DOCUMENT RESUME

ED 278 986 CS 008 728

AUTHOR	Wilkinson, Ian; And Others
TITLE	Silent Reading Reconsidered: Reinterpreting Reading Instruction and Its Effects. Technical Report No. 390.
INSTITUTION	Bolt, Beranek and Newman, Inc., Cambridge, Mass.; Illinois Univ., Urbana. Center for the Study of Reading.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	Oct 86
CONTRACT	400-81-0030
NOTE	21p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Elementary Education; Learning Disabilities; Learning Processes; *Models; Oral Reading; *Reading Achievement; Reading Difficulties; *Reading Instruction; *Reading Research; *Silent Reading; *Time Factors (Learning)

ABSTRACT

Results reported by Leinhardt, Zigmond, and Cooley (1981) have been interpreted as support for increased silent reading in classroom reading instruction. G. Leinhardt and colleagues examined a causal model of classroom processes influencing reading achievement and found that time spent in silent, rather than oral, reading was positively related to gains in reading achievement. To clarify the interpretation of these results, a study reanalyzed the Leinhardt data, which used students in 11 elementary classrooms for learning disabled students. Using linear structural equation modeling, the reanalysis showed that students' entry-level reading abilities had a significant direct effect on time spent in silent reading, but no such effect on time spent on oral or "indirect" reading. When entry-level abilities were more adequately controlled by incorporating measurement error into the model, silent reading no longer showed a significant effect on posttest reading performance. Indeed, under alternative models of this data, there was even the suggestion that time spent in oral reading had more effect on final reading achievement. These finding have important implications for the oral versus silent reading debate, as well as for the more general question of the relationship between time spent in reading and student achievement. (Author/FL)

CENTER FOR THE STUDY OF READING

Technical Report No. 390

SILENT READING RECONSIDERED: REINTERPRETING READING INSTRUCTION AND ITS EFFECTS

Ian Wilkinson, James L. Wardrop
Richard C. Anderson

University of Illinois at Urbana-Champaign

October 1986

University of Illinois
at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

Bolt Beranek and Newman Inc.
10 Moulton Street
Cambridge, Massachusetts 02238

The work upon which this publication is based was performed pursuant to Contract No. 400-81-0030 of the National Institute of Education. It does not, however, necessarily reflect the views of this agency. The authors are indebted to William W. Cooley and Richard G. Lomax for supplying data and other information necessary for the analyses described in this paper. A version of this paper is to appear in the American Educational Research Journal.

Abstract

Results reported by Leinhardt, Zigmond, and Cooley (1981) have been interpreted as support for increased silent reading in classroom reading instruction. Leinhardt et al. examined a causal model of classroom processes influencing reading achievement and showed that time spent in silent reading, rather than oral reading, was positively related to gains in reading achievement. The present study reanalyzed the Leinhardt et al. data in an attempt to clarify the interpretation of their results. By means of linear structural equation modeling we show that students' entry-level reading abilities had a significant direct effect on time spent in silent reading, but no such effect on time spent on oral or "indirect" reading. Any attempt at examining the role of silent reading needs to take this into account. When entry-level abilities were more adequately controlled by incorporating measurement error into the model, silent reading no longer showed a significant effect on posttest reading performance. Indeed, under alternative models of the data, there is even the suggestion that time spent in oral reading had more effect on final reading achievement. These findings have important implications for the oral versus silent reading debate, as well as for the more general question of the relationship between time spent in reading and student achievement.

Silent Reading Reconsidered: Reinterpreting Reading Instruction and its Effects

> I have steadily endeavoured to keep my mind free so as to give up any hypothesis, however much beloved ... as soon as facts are shown to be opposed to it ... for with the exception of the Coral Reefs, I cannot remember a single first-formed hypothesis which had not after a time to be given up or greatly modified. (Charles Darwin, 1888, p. 83)

Research on time spent in silent reading during classroom instruction and its effect on students' reading achievement is largely equivocal. In elementary school classrooms and classrooms for learning-disabled students, time spent in silent reading has been shown to be positively related to gains in reading achievement (Clark, 1975; Clark & Spath, 1979; Leinhardt, Zigmond, & Cooley, 1981). In contrast, research with students in special education resource rooms has shown no relationship between silent reading and gains in reading achievement (Haynes & Jenkins, 1986), and investigation in secondary remedial reading classrooms has even shown the relationship to be negative (Stallings, 1980). In Stallings' study, gains in reading achievement were found to be positively associated with time spent in oral reading.

While differences in population and associated instructional variables may provide a partial account of these findings, reviewers of this research tend to favor the results of Leinhardt et al. and interpret the general effects of silent reading on achievement to be positive. Leinhardt et al.'s study focused on classrooms for learning-disabled students and is important for at least two reasons. First, unlike some of the studies, the research is of high methodological quality. Leinhardt et al. present a causal model of classroom processes influencing reading achievement and test their model using students from a number of classes. There is a high regard for validity and reliability of measurement, a priori specification of the theoretical model, and relatively sophisticated statistical analysis. Second, their results indicate that silent reading may have a large effect on student performance and that relatively small increments in reading time may result in substantial gains in reading achievement. It is perhaps because of these reasons that reviewers frequently cite the Leinhardt et al. study when considering the available evidence in the oral versus silent reading debate (e.g., Allington, 1983, 1984; Hiebert, 1983; Reutzel, 1985), as well as the more general question of the relationship between time spent in reading and student achievement (e.g., Anderson, Mason, & Shirey, 1984; Englert, 1984; Vallecorsa, Zigmond, & Henderson, 1985).

The purpose of the present paper is to make the case that Leinhardt et al.'s data do not warrant the conclusion concerning the merits of silent reading. In interpreting their results, special consideration needs to be given to the effects of measurement error on parameter estimates in regression and to the variables used by Leinhardt et al. in specifying a model for explaining reading achievement.

The Original Study

Leinhardt et al. examined classroom processes and reading achievement in 11 elementary classrooms for learning-disabled (LD) students. Students in LD classes exhibit a wide range of abilities and instruction is usually individualized. The choice of the LD population afforded the opportunity to capitalize on this variation and so enhance the likelihood of obtaining stable parameter estimates.

The sample consisted of 105 students ranging in age from 6 to 12 years with a mean age of 8.7. The students and teachers in each classroom were observed for an average of 30 hours over a 20-week period, and pre- and posttest measures of reading performance were obtained for all students.

