoretical arguments and to clarity of presentation. Requests for reprints should be sent to the author at the Department of Psychology, Yale University, 333 Cedar Street, New Haven, Connecticut 06510.

Definition

Consider a group of subjects for whom we have information on two variables, x and y. For illustrative purposes, let us operationalize these variables as the amount of aggression observed in nursery school children in a standard experimental situation. Let x be amount of aggression displayed by a subject at Time 1 (measured by behavior observation techniques), and let y be the amount of aggression displayed by the same subject at a later Time 2. Let $\bar{X}$ represent the mean level of aggression for the entire group at Time 1, and let $\bar{Y}$ represent the mean level of aggression of the group at Time 2. Suppose that (*a*) we have perfectly reliable (errorless) measures of aggression (e.g., two observers agree perfectly when observing the same subject on the same occasion) and (*b*) the correlation between x and y is less than 1.0; for simplicity of presentation, suppose further that (*c*) the distributions of the two variables are normal and (*d*) the standard deviations of the two variables are equal (i.e., $s_x = s_y$). This situation is represented graphically in Figure 1, where each subject may be located according to his x and y scores. The ellipse encloses all the cases in the scatter plot (i.e., all of the subjects), the solid line represents the best-fitting line for y regressed on x (whose slope is r_{xy}, the correlation between x and y), and the broken line represents $x = y$ (i.e., the regression line if there were a perfect correlation between x and y).

Now let us choose a certain aggression level at Time 1 (any level other than the mean) and call it x'. Next take all those subjects whose x score is x', and calculate their mean score on y which is called y' (i.e., for all those subjects who displayed x' amount of aggression at Time 1 we calculate their mean amount of aggression displayed at Time 2). We find that the mean level of aggression at Time 2 for these subjects is actually closer to the overall group mean aggression level at Time 2 than their level of aggression at Time 1 was to the overall group mean at Time 1. This is represented in Figure 1 by the fact that x' is farther from $\bar{X}$ than y' is from $\bar{Y}$. If there had been no regression effect, the mean y score for these subjects would have been y^*, which is just as far from $\bar{Y}$ as x' is from $\bar{X}$.

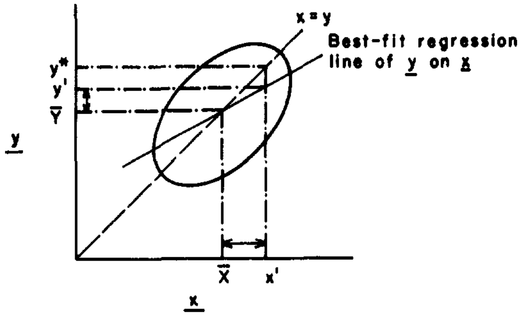

FIG. 1. Graphic representation of regression toward the mean.

This phenomenon has come to be known as "regression toward the mean" because x and y often represent the same operational variable (as in our example where both are measures of observed aggression) but are measured on different occasions. Subjects with a given aggression level x' appear to "regress toward the mean" when they are measured some time later on the same variable. Another way to view this graphically is presented in Figure 2, which contains exactly the same scatter plot as Figure 1. All of the subjects whose initial aggression level was x' can be found somewhere along the heavy solid line, since all subjects are contained within the ellipse. These subjects' y scores can all be found within the projection of the heavy solid line on the y axis, and their mean y score is y'; y' is closer to $\bar{Y}$ then x' is to $\bar{X}$.

Although we have assumed equal standard deviations for simplicity of presentation, the regression effect still obtains when $s_x \neq s_y$, and the more general case may be expressed

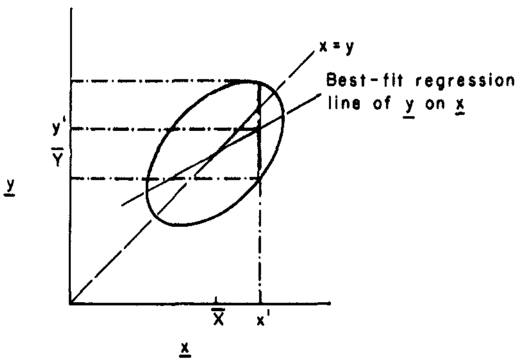

FIG. 2. Range of y scores for all subjects whose x score is x'.

as follows: for a given score on x (e.g., x'), the corresponding mean score on y (e.g., y') is closer to $\bar{Y}$ in standard deviation units than x' is to $\bar{X}$ in standard deviation units.

