DOCUMENT RESUME

ED 320 942 TM 015 205

AUTHOR Gaines, Margie L.; Davis, Margaret
TITLE Accuracy of Teacher Prediction of Elementary Student Achievement.
PUB DATE Apr 90
NOTE 32p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement; *Educational Diagnosis; Elementary Education; Elementary School Students; *Elementary School Teachers; Grade 2; Grade 4; Grade 6; *High Risk Students; Language Tests; Literacy; *Prediction; Racial Bias; Reading Tests; Remedial Instruction; Standardized Tests; Teacher Attitudes; *Teacher Expectations of Students
IDENTIFIERS Iowa Tests of Basic Skills

ABSTRACT

Whether teachers, in the absence of standardized test scores, can adequately identify students who require remedial services or are at risk of failing a literacy test was studied. During the 1987-88 school year 22 fourth-grade teachers were asked to identify which of 530 students of various ethnic backgrounds would score in the top or bottom quartiles of standardized reading and language tests. In a second study, 29 second-grade, 29 fourth-grade, and 25 sixth-grade teachers predicted which students (n=2,070) would score in percentiles 1 to 15, 16 to 35, 36 to 50, and above on standardized reading and mathematics tests. Students took the Iowa Tests of Basic Skills and a reading test. Fourth-grade teachers had similar accuracy rates for minority and white students, but had lower expectations for students who had been retained. The second study demonstrated that the teachers were better at identifying students who were truly at risk, and that the teachers had difficulty identifying students below average but not seriously at risk. Overall, the teachers were more likely to make pessimistic judgments about minority and low socioeconomic students, suggesting that the teachers were not conscious of how non-academic factors biased or mediated their expectations for students. An appendix contains 15 data tables. (SLD)

Accuracy of Teacher Prediction of Elementary Student Achievement

Margie L. Gaines, Ph.D.
Director of Program Evaluation

Margaret Davis, C.A.S.
Reading Specialist

York County Public Schools
302 Dare Road
Grafton, Virginia 23692

Paper presented at the annual meeting of the
American Educational Research Association
Boston, April, 1990

Abstract

Concern over the reliability of teacher judgment as a selection criteria for identifying students who need remedial services or who are at-risk of failing a literacy test raised the question of whether teachers can accurately identify such students without knowledge of standardized test scores. Two studies were conducted in a mid-sized suburban school district asking teachers to identify 1) students who would score in either the top or bottom quartile on a standardized reading test and a language test, and 2) students who would score in the 1st to 15th, 16th to 35th, 36th to 50th percentile range or above the 50th percentile on standardized reading and mathematics test. Predictions of standardized test performance were made concurrent with test administration. Overall accuracy rates for identifying students who scored in the bottom quartile or within the at-risk range in grades 2, 4, and 6 were examined. Differential accuracy rates for subgroups of students based on ethnicity, socioeconomic status, and retention status were also examined. Patterns of over- and under-predictions of performance were examined to determine if there was evidence supporting the hypothesis that teachers hold differential expectations for subgroups of students. Study 1 found that teachers had similar accuracy rates for minority and White students. However, retention status was found to be a biasing factor. Teachers held lower expectations for students who had been retained than for students who had never been retained. The results of Study 2 indicated that teachers were better at identifying students who truly were at risk but had difficulty identifying students who were below average but not seriously at risk. Frequency patterns of "hit" and "miss" predictions were not biased by ethnicity or socioeconomic characteristics. However, when the types of errors were examined by subgroup, second- and sixth-grade teachers were more likely to make pessimistic errors for minority and low-SES students than for their White and high-SES counterparts. Fourth-grade teachers, who had participated in Study 1 the previous year, made optimistic errors for low-SES and minority students. The results of the studies suggested that teachers were not conscious of how nonacademic factors biased or mediated their expectations for students. Inservice or feedback may be necessary before teacher judgment is a reliable criteria for selecting students to enter or exit remedial programs.

Introduction

The ability of teachers to accurately assess students' performance on an informal, on-going basis for the purpose of individualizing instruction to meet specific needs is critical. Especially in this time of increased need to remediate deficits with decreased resources, the role of teacher judgment is paramount in correctly identifying at-risk students. In Virginia, the state assessment program was revised to include a competency test at sixth grade. Beginning with the sixth-grade class of 1990, students must pass the reading comprehension, mathematics, and writing components of the literacy tests by the end of eighth grade to be classified as ninth graders. With the lowering of the competency gate from eleventh to sixth grade came increased emphasis upon identifying students at risk of failing the literacy tests. Revisions in the *Standards for Accrediting Public Schools in Virginia* (June, 1987) include the requirement for remediating students who score in the bottom national quartile on the reading, language, or mathematics portion of the state-adopted standardized test (Cognitive Abilities Test, Iowa Tests of Basic Skills, or Tests of Achievement and Proficiency).