The variables under study were determined by the structural (causal) model for explaining reading achievement shown in Figure 1. There are two parts to this model, and each was examined separately. First, the model assumed that the posttest reading performance of a student was attributable to his or her pretest performance, the overlap between curriculum and test, and the reading behaviors in which the student engaged represented by times spent in silent, oral, and "indirect" reading. Indirect reading included those [illegible] assumed to be related to reading but in which the student was not directly engaged in resp[illegible] print (e.g., discussing a story, circling pictures with a common phonetic element, listening, [illegible] Second, the model assumed that the students' reading behaviors (total time spent in the th[illegible]ities) could be accounted for by the students' pretest performance, teacher instruction and aff[illegible]ontact, and the pacing of instruction. Affective contact included the amount of praise (reinf[illegible]) received by each student and the academic focus (cognitive press) exerted by the teacher towar[illegible] student. It is important to note that in the first part of the model three separate reading variables were used, whereas in the second, a combined reading variable was used (the aggregate of time spent in silent, oral, and indirect reading).

[Insert Figure 1 about here.]

Leinhardt et al. tested the two parts of their model by multiple regression with variables being entered simultaneously into each equation. The results are shown in Table 1. The first regression indicated that posttest performance was significantly influenced by pretest, overlap, and time spent in silent reading, but not by time spent in oral or indirect reading. The second regression indicated that total time in the three behaviors was significantly influenced by all independent variables except pacing.

[Insert Table 1 about here.]

These results provide overall support for the parts of the model, and point strongly to the beneficial effects of silent reading. According to Leinhardt et al., "these results suggest that an average of one minute per day of additional silent reading time increases posttest performance by one point. An increase of five minutes per day would be equivalent to about one month (on a grade-equivalent scale) of additional reading achievement" (p. 355). Given this finding, it is easy to understand how reviewers might interpret this research as providing support for increased silent reading during classroom reading instruction.

Reanalysis

The present reanalysis addresses a key problem with the interpretation of the Leinhardt et al. results. Because their analysis assumed error-free measurement, they were unable to control fully for the differential relations between pretest performance and the reading activities in which the students engaged. Unlike oral and indirect reading, time spent in silent reading was highly correlated with pretest performance ($r = .63$), suggesting that students' initial abilities may have had a direct effect on time allocated to silent reading; the better the entry-level ability of a student, the more likely the teacher might be to assign him/her to this type of activity (cf., Allington, 1983). The extent of this direct effect was not assessed in the original study because Leinhardt et al. did not test their model in its entirety. If measurements were error-free, the attempt at partialing-out the influence of pretest performance by including this variable in the equation would have controlled for this confounding. However, to the extent that the pretest measure was fallible, it is unlikely that confounding was avoided (see Linn & Werts, 1982). Some portion of the posttest variance attributed to silent reading may have been due to the indirect effect of students' initial abilities on posttest performance.

In the present reanalysis of the data we show that (a) students' entry-level reading abilities did indeed have a significant direct effect on amount of time spent in silent reading, (b) when measurement error is taken into account and students' initial abilities more adequately controlled, the effect of silent reading on posttest performance is non-significant, though still greater than that of oral reading, and (c) under alternative models of the data, the effect of silent reading is substantially reduced, and results may even suggest that time spent in oral reading had more effect on final reading achievement.

Method

Data Analysis

Our analysis was undertaken using linear structural equation modeling implemented through LISREL VI (Joreskog & Sorbom, 1984). The input was the correlation matrix for observed variables. All except four of the correlations were obtained from original source documents supplied by William W. Cooley of the Learning Research and Development Center, University of Pittsburgh, and these were accurate to five decimal places. The four correlations relating indirect reading to teacher instruction, cognitive press, reinforcers, and pacing could not be obtained from source documents and were taken from the published article. These were accurate to two decimal places. For all LISREL analyses, the diagonal elements of the factor variance-covariance matrix were free to be estimated and disturbances were assumed to be uncorrelated. All structural coefficients reported throughout this paper are those for standardized solutions.

Models

We conducted the analysis in three stages. First, the original model was tested in its entirety. This enabled examination of the separate effects of pretest performance on the three reading behaviors (not assessed by Leinhardt et al.) as well as comparison of the LISREL results with those of the original analysis. Second, estimates of the reliability of the measures were obtained and the full model incorporating measurement error was tested. Finally, the effects of silent, oral, and indirect reading were examined under alternative models of the data.

Measures and Reliabilities

Pretest. The pretest was a composite measure consisting of scores from six subtests of the Diagnostic Reading Scales (Spache, 1972) combined with the Level I Reading Subtest of the Wide Range Achievement Test (WRAT) (Jastak, Bijou, & Jastak, 1976). Because accurate estimation of its reliability was crucial to the analyses, several methods of estimation were employed. Lomax and Cooley (1980) report a coefficient alpha estimate of reliability for four components of the composite represented by the Spache subtests (using the same sample as Leinhardt et al.). When correction was made for the inclusion of the WRAT component, this estimate yielded a reliability of .91 (Spearman-Brown formula).

The pretest measures were also administered as a posttest and so another estimate of reliability was provided by the correlation between scores on the composite at pretest and posttest. Tests administered on different occasions most likely constitute parallel or tau-equivalent forms. Although there were insufficient degrees of freedom to test the relative fit of the two models, LISREL analyses were used to examine the correlation between observed- and true-score components under the alternative models. The reliability estimates resulting from both models showed no departure from the observed correlation (.91) and, thus, agreed with the previous estimate.

A further estimate of reliability was obtained from the intercorrelations among the seven subtests making up the composite reported in Lomax (1980). At pretest, the correlation matrix reported is for $N = 120$ and at posttest, for $N = 101$ (cf., the $N = 105$ in the Leinhardt et al. study). Treating these data as variance-covariance matrices of standardized scores and assuming five test components--based on a liberal interpretation of the construction of the composite as described in Lomax and Cooley (1980)--we computed coefficient alpha estimates of reliability. For the pretest, coefficient alpha is .93 and for the posttest, .91. The discrepancy appears due to sampling differences, rather than to any inherent measurement characteristic, and the average was taken as the estimate of reliability (.92).

Posttest. In addition to repeating the pretest, the reading subtests of the Comprehensive Test of Basic Skills (CTBS, CTB/McGraw-Hill, 1974a) were used as a posttest. Level B, C, or 1 was administered to each student based on age and expected grade-level in reading. In order to obtain an overall reliability estimate, the reliabilities (KR20) for total reading scores reported in *Technical Bulletin No. 1* (CTB/McGraw-Hill, 1974b) were averaged across levels and grades within level. This yielded an estimate of .94.

Student behaviors. Classroom observation was conducted using a time sampling procedure termed the Student-level Observation of Beginning Reading (SOBR). Details of this procedure are described in Leinhardt and Seewald (1981a). The system enables the categorization of student reading behaviors into direct (silent, oral) and indirect reading behaviors at the level of letters, words, sentences, and paragraphs. The measures of reading used in this and the original analysis were the number of minutes per day a student was reading silently, reading aloud, or engaged in indirect reading.