Since our measurement of aggression was errorless (by assumption), the observed regression effect must be due to real changes in aggression: Subjects who were more aggressive than average at Time 1 actually tended to get less aggressive, and those less aggressive than average at Time 1 tended to get more aggressive by Time 2. If Time 1 and Time 2 were only milliseconds apart, then the correlation between x and y would be perfect and we would not observe the regression effect. On the other hand, if our measuring instruments were not perfectly reliable (as is most often the case in psychology), then even if the two measures were only milliseconds apart we would obtain regression effects.

In sum, a less than perfect correlation between x and y leads to regression effects, and such a less than perfect correlation can be due either to unreliability in x and y or to real differential change in subjects, or to both. In most developmental studies, we have both unreliable measuring instruments and two variables which are separated by a significant period of time and which are not perfectly correlated with each other. We can thus expect regression toward the mean both from measurement error and from real changes in our subjects. But what do these regression effects mean? Why do they exist?

Sources and Meaning of Regression toward the Mean

Consider first the case of perfectly reliable measures. The farther a score is from the mean, the more extreme it is. The more extreme a score, the rarer it is and the more likely it is to have been the result of a very rare combination of factors. If I now compare an extreme score on variable x, for example, with a score for the same person on another variable y, it is highly unlikely that this person will also have the necessary rare combination of factors determining y so that he will have an extreme score on y as well. It is unlikely for any person to have an extreme score on y, including a person who has an extreme score on x.

If there were no significant correlation between x and y, we would expect regression all the way to the mean. That is, for all those individuals with a given extreme x score, their mean y score would be equal to the overall group mean $\bar{Y}$. However, psychological variables are often moderately (although not perfectly) correlated with each other, and thus instead of having regression *to* the mean, we have regression *toward* the mean. It makes intuitive sense that some of the variables that determine x also determine y (this is usually what a nonzero correlation means). When x is a very extreme score, then all of the factors determining it must have been extreme (and rare). Some of those factors probably also determine y (if x and y are correlated), and thus y tends to be in the same direction from the mean $\bar{Y}$ as the x score is from its mean $\bar{X}$. However, there are also some other factors that determine y which have nothing to do with x. For each of those factors, extreme values are very rare, and it is therefore unlikely that all of the factors determining y are extreme and produce a score as extreme as the x score (for those subjects with extreme x scores). For explanatory purposes I speak of "extreme" scores and "very rare" determining factors. But this is all a matter of degree, and the same argument holds for any scores that are different from the mean—but the magnitude of the regression effect varies as a function of the distance of the x score from its mean.

Example

It may be useful to consider the reasoning of the preceding paragraphs in terms of a concrete example. Suppose that x equals the observed level of exploratory behavior in 3-year-old children in a free-play situation, and that y equals the amount of interpersonal aggression in the same 3-year-olds as measured by behavior observations in nursery school. If exploratory behavior and interpersonal aggression are correlated, it makes intuitive sense that some of the factors that determine exploratory behavior also determine interpersonal aggression. In this case, it may be that activity level has a determining influence on both exploratory behavior and aggressive behavior. Now, if a given child's exploratory behavior is extremely high, then all of the factors determining exploratory behavior must be extremely high, including activity level. Since aggressive behavior also is determined partially by activity level, it also tends to be high for this particular child. However, there are also some other factors that determine aggressive behavior and that have nothing to do with determining exploratory behavior. For illustrative purposes, suppose that bicep strength and frustration level (defined as the degree to which situational characteristics block goal-directed behavior) are two factors which determine aggressive behavior, but are unrelated to exploratory behavior. Extremely high values on any factors, including bicep strength and frustration level, are very rare. It is therefore unlikely that for any given individual (including those with high exploratory scores) all three of the factors determining aggression will be extremely high. For a child whose exploratory behavior is very high, activity level must be high. But it is unlikely that both bicep strength and frustration level are equally high. Thus it is unlikely that a child with an extremely high exploratory score will have an equally extreme aggression score.