Teacher prediction of student achievement begins early in kindergarten when classroom teachers assess readiness for reading and make recommendations for reading group placements in first grade. By the end of kindergarten teachers utilize standardized test data, criterion referenced curriculum based assessments, and informal assessment methods to identify students who are "at risk" and, therefore, require remediation or may qualify for special services. However, the earlier teachers can identify at-risk students, the earlier intervention may begin and the greater the likelihood of positive effects. The underlying question or problem which gave rise to this study is whether or not teachers can accurately identify those students who will score in the bottom national quartile on a standardized test. The accuracy with which teachers can correctly identify at-risk students impacts the ability of a school system to provide effective early intervention and appropriate modifications for students who may fail to profit from the regular instructional design. Relying on teacher judgment to identify

at-risk students raises the concern that over-identification of the at-risk population can reduce the effectiveness of any remedial program by diluting its efforts.

Because teacher judgment, whether formalized or informal, is an important component in the identification matrix for selection for remedial services and programs, the research questions of the studies were :

Study 1:

- What is the accuracy of teacher prediction of student achievement in reading comprehension at the fourth grade level?
- Is the number of years of experience in teaching and/or years of experience teaching fourth grade a factor in increasing the accuracy of predictions?
- Are ethnicity, socioeconomic status, or a student's status as a retainee a source of prediction bias?

Study 2:

- Is there a significant difference in teacher prediction accuracy between reading and mathematics, or among grade levels?
- Does experience in administering the criterion test instrument affect the accuracy of predictions?
- Does previous experience in predicting and feedback from previous predictions affect prediction accuracy?
- Can teachers make finer discriminations among levels of test performance than "bottom quartile" when making predictions?

Review of the Literature

The formation of expectations by teachers for students is a naturally occurring phenomenon in classrooms. Based upon available information, such as previous level of achievement, in-class performance, behavior patterns, ethnicity, socioeconomic factors, or physical appearance, teachers evaluate student performance and arrive at inferences and judgments about how particular students are likely to perform in the future. Teachers develop cognitive schemas over time to account for the complexity and simultaneity of the classroom context which help the teacher to reduce complexity and to deal

effectively and efficiently with classroom events and student behaviors (Doyle, 1979). As teachers' propositional knowledge aids them as classroom managers, teachers' expectations and judgments help them to manage uncertainty and exert a certain amount of control by "categorizing" students according to anticipated performance levels (Cooksey & Freebody, 1983). Expectancies, attributions and judgments based on formal and informal evaluations of classroom events and student performance are essential components of a teacher's instructional managerial skills. Teachers must constantly evaluate student performance and assess needs to adjust instructional techniques, grouping, or pacing of instruction.

The accuracy of teacher judgment in predicting student achievement has been an active field of study for over forty years beginning with early investigations of kindergarten and first grade teachers' accuracy in predicting success in first grade reading (Henig, 1949; Kottmeyer, 1947). These early studies established the predictive validity of teacher ratings as being equal to or greater than that of a standardized readiness test. Later studies found that teachers are able to make valid predictions of achievement in the elementary grades (Doherty, 1985; O'Connell, Dusek, & Wheeler, 1974; Oliver, 1978), and they are able to make valid long-range predictions covering a few years (Ebbesen, 1968; Keogh & Smith, 1970) or longer periods of time (Stevenson, 1986). Studies by Rosenthal (1968, 1973) posited that teacher expectancy is itself a causal factor in subsequent student achievement.

Although investigations into teachers' expectations have been extensive, research into teachers' cognitive thought processes is relatively new. Of interest are the information bases, theories and beliefs, and teachers' interactive thoughts and decisions that influence attributions and judgments about students. While early studies attempted to establish the value and reliability of teacher predictions of achievement, more recent studies have focused on identifying specific variables which influence or bias teachers' judgments.

The role of teacher presage variables in predictions of achievement has been investigated. One line of research has focused on teaching experience as a factor

influencing prediction or diagnostic proficiency (Burnett, 1963; Guskey, 1975; Kermonian, 1962; Shavelson, 1977). Burnett posited that prediction accuracy is a function of teaching experience or special training. Calderhead (1983) reported that experienced teachers use available information about students' backgrounds to anticipate behavior problems and which students would need special help. Novice teachers differ from experienced teachers in the types of student cues they attend to in their thinking and decision-making processes. Fogarty et al. (1982) found that novice teachers reported focusing on students' disruptive behaviors most frequently, whereas experienced teachers reported disruptive behavior infrequently in their reports of cues that led them to make decisions.