A generalizability study of the SOBR (Lomax, 1982) revealed a high level of stability and interobserver agreement for the instrument (coefficients of .90). However, because most of the estimated variance components were for single-facet designs, no further information on the reliability of the measures could be gleaned from this study. Despite arguments against the use of observer agreement indices in estimating reliability (e.g., McGaw, Wardrop, & Bunda, 1972), the coefficient of .90 was used as the best available estimate of reliability for the measures of silent, oral, and indirect reading. Importantly, there is no suggestion in the generalizability study that the reliabilities for silent and oral reading were appreciably different.

Teacher behaviors. Leinhardt et al. divided these into two areas: teacher instruction and teacher affective contact. Teacher instruction included model presentation, explanation, feedback, cueing, and monitoring, and was also recorded using the SOBR. Times spent in these activities were combined into a single estimate of the number of minutes per day a student received teacher instruction. Affective contact included reinforcers and cognitive press. Reinforcers were measured as the number of positive statements received by a student per day. Cognitive press was measured as a rating of the degree to which a student was focused on academic material and the degree to which the teacher supported that focus, recorded during each observational session in which a student was supposed to be engaged in academic activities other than reading. Again, in the absence of any better information, a reliability of .90 was assumed for each of the three measures of teacher behavior.

Overlap. Overlap was an estimate of the relationship between the curriculum content covered and the posttest measure of reading performance. It was measured as the number of items on the CTBS for which the content had been taught, and was obtained through a teacher estimate for each student. Unfortunately, no information on the reliability of this measure was available. Leinhardt and Seewald (1981b) report a correlation of .71 between this measure and a computer-based estimate of curriculum test overlap, using the same sample as Leinhardt et al., and this figure was taken as an approximation to the reliability. It is almost certainly an underestimate of the true reliability of the measure.

Pacing. The pacing variable was an estimate of the rate of movement through the reading material and was measured by counting the number of words in texts and workbooks assigned to be read by a student over three consecutive days. The natural log of this variable was used in this and the original analysis. The measure was assumed to be error-free (reliability = 1.0).

Original Model

In the first stage of the analysis, Leinhardt et al.'s original model was tested in its entirety (i.e., the two parts combined) by specifying the three measures of student reading behaviors as separate endogenous variables. The model examines relationships among the observed variables and, hence, the analysis is essentially a path analysis. As in the Leinhardt et al. study, only the CTBS was used as the criterion measure of reading performance. The posttest composite measure was not included in their original analysis because the measure of curriculum-test overlap was not calibrated for this measure.

Figure 2 presents the path diagram and results of the analysis. The goodness-of-fit statistics for the original model indicate a moderately good fit to the data ($X^2 = 22.84$, $df = 10$, $p = .01$; RMSR = .025). As expected, students' pretest performance has a significant direct effect on amount of time spent in silent reading ($t = 6.92$, $df = 99$, $p < .05$), but no such effect on the amount of time spent in oral reading ($t = -1.02$, $df = 99$, $p > .05$) or indirect reading ($t = .37$, $df = 99$, $p > .05$). The coefficients for the structural equation for the posttest are identical to the standardized regression weights obtained in Leinhardt et al.'s first regression equation (and this despite the use of maximum-likelihood estimation rather than ordinary-least-squares).

[Insert Figure 2 about here.]

In order to demonstrate the extent of confounding associated with students' entry-level abilities, the model was retested with the path between pretest and posttest removed. Predictably, this more restrictive model showed a significantly poorer fit to the data (difference $X^2 = 65.15$, $df = 1$, $p < .05$). More interesting is the change in the structural coefficients. The coefficient relating silent reading and posttest increases to .54, and the coefficient relating overlap and posttest increases to .34. There were no major changes in any other coefficients. Thus, when there is no control for entry-level reading abilities, silent reading as well as overlap absorb the variance in the posttest attributable to pretest performance. This finding reinforces our suspicion concerning the confounding of measures, and emphasizes the need to control fully for students' entry-level abilities.

Model Assuming Measurement Error

In the second stage of the analysis, the attempt was made to control more fully for entry-level ability by incorporating measurement error into the model. The model was tested with each observed variable serving as an indicator of a latent variable. Factor loadings and error variance components of observed variables were calculated from their reliabilities and the entire measurement model was fixed. For the pretest, the reliability calculated from the intercorrelations among the test components (.92) was taken as the best estimate. Again, only the CTBS was used as the posttest measure of reading performance.

Figure 3 presents the path diagram and results of the analysis. Overall goodness-of-fit was reasonable, albeit modest ($X^2 = 30.49$, $df = 10$, $p = .001$; RMSR = .023). Of special interest is the change in the coefficients for the posttest structural equation. The effect of pretest on posttest ($t = 7.41$, $df = 99$, $p < .05$) increases, and the effect of silent reading on posttest is now smaller and non-significant ($t = 1.24$, $df = 99$, $p > .05$), although still greater than that of oral reading ($t = 1.12$, $df = 99$, $p > .05$) or indirect reading ($t = -.15$, $df = 99$, $p > .05$). The latter two coefficients remain almost unchanged. The effect of overlap on posttest also becomes smaller and non-significant ($t = 1.35$, $df = 99$, $p > .05$), again indicating that this measure was confounded with pretest performance.

[Insert Figure 3 about here.]

These results are predicated on the accuracy of the estimate of pretest reliability. The estimate of .92 was our best guess, based on the construction of the composite measure, although the actual range of reliability estimates was .91 to .93. In order to examine the effect of silent reading over this range and beyond, a series of LISREL analyses was conducted in monte carlo fashion by varying the factor loading and error variance for the pretest and holding all other parameters in the measurement model at the previously established values. Pretest reliabilities were varied from 1.0 through .80 in decrements of .01, and the beta coefficient relating silent reading to posttest performance was noted at each step. The curve describing the function is shown in Figure 4. Given the range of reliability estimates, the curve indicates that beta for silent reading falls between .10 and .12 and that even at the upper-end of the range the coefficient fails to reach significance (alpha = .05). Indeed, it is only at a value of .18 that beta becomes significant, and then the reliability required is .98!

[Insert Figure 4 about here.]

It could be argued that these values of beta may be unduly influenced by the error variance components in the other measures. If the reliabilities of these other measures have been incorrectly estimated, then our finding concerning the status of the beta coefficient for silent reading might be in error. An extreme test of this possibility was provided by assuming error-free measurement in all measures except the pretest, and again performing the series of monte carlo runs for changes in the reliability of the pretest (1.0 through .80). The curve describing the function is also given in Figure 4. For reliabilities above .86, the magnitude of beta is attenuated due to the effects of measurement error. Moreover, a reliability of .99 is now needed if the coefficient is to reach significance (alpha = .05). To the extent that there is measurement error in the other measures, the standard error of the estimate (beta) is increased, and our basic finding concerning the status of the silent reading coefficient remains the same.