Errors of Measurement

When we have unreliable measures, exactly the same reasoning applies to "errors of measurement" as applies to "factors determining" x and y. Errors of measurement contribute to the less than perfect correlation between x and y. This is because the error in a subject's x score is uncorrelated with either the absolute value of his x score or with the error in his y score. If this were not the case, the measuring instrument would be biased, and in this discussion we are dealing only with random, unbiased errors. Those subjects with large positive error contributing to their scores on x are likely to have higher x scores on the average than those subjects with negligible or large negative error contributing to their scores. However, it is highly unlikely that these subjects with large positive error in their x scores will also have large positive error in their y scores (since error in x is uncorrelated with error in y). Therefore, their y scores tend to be lower (closer to the mean) than their x scores. (The analogous but opposite statement is true for large negative error in x.)

The major portion of this article focuses on the case of errorless measures since it is the more interesting and less understood. In fact, "the more widely accepted interpretation" of regression effects deals only with regression caused by measurement error (Baltes et al., 1972). A straightforward explanation (in terms of psychological cause and effect relationships) of why we also observe regression for errorless measures cannot be found.[2] Error of measurement can be considered as just a specific case of the more general "factors determining a score," which is the major concern of this discussion.

Relation between x *and* y

While regression toward the mean is often discussed in the context of developmental changes in the same variable over time, it is actually a more general phenomenon. It is possible for x and y to be related to one another in a variety of ways as long as they both can be referred to each of the individuals under study (on the same numerical scale). In the most common case, x and y are the same operational measure administered to the same person on different occasions (as in the example of aggression at Time 1 and Time 2). On the other hand, x and y may also be quite different measures of the same person (as in the example where x equals exploratory behavior and y equals aggressive behavior). Sometimes x and/or y are quite indirectly related to the individual himself, as in the often misinterpreted comparisons of children's IQs (y) with IQs of their parents (x). No matter what the two variables compared, the correlation between them must be less than 1.0 in order to obtain a regression effect, and the smaller the r_{xy}, the greater the regression toward the mean.

[2] The common and unique factors determining x and y have been related to regression phenomena in a less direct and more mathematically sophisticated way by Campbell and Erlebacher (1970), Lord (1960), and Porter (1967) in their discussions of covariance analysis, and by Thorndike (1942) in his discussion of the matched groups experiment.

Interchangeability of x and y

An important but seldom recognized characteristic of regression toward the mean is the fact that x and y can be interchanged and the regression effect is still obtained. In other words, for a given score on y (call it y'') the corresponding mean x score (call it x'') is closer to $\bar{X}$ than y'' is to $\bar{Y}$. This phenomenon is represented in Figure 3 which contains exactly the same scatter plot as Figures 1 and 2, except that the light solid line here represents the best-fitting line for x regressed on y. The heavy solid line contains all of those subjects with a score of y''; it can be seen that their mean x score is x''. While this phenomenon may seem paradoxical, it makes intuitive sense if we refer to the preceding explanation of the meaning and sources of regression, just reversing x and y: The more extreme a person's score on y, the more unlikely that he will also have the necessary rare combination of factors determining x so that he will have an equally extreme score on x. When x and y represent the same measure obtained on two different occasions, regressing x (Time 1) on y (Time 2) constitutes "time-reversed analysis," suggested originally by McNemar (1940), and more recently by Campbell and Stanley (1963). This procedure has rarely been considered for analyzing developmental data despite the extreme usefulness of such a technique for separating regression effects from other sources of longitudinal changes (see Baltes et al., 1972).

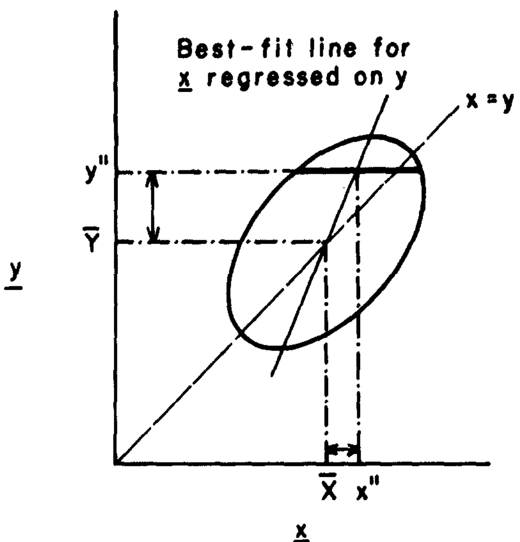

FIG. 3. Regression toward the mean when x and y are interchanged.