Researchers have hypothesized that student factors also affect teachers' attributions for the causes of students' performance. The role of student characteristics, including race, social class, and sex in affecting teachers' attributions have been investigated. Research findings of sex bias have not been uniform. Nonsignificant findings for the effect of sex on the causal attributions that teachers made for students' performance have been found (Wiley and Eskilson, 1978; Hanes, 1979). Sex bias in favor of girls appears to affect predictions of later achievement (Tobiessen & Duckworth, 1971; Keogh, 1970; Stevenson, 1976). Additional sources of bias found in the literature include time of year and halo effects (Guskey, 1975; Sullivan, Smith, & Lopez, 1989) and context clues such as the variance of a given class and norms for behavior within that class (Kagan, 1988).

While the bases teachers use to form expectancies are not clearly identified, some data suggest that teachers' expectancies are based on academic performance rather than on social criteria (O'Connell, Dusek, & Wheeler, 1974). However, several studies support hypotheses that teacher predictions are influenced by social factors such as students' ethnicity and socioeconomic status rather than just academic variables (Cooper, Baron, & Lowe, 1975; Doherty & Connolly, 1985; Goodwin, 1969; Payne, 1989; Rist, 1970). The attributions teachers make for the success or failure of White or Black students or of middle-class or low-income students are not clearly understood. Whereas some studies have found that teachers attribute academic failures of Black and low-SES students to internal

factors like ability or effort (Cooper & Burger, 1980; Rosenthal and Jacobson, 1968), other studies have found that Black and lower-class students were held less personally responsible for failure than were White middle-class students (Cooper, Baron, & Lowe, 1975; Wiley & Eskilson, 1978).

Wittingly or not, teachers form different expectations for groups of students who deviate from the norm. Applied to identified learning disabled or handicapped children, teachers viewed "labeled" students more negatively and thought they would have more problems than unlabeled students exhibiting the same behaviors (Foster & Ysseldyke, 1976; Foster, Schmidt, & Sabatino, 1976). It seems reasonable to suggest that labels may induce teachers to adjust their expectations for labeled students and to make different attributions for their successes or failures. Status as a retained student is another kind of label. Payne (1989) reported that a student's status as a retainee may bias teacher prediction of performance.

The review of the literature has revealed that teacher expectancy plays a role in influencing teacher behavior and attributions for student performance. Process-product research into educational settings has demonstrated clear links between teachers' expectations, attributions, and behaviors (Brophy & Evertson, 1981; Brophy & Good, 1974). In turn, teachers' cognitive assessments influence their decisions in areas of selection of materials, instructional techniques, grouping practices, pacing, and the amount of drill and practice needed (Brophy & Good, 1974; Dusek, 1975; Shavelson, 1973). While teacher prediction of achievement appears inextricably bound with teacher expectancy, the focus of the studies presented here is upon examining the accuracy of teacher predictions and drawing inferences about influential factors or sources of bias rather than probing for the content of teacher thinking.

Two studies were conducted in a mid-sized suburban school system of 9,000 students in kindergarten to grade 12. The nature of the studies was more exploratory than experimental in order to investigate overall prediction accuracy and accuracy for subgroups of students. Based on the literature, hypotheses for the studies predicted that teachers would underestimate the performance of

minority and low-SES students and overestimate the performance of high-SES and White students.

The studies were designed to provide information in response to practical concerns facing the school system: in the absence of standardized test scores for determining qualification for remedial services or identifying students at risk of failing the sixth grade literacy tests, can teacher judgment be a reliable alternative to test scores?

Method

Subjects

Teachers.

Study 1: During the 1987-88 school year, the 30 fourth-grade teachers in a suburban school system were invited to participate in the study; 22 (73%) agreed to participate. Each of the nine elementary schools was represented in the voluntary sample. Teachers' total years of teaching experience ranged from one to 23 years with a mean of 12.7 years, and experience teaching fourth grade ranged from one to 17 years with a mean of 8.3 years.

Study 2: During the 1988-89 school year, elementary teachers in the same system in grades 2, 4, and 6 were encouraged to participate in the study. Of the 84 regular classroom teachers at the three grade levels, only one teacher (grade 6) chose not to participate. There were 29 second grade, 29 fourth grade, and 25 sixth grade teachers who participated in the study.

Students. Students, by virtue of being in the classroom of a participating teacher, were subjects of the study.

Study 1: The sample included 530 students in grade 4. The ethnic distribution was 78% White, 18% Black, 3% Asian, and 1.5% Hispanic or American Indian/Alaskan Native. Sixteen percent of the sample were on the free or reduced-price Federal lunch program; 20% of the students had been retained at least once.