Alternative Models

In the third stage of the analysis, we decided to examine the effect of removing the measure of curriculum-test overlap. This measure had been shown to be confounded with students' pretest performance in the previous analyses and, in any case, there was little substantive interest in the variable. It was included in the Leinhardt et al. model as a control for differences in content coverage of items on the CTBS. The decision to remove overlap was further prompted by doubts about its adequacy as an estimate of the content covered in relation to the items on the CTBS. Inspection of the correlations among observed variables revealed that overlap was correlated just as highly with the posttest composite (.51) as it was with the CTBS (.50), despite the fact that it had been calibrated only for items on the latter measure.

The removal of overlap from the model assuming error-free measurement (our original model) yielded a X^2 of 5.95 ($df = 7, p = .55$; RMSR = .017) and from the model assuming measurement error, a X^2 of 9.05 ($df = 7$, p = .25; RMSR = .018). Clearly, when these results are compared with those for the two corresponding models in which overlap was included, it is apparent that this modification resulted in a markedly superior fit to the data (difference $X^2 = 16.89$, $df = 3$, $p < .05$ for models assuming error-free measurement; difference $X^2 = 21.44$, $df = 3$, $p < .05$ for models assuming measurement error). Moreover, there were no major changes in the structural coefficients in either case.

With the overlap variable removed, it now made sense to examine the use of the alternative measure of posttest reading performance (the composite measure). The posttest composite correlated more highly with other observed variables than did the CTBS, and was thought to provide a better criterion measure.

[Insert Figure 5 about here.]

The use of the posttest composite was examined in a series of regression models because correlations with this composite were available only for variables in the posttest structural equation. These regression models correspond to Leinhardt et al.'s first regression equation (predicting the posttest) except that in the present analyses measurement error was incorporated using the previously established estimates of reliability. Three models were run (see Figure 5). The first model used only the CTBS as the criterion (reliability = .94), and was analyzed to provide an appropriate basis for comparison for the other two models. The second model used only the posttest composite as the criterion (reliability = .92). The third model used both the CTBS and the composite as multiple indicators of posttest performance. The latent variable so constructed was thought to provide the most valid and reliable criterion. In this analysis, the error variance of the posttest composite measure was fixed according to its reliability (.92), and the variance of the latent variable was scaled to that of the true-score component of the composite, leaving the factor loading and error variance of the CTBS free to be estimated. The results of all three analyses appear in Table 2.

[Insert Table 2 about here.]

At issue here is the magnitude of the parameter estimates rather than the goodness-of-fit of the models (the measurement parameters for the first two regression models are fixed and df = 0). The results from the first model show only minor changes from those for the full model. The results from the second and third models, on the other hand, show that the effect of silent reading is reduced. Indeed, under the most favorable conditions for valid and reliable measurement of the criterion (model 3), it is the coefficient for oral reading which approaches significance, perhaps suggesting that time spent in oral reading had more effect on final reading achievement. The pattern of results obtained in these three analyses showed no change even when overlap was retained in the models.

Conclusion

The conclusion to be drawn from this reanalysis seems inescapable. Contrary to Leinhardt et al.'s finding, there is no persuasive evidence that silent reading had an effect on students' reading achievement. As expected, students' entry-level abilities had a significant direct effect on time allocated to silent reading but no such effect on time allocated to oral or indirect reading. Any attempt at examining the role of silent reading, therefore, needs to take this into account. When measurement error was incorporated into the model and initial abilities more adequately controlled, silent reading no longer showed a significant effect on posttest performance. Under alternative models of the data, there is even the suggestion that oral reading may have had more effect on final reading achievement.

The contrast between the results from Leinhardt et al. and our own analysis cannot be attributed to differences in the methods of estimation. In testing the original model, we were able to reproduce exactly the coefficients obtained by Leinhardt et al. in their first regression equation. Nor can the result easily be attributed to inaccurate estimates of the reliability of the measures. In testing the model with measurement error, the coefficient for silent reading failed to reach significance even when we allowed for some slippage in our estimate of pretest reliability and error-free measurement in the other measures. Lest this seem to be playing games with an arbitrary criterion for significance, the analysis of the alternative models showed that the coefficient is not only non-significant but also relatively small. Our regression models are comparable to Leinhardt et al.'s first regression equation and show the beta coefficient relating silent reading to posttest performance to be *less than half* the size of their original estimate.

The implications of these results for the research literature are substantial. Unfortunately, the data provided by Leinhardt et al. do not warrant the conclusion concerning the merits of silent reading over oral reading as suggested in reviews by Allington (1983, 1984), Hiebert (1983), Reutzel (1985) and others. At best, such an interpretation gives sanction to a very fragile finding. At worst, it is probably wrong. More generally, the results of the reanalysis also call into question the interpretations of the data with respect to the relationship between time spent in reading and student achievement (e.g., Anderson, Mason, & Shirey, 1984; Englert, 1984; Vallecorsa, Zigmond, & Henderson, 1985). The finding in the present study concerning the effect of oral reading is only tentative and no firm conclusion can be drawn from it.

The present reanalysis is not without its limitations. The raw data from the original study were not available and so we had to assume the distributions of observed variables were approximately normal. Judging by the mean and standard deviation, the distribution of silent reading time (mean= 13.68, SD = 8.82) may be slightly skewed (cf., oral reading time; mean = 13.40, SD = 7.52). Hence, our analysis may have slightly underestimated the relationship between silent reading and posttest performance. While this is possible, the extreme sensitivity of the silent reading coefficient to the effects of measurement error in the model (see Figure 4) suggests that skewness alone cannot account for the poor showing of the effect of silent reading.

Few people will be as disappointed in these results as we were. Given the promise of substantial gains to be made from increased silent reading in classroom reading instruction, it is indeed unfortunate that no convincing evidence for its positive effect can be found (see also, Clark, 1975; Clark & Spath, 1979; Haynes & Jenkins, 1986; Stallings, 1980). Other researchers concerned with the potential merits of silent over oral reading may find it hard to accept such an outcome. When facts and favor are at odds, however, empiricist traditions necessitate our reliance on the data. The words of Charles Darwin (1888) quoted at the beginning of this article provide a vivid reminder in this regard: "I have steadily endeavoured to keep my mind free so as to give up any hypothesis, however much beloved ... as soon as facts are shown to be opposed to it ... for with the exception of the Coral Reefs, I cannot remember a single first-formed hypothesis which had not after a time to be given up or greatly modified" (p. 83). To be sure, there are very few "coral reefs" in educational research--and the merits of silent over oral reading is probably not one of them. Having cast doubt on the findings of one of the better studies, the need for good empirical work in the area is now even more urgent than before.