Implications for Interpreting Regression toward the Mean

The purpose of this article has been to explain the fundamental nature and sources of regression toward the mean. The ultimate goal is that developmental psychologists understand regression effects well enough so that they will not make erroneous interpretations of such effects in their empirical data. This concluding section is limited to several very salient examples of practical application of the principles exposed above. This is not and could not be an exhaustive review of all the possible implications for interpreting regression effects in development analyses.

Causal Determinants of Variables

Regression effects have recently been interpreted as supporting evidence for genetic sources of individual differences in the variables under consideration. A case in point is the obtained regression toward the mean when children's IQs (y) are compared with the IQs of their parents (x). The existence of a regression effect has been interpreted as supporting the importance of hereditary determinants of IQ (Eysenck, 1971; Jensen, 1968). However, such an interpretation ignores the fact that environmental determinants may just as well lead to regression effects as genetic ones.

For the case of intelligence, it makes intuitive sense that a parent who has an extreme value on every environmental factor determining intelligence will pass on some of those extremes to his child. For example, the parent is likely to imitate the child-rearing practices of his own parents, live in a similar neighborhood, etc. Assuming for the moment that the environment plays a significant role in determining individual differences in IQ, this would account for the fact that if a parent's IQ is above the mean, his child's IQ also is likely to be above the mean. However, for a parent who had extremes on many environmental factors, it is unlikely that his child will have as many extremes since some of them depend partly on chance.

Thus, if IQ differences were significantly determined by environmental factors we would still expect a regression effect. Regression toward the mean may just as well indicate traits or variables that have environmental causes as those that have genetic determinants.

In a recent superb discussion of the race-intelligence controversy, Scarr-Salapatek (1971) suggested that differential regression effects for blacks and for whites would be expected if their population mean genotypes for intelligence actually differed. More specifically, the children of high-IQ white parents should regress less than children of equally high-IQ black parents if the population mean genotype were lower for blacks than for whites. While this analysis is perfectly sound, one may not go further and claim that differential regression effects (if found) would be evidence for genetic differences in intelligence between races. Differential regression also could result from different mean environments between the races. As indicated above, even if IQ were determined solely by environmental factors we would expect regression toward the mean. And if the mean environment for the black population were inferior (in IQ-determining aspects) to that for the white population, then for a given high parental IQ score, we would expect more regression by a black child than by a white child (see Figure 4 for a graphic representation of this phenomenon). Conversely, for a given low parental IQ score, we would expect more regression from the white child than from the black.

On the other hand, smaller than expected differential regression effects (i.e., the means toward which the two groups regress differ by substantially less than the observed mean IQ difference between blacks and whites of 15 IQ points) would be evidence for equal population mean genotypes, different popu-