References

Allington, R. L. (1983). The reading instruction provided readers of differing abilities. *The Elementary School Journal, 83,* 548-559.

Allington, R. L. (1984). Oral reading. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 829-864). New York: Longman.

Anderson, R. C., Mason, J., & Shirey, L. (1984). The reading group: An experimental investigation of a labyrinth. *Reading Research Quarterly, 20,* 6-38.

Clark, R. M. (1975). *A study of teacher behavior and attitudes in elementary schools with high and low pupil achievement.* Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

Clark, R. M., & Spath, G. (1979). *Exemplary schools--The statistical problems of identifying them, and the results of the study.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Comprehensive Test of Basic Skills. (1974a). Monterey, CA: CTB/McGraw-Hill.

Comprehensive Test of Basic Skills. (1974b). Technical Bulletin No. 1. Monterey,CA: CTB/McGraw-Hill.

Darwin, C. (1888). Recollections of the development of my mind and character. In F. Darwin (Ed.), *The life and letters of Charles Darwin* (Vol. 1, pp. 25-86). New York: Appleton and Co.

Englert, C. S. (1984). Measuring teacher effectiveness from the teacher's point of view. *Focus on Exceptional Children, 17,* 1-15.

Haynes, M. C., & Jenkins, J. R. (1986). Reading instruction in special education resource rooms. *American Educational Research Journal, 23,* 161-190.

Hiebert, E. H. (1983). An examination of ability grouping for reading instruction. *Reading Research Quarterly, 18,* 231-255.

Jastak, J. F., Bijou, S. W., & Jastak, S. R. (1976). *Wide Range Achievement Test.* Wilmington, DE: Jastak Associates.

Joreskog, K. G., & Sorbom, D. (1984). *LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods.* Mooresville, IN: Scientific Software.

Leinhardt, G., & Seewald, A. M. (1981a). Overlap: What's tested, what's taught? *Journal of Educational Measurement, 18,* 85-96.

Leinhardt, G., & Seewald, A. M. (1981b). Student-level observation of beginning reading. *Journal of Educational Measurement, 18,* 171-177.

Leinhardt, G., Zigmond, N., & Cooley,W. W. (1981). Reading instruction and its effects. *American Educational Research Journal, 18,* 343-361.

Linn, R. L., & Werts, C. E. (1982). Measurement error in regression. In G. Keren (Ed.). *Statistical and methodological issues in psychology and social sciences research* (pp. 131-154). Hillsdale, NJ: Erlbaum.

Lomax, R. G. (1980). *Testing a component processes model of reading comprehension development through linear structural equation modeling.* Unpublished doctoral dissertation, University of Pittsburgh.

Lomax, R. G. (1982). An application of generalizability theory to observational research. *Journal of Experimental Education, 51,* 22-30.

Lomax, R. G., & Cooley, W. W. (1980). *A scaling of Chall's stages of reading development.* Unpublished manuscript, University of Pittsburgh, Learning Research and Development Center.

McGaw, B., Wardrop, J. L., & Bunda, M. A. (1972). Classroom observation schemes: Where are the errors? *American Educational Research Journal, 9,* 13-27.

Reutzel, D. R. (1985). Reconciling schema theory and the basal reading lesson. *The Reading Teacher, 39,* 194-197.

Spache, G. D. (1972). *Diagnostic reading scales.* Monterey, CA: CTB/McGraw-Hill.

Stallings, J. (1980). Allocated academic learning time revisited, or beyond time on task. *Educational Researcher, 9,* 11-16.

Vallecorsa, A. L., Zigmond, N., & Henderson, L. M. (1985). Spelling instruction in special education classrooms: A survey of practices. *Exceptional Children, 52,* 19-24.

Table 1

Results from Regression Analyses by Leinhardt et al. (1981)

	Regression Weights			
	Raw	Standardized	Std Error of Raw	F
Equation 1: Predicting Posttest				
Pretest	6.24	.66	.67	86.2*
Overlap	.40	.18	.12	10.2*
Silent reading	1.00	.15	.47	4.5*
Oral reading	.50	.06	.43	1.3
Indirect reading	-.09	-.02	.27	.1
		Adjusted $R^2 = .72$		
Equation 2: Predicting Total Reading Behaviors				
Pretest	.70	.23	.22	10.1*
Teacher instruction	1.03	.44	.16	43.4*
Reinforcers	.04	.35	.01	26.4*
Cognitive Press	5.02	.22	1.63	9.4*
Pacing	2.54	.10	1.98	1.6
		Adjusted $R^2 = .59$		

* $p < .05$

Table 2

Results from Three Regression Models with Overlap Removed and Incorporating Measurement Error (Standardized Solution)

	Structural Coefficient	Std Error	t
Model 1: Predicting CTBS			
Pretest	.81	.09	9.04*
Silent Reading	.10	.09	1.05
Oral reading	.09	.06	1.38
Indirect reading	-.01	.06	-.22
Model 2: Predicting Composite			
Pretest	.94	.07	13.81*
Silent reading	.05	.07	.77
Oral reading	.09	.05	1.83
Indirect reading	.05	.05	1.10
Model 3: Predicting CTBS & Composite Latent Variable			
Pretest	.92	.07	13.26*
Silent reading	.07	.07	.98
Oral reading	.09	.05	1.87
Indirect reading	.03	.04	.70

* $p < .05$

Figure Captions

FIGURE 1. Causal model for explaining reading achievement analyzed by Leinhardt, Zigmond, and Cooley (1981).

FIGURE 2. Standardized solution for original model (* $p < .05$, numbers in parentheses are residual variances).

FIGURE 3. Standardized solution for model assuming measurement error (* $p < .05$, numbers in parentheses are residual variances).

FIGURE 4. Beta coefficient for silent reading as a function of reliability of pretest.

FIGURE 5. Three regression models with overlap removed and incorporating measurement error.

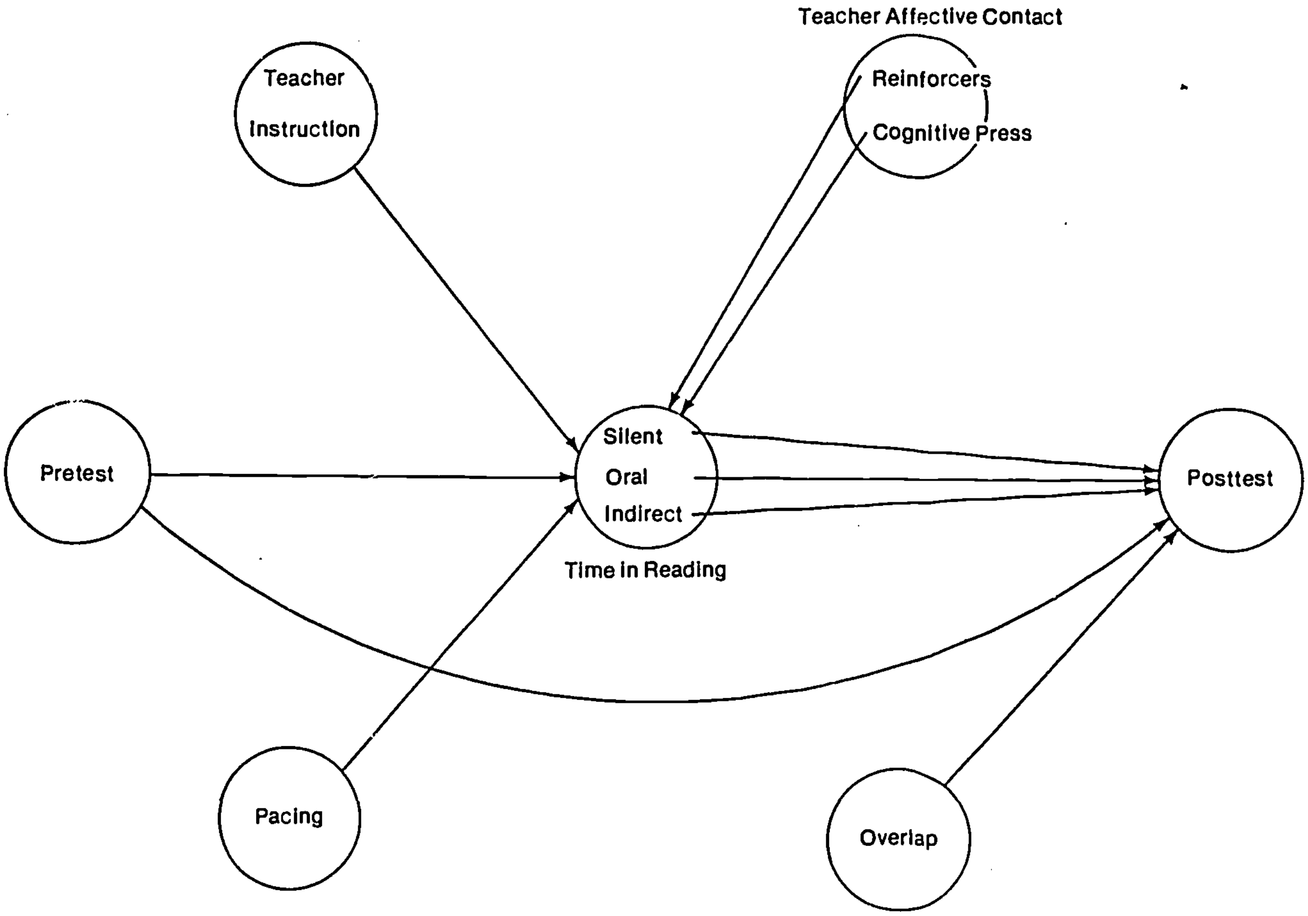

Teacher Affective Contact
Teacher
Instruction
Reinforcers
Cognitive Press
Pretest
Silent
Oral
Indirect
Time in Reading
Posttest
Pacing
Overlap