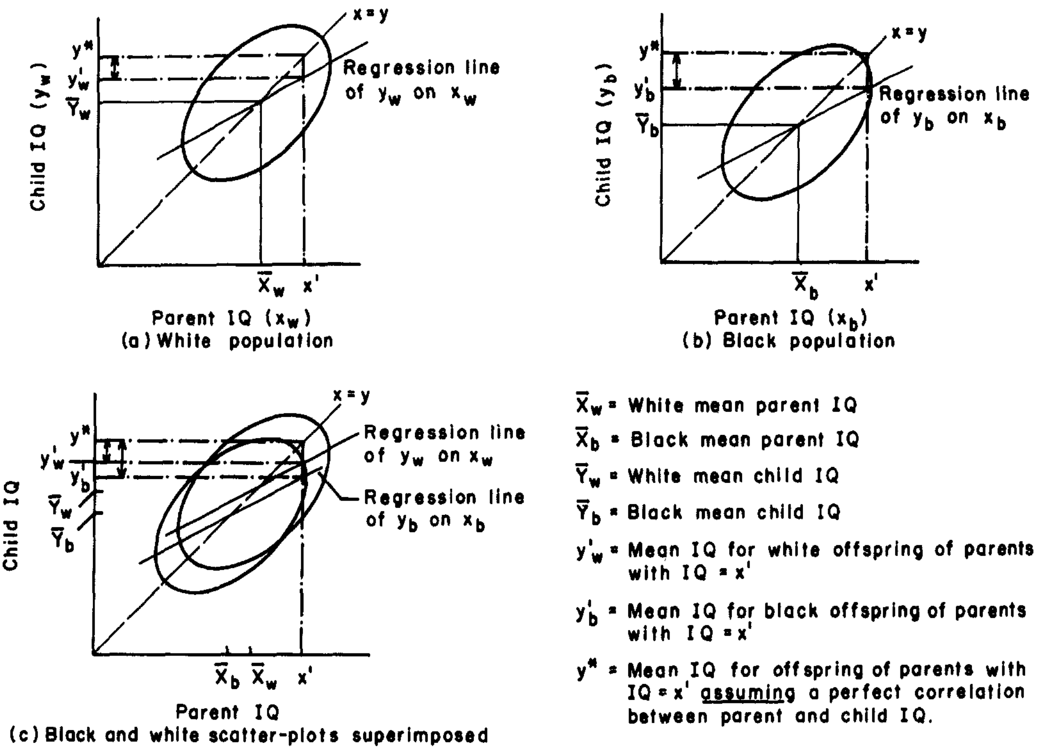

FIG. 4. Differential regression toward the mean for hypothetical IQ distributions of blacks and of whites (The only difference between a and b is that the population mean IQ is assumed to be lower in b than in a. y_b' has regressed more than y_w', but percentagewise the two have regressed the same amount, that is, halfway to their respective population means; it is the absolute amount of regression that is greater in b than in a.)

lation mean environments, and a significant role of both heredity and environment in determining within-group IQ differences. The latter would also suggest that environmental differences account for the observed racial differences in IQ. But it would still be conceiveable that the mean population genotypes differ between the races but by less than the observed difference in phenotypes of 15 IQ points.

Correlates of Change or Difference Scores

One must be extremely cautious in interpreting which subjects have changed how (i.e., which subjects have regressed in what direction). For example, since IQ is correlated with SES, the regression effect discussed above (child IQ regressed on parental IQ) can be stated in different terms. We have seen that children of high-IQ parents tend to be less intelligent than their parents; but, alternatively, children of high-SES parents tend to be less intelligent than their parents. While this statement is perfectly correct, some authors (notably, Eysenck, 1971) have gone further and claimed that this regression effect reflects a negative relation between IQ and SES-determined environmental factors.

Eysenck's interpretation is totally unjustified. The regression of child's IQ when compared to parental IQ obtains regardless of SES, but Eysenck is interpreting SES as a causal factor where we have no evidence on causation. In fact, I can easily turn the argument around by simply reversing the comparison and regressing parental IQ on child IQ (i.e., by interchanging x and y as discussed above). Then, given a certain child IQ (call it x'''), the mean IQ for parents of all those children with an IQ of x''' is halfway closer to the mean of the whole population of parents than is x''' to the mean IQ of the entire population of children. Thus, the most intelligent children come from parents who are less intelligent than their children but still above the mean, while the least intelligent children come from parents who are more intelligent than their children but below the population mean. Using Eysenck's logic for interpreting regression effects, we could conclude the following: Since the most intelligent children come from smarter than average parents, and since smarter than average parents also are higher than average SES, then we could claim that high-SES parents supply their children with an environment that makes the children even smarter. Similarly, the least intelligent children come from lower than average IQ parents who are also lower than average SES. Thus, we could claim that low-SES parents supply their children with an environment that makes the children even less intelligent. It then would appear that we have proved a positive relation between IQ score and environmental factors determined by SES.