Figure 1

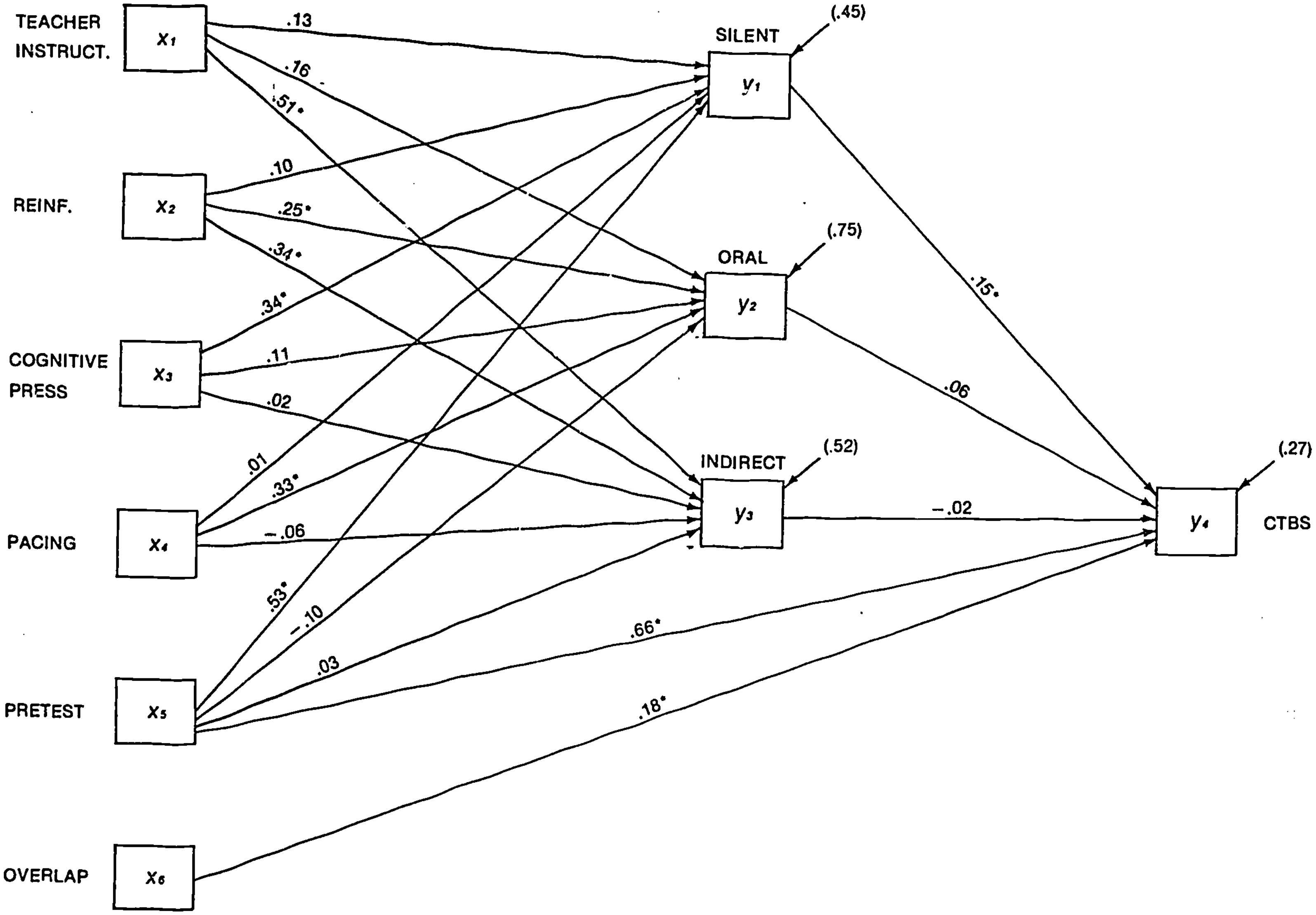
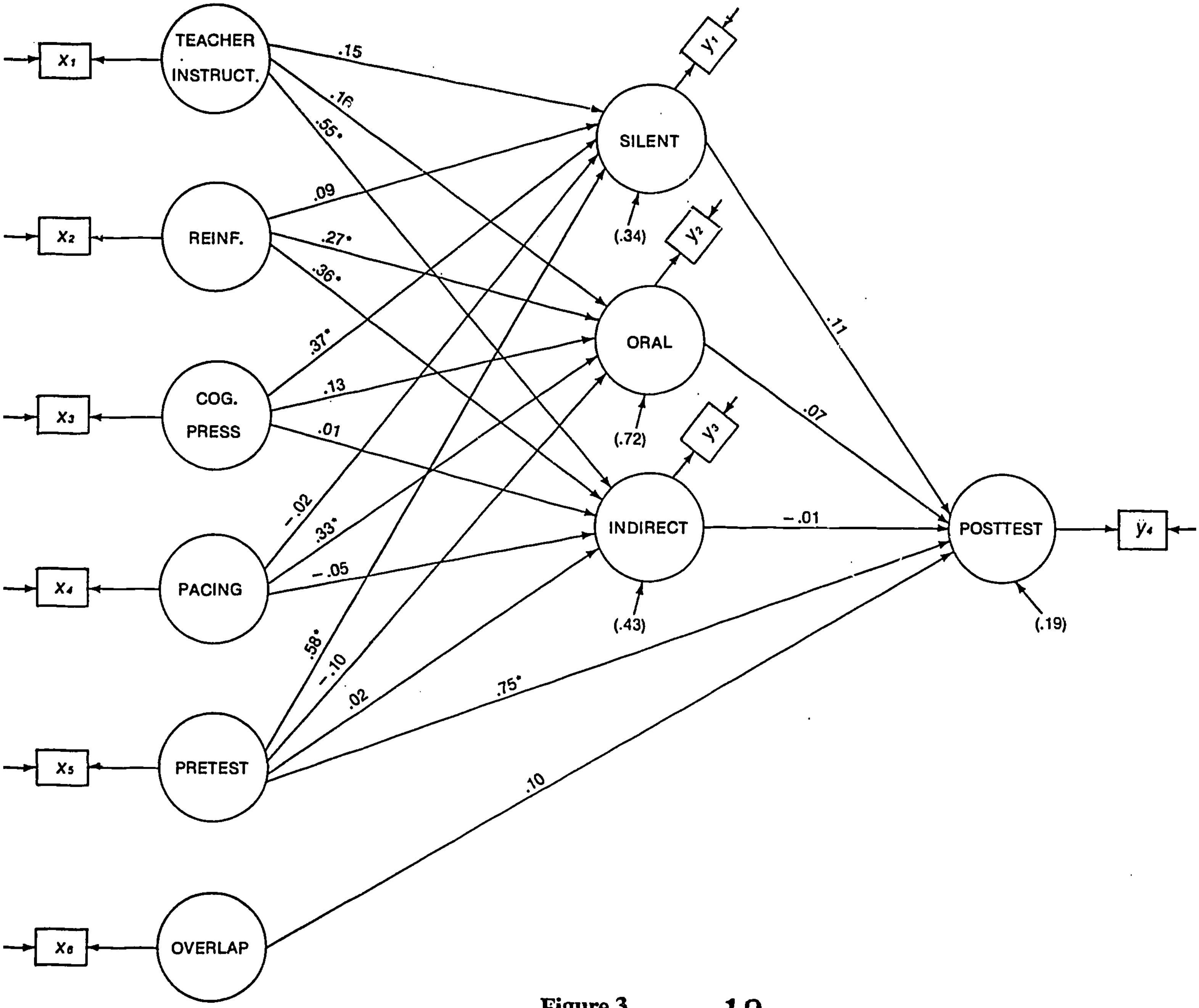

TEACHER INSTRUCT.
x_1
REINF.
x_2
COGNITIVE PRESS
x_3
PACING
x_4
PRETEST
x_5
OVERLAP
x_6
SILENT
y_1
(.45)
ORAL
y_2
(.75)
INDIRECT
y_3
(.52)
y_4
CTBS
(.27)
.13
.16
.51*
.10
.25*
.34*
.34*
.11
.02
.01
.33*
−.06
.53*
−.10
.03
.66*
.18*
.15*
.06
−.02

Figure 2



X_1
TEACHER INSTRUCT.
X_2
REINF.
X_3
COG. PRESS
X_4
PACING
X_5
PRETEST
X_6
OVERLAP
.15
.16
.55*
.09
.27*
.36*
.37*
.13
.01
−.02
.33*
−.05
.58*
−.10
.02
.75*
.10
SILENT
y_1
(.34)
ORAL
y_2
(.72)
INDIRECT
y_3
(.43)
.11
.07
−.01
POSTTEST
y_4
(.19)

Figure 3

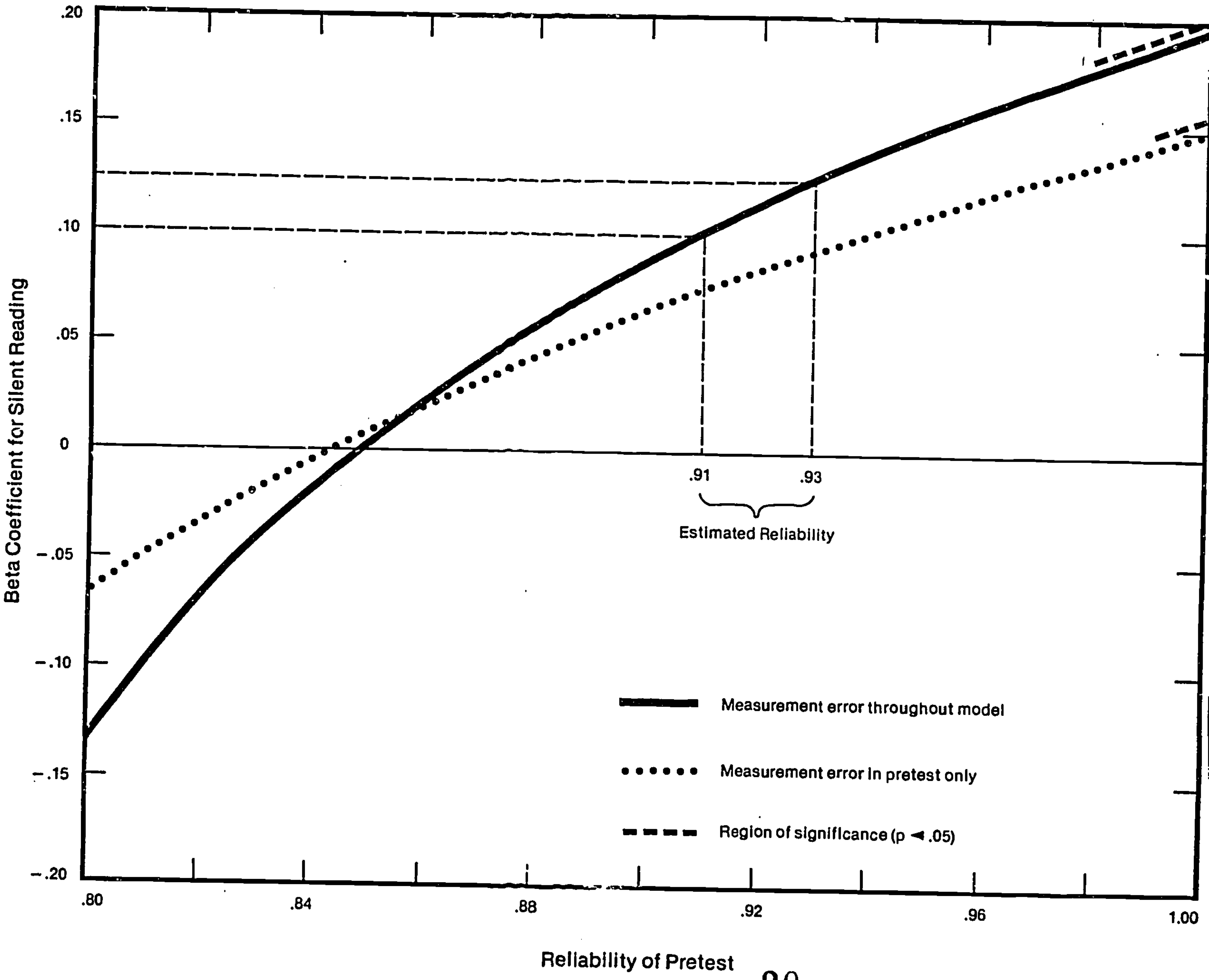
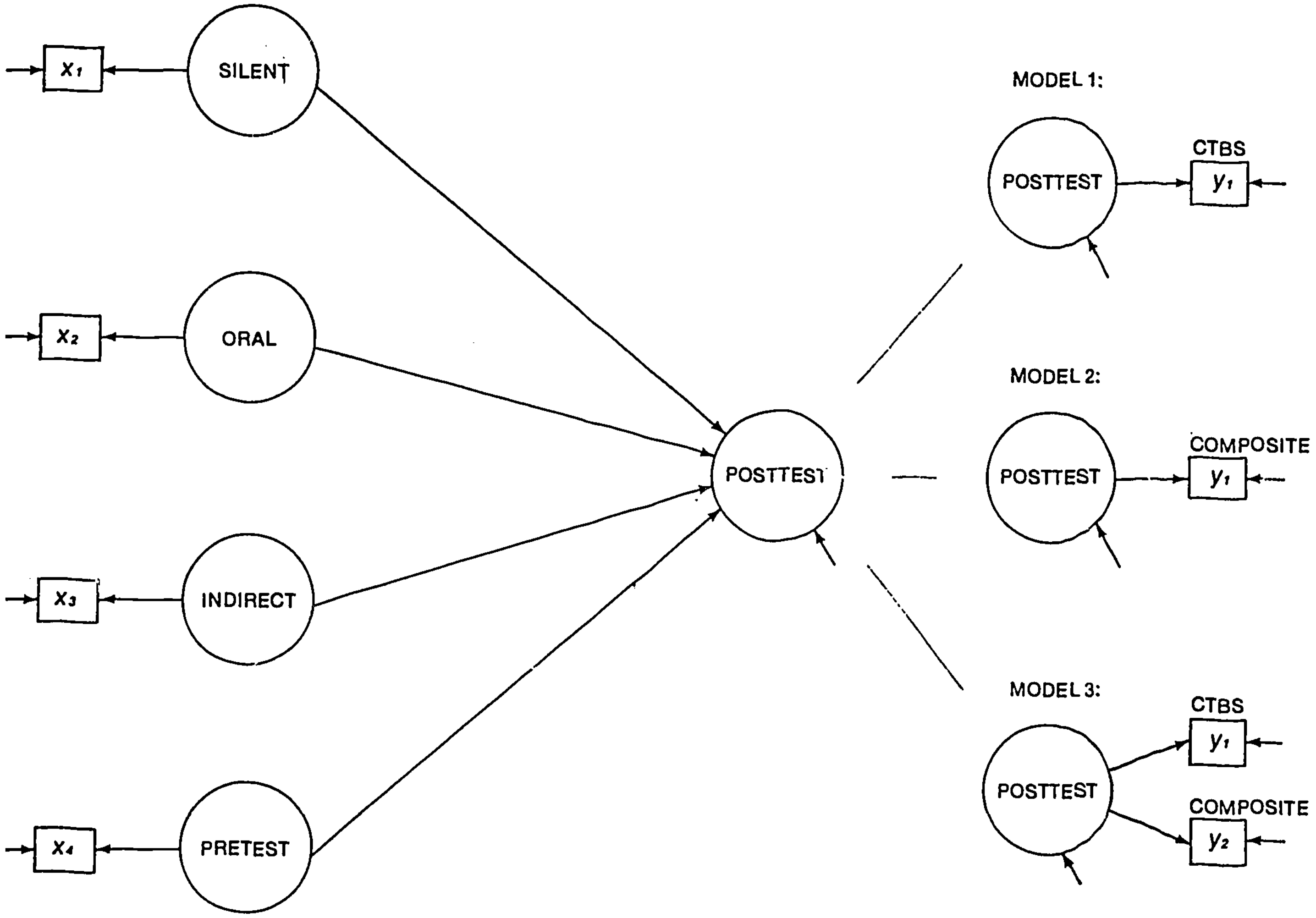

.20
.15
.10
.05
0
−.05
−.10
−.15
−.20
Beta Coefficient for Silent Reading
.80
.84
.88
.92
.96
1.00
Reliability of Pretest
.91
.93
Estimated Reliability
Measurement error throughout model
Measurement error in pretest only
Region of significance (p ◄ .05)

Figure 4



X_1
SILENT
X_2
ORAL
X_3
INDIRECT
X_4
PRETEST
POSTTEST
MODEL 1:
POSTTEST
CTBS
y_1
MODEL 2:
POSTTEST
COMPOSITE
y_1
MODEL 3:
POSTTEST
CTBS
y_1
COMPOSITE
y_2

Figure 5