Obviously the above interpretation is erroneous, as is Eysenck's. When regression toward the mean is present, the change or difference between x and y is necessarily correlated with the initial x score: The farther a given x score is from the mean $\bar{X}$ the greater the amount of regression (in absolute value). The fact that another variable such as SES is also correlated with the difference between x and y may be due only to the fact that the variable in question (e.g., SES) is correlated with initial x value (e.g., parents' IQ). This poses a classic problem in change and difference scores: Should, and how should, we measure differences while holding initial level constant, and what are the correlates of change? Solutions to this problem have been discussed elsewhere (Campbell & Erlebacher, 1970; Cronbach & Furby, 1970; DuBois, 1957; Harris, 1963; Lord, 1956, 1958, 1960; McNemar, 1958; O'Connor, 1970; Porter, 1967; Tucker, Damarin, & Messick, 1966; Werts & Linn, 1970), and the reader would do well to consult these references for possible answers to "what to do with the data" questions. In contrast, the main purpose of the present discussion is to explain why regression toward the mean exists and why it poses difficulties for data interpretation. It is sufficient here that the reader understand (*a*) the sources of regression as discussed above, (*b*) that x-y differences (i.e., amount of regression toward the mean) are related to initial x level (and that when we interchange the axes the differences are then related to y level), and (*c*) that many variables are related to the x-y difference only because they are related to initial x score.

In 1941 Rulon could state that "the list of studies in which the regression factor has been neglected grows monotonous, as well as distressing [p. 222]." Thirty years later the situation is even worse. Not only is regression often neglected, but it is also seriously misinterpreted, as demonstrated in the above examples. It is hoped that this discussion will help investigators to obtain a basic understanding of the sources of regression, and thereby to neither neglect nor misinterpret such a pervasive phenomenon.

REFERENCES

BALTES, P. B., NESSELROADE, J. R., SCHAIE, K. W., & LABOUVIE, E. W. On the dilemma of regression effects in examining ability level-related differentials in ontogenetic patterns of intelligence. *Developmental Psychology,* 1972, **6,** 78–84.

CAMPBELL, D. T., & ERLEBACHER, A. How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), *Disadvantaged child.* Vol. 3. *Compensatory education: A national debate.* New York: Brunner/Mazel, 1970.

CAMPBELL, D. T., & STANLEY, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching.* Chicago: Rand McNally, 1963.

CRONBACH, L. J., & FURBY, L. How should we measure "change"—or should we? *Psychological Bulletin,* 1970, **74,** 68–80.

DUBOIS, P. H. *Multivariate correlational analysis.* New York: Harper, 1957.

EYSENCK, H. J. *The IQ argument.* New York: Library Press, 1971.

HARRIS, C. W. *Problems in measuring change.* Madison: University of Wisconsin Press, 1963.

JENSEN, A. R. Social class, race and genetics: Implications for education. *American Educational Research Journal,* 1968, **5,** 1–42.

LORD, F. M. The measurement of growth. *Educational and Psychological Measurement,* 1956, **16,** 421–437.

LORD, F. M. Further problems in the measurement of growth. *Educational and Psychological Measurement,* 1958, **18,** 437–454.

LORD, F. M. Large-sample covariance analysis when the control variable is fallible. *Journal of the American Statistical Association,* 1960, **55,** 307–321.

LORD, F. M. Elementary models for measuring change. In C. W. Harris (Ed.), *Problems in measuring change.* Madison: University of Wisconsin Press, 1963.

MCNEMAR, Q. A critical examination of the University of Iowa studies of environmental influence upon the IQ. *Psychological Bulletin,* 1940, **37,** 63–92.

MCNEMAR, Q. On growth measurement. *Educational and Psychological Measurement,* 1958, **18,** 47–55.

O'CONNOR, E. F. *Extending classical test theory to the measurement of change.* (CSE Rep. No. 60) Los Angeles: Center for the Study of Evaluation, 1970.

PORTER, A. C. The effects of using fallible variables in the analysis of covariance. Unpublished doctoral dissertation, University of Wisconsin, 1967.

RULON, P. J. Problems of regression. *Harvard Educational Review,* 1941, **11,** 213–223.

SCARR-SALAPATEK, S. Unknowns in the IQ equation. *Science,* 1971, **174,** 1223–1228.

THORNDIKE, R. L. Regression fallacies in the matched groups experiment. *Psychometrika,* 1942, **7,** 85–102.

TUCKER, L. R., DAMARIN, F., & MESSICK, S. A base-free measure of change. *Psychometrika,* 1966, **31,** 457–473.

WERTS, J. F., & LINN, R. L. A general linear model for studying growth. *Psychological Bulletin,* 1970, **73,** 17–22.

(Received April 21, 1972)