Study 2: The student sample included: 678 in grade 2, 656 in grade 4, and 736 in grade 6. The ethnic distribution of students was approximately 75% White

(80% at grade 4), 20% Black (17% at grade 4), and less than 4% other minority. The percentage of low-income students on free or reduced-price lunch averaged 20%. Retention status data were not collected as part of the second study.

Instrumentation

The Iowa Tests of Basic Skills (ITBS), Form G (grade 4) and Form H (grade 2 and 6), with 1985 norms published by the Riverside Publishing Company was administered to students in grades 2, 4, and 6 in March, 1988 and 1989. In the first study, predictions were also obtained for the Degrees of Reading Power (DRP) test administered immediately following the ITBS administration.

Procedures

Study 1: In April, 1988 after the ITBS was administered yet before results were returned, all fourth-grade teachers were sent a letter inviting them to participate in the research study. Teachers were informed that the purpose of the study was to identify at-risk students and to determine if teachers can accurately identify at-risk students without knowledge of test results. Teachers who agreed to participate returned a questionnaire which asked for the names of students they believed would score in the bottom national quartile (percentiles 1 to 25) or the top national quartile (percentiles 76 to 99) on the Total Language score on the ITBS or in the top or bottom quartile on the DRP using the instructional level reading score. Teachers also made brief written comments in response to the question, "What information did you consider in making the predictions?" The only teacher background information requested was total years teaching experience and years of teaching fourth grade.

Study 2: At the time the ITBS tests were administered, teachers were given instructions for marking special codes on students' answer sheets to indicate in which percentile range they predicted each student would score in reading comprehension or total mathematics. In one field, teachers were instructed to mark code 1 if they predicted the student would score in the 36th to 50th

percentile in reading comprehension, code 2 if they predicted the 16th to 35th percentile, and code 3 if they predicted the 1st to 15th percentile. The field was left blank if the teacher predicted the student would score above the 50th percentile. Mathematics predictions were coded similarly in a second field. Teachers were encouraged to cooperate in the research study by the following statements:

> As part of the Remedial Education Plan (STARS), especially the development of criteria for identifying students who need remediation, teachers at grades 2, 4, and 6 are being asked to predict who in their class will score at or below the 50th percentile in Reading Comprehension and Mathematics Total.
>
> The state requires that students who score in the bottom quartile (percentiles of 1 to 25) on a standardized test receive remediation. Teacher judgment is also important in identifying students who need assistance. Because York County does not test students at every grade level, teacher judgment will be especially important at those grade levels where the ITBS or TAP are not administered.
>
> By predicting (1) which students will score below the national average on this test and (2) the percentile rank range in which you predict they will score, you will be helping the division identify the characteristics of at-risk students.
>
> Teachers with a high rate of accuracy for prediction will be interviewed to learn how they predict performance—which characteristics and behaviors they consider and which they disregard. This information will be used in the development of county identification criteria and strategies for remediation.
>
> Your cooperation in providing this information is greatly appreciated.

Predictions were compared with actual percentile scores in reading comprehension and total mathematics. Chi Square analyses were used to examine the association between student characteristics and either hit or miss predictions or whether predictions were accurate, overestimates, or underestimates. Because the target group of interest was those students who were at-risk for academic difficulties, e.g., those who scored in the bottom quartile (plus ten percentile points to allow for standard error of measurement),

those who scored above the 35th percentile were excluded from some analyses. Strict and liberal applications of the criteria for prediction accuracy were used. Strictly speaking, an accurate prediction is a single cell, e.g., a prediction of 16th to 35th percentile for a student who scored in that range. Liberally applied, if any prediction (1-15, 16-35, or 35-50) was made for a student who scored within the 1st to 35th percentile it was considered a "hit." The liberal criteria was used because standardized tests have a standard error of measurement, and for instructional purposes it is more important that a teacher recognizes "at-riskness" than the degree of "at-riskness."

Results

Study 1: Overall, teachers made accurate predictions for 63% of the students who actually scored in the top quartile on the Degrees of Reading Power test and for 68% of the top-quartile students on Total Language on the ITBS. Teachers accurately identified 61% of the bottom-quartile students on the DRP and 64% who scored low on the ITBS.

A significant correlation was found between total years of teaching experience and accuracy rates for the DRP ($r=.45$, $p<.025$ for a one-tailed test); however, the correlation between total years experience and accuracy on the ITBS was not significant. On the other hand, the correlation between years of experience teaching fourth grade and prediction accuracy on the ITBS was significant ($r=.428$, $p<.05$ for a one-tailed test), but years of experience did not correlate with accuracy in predicting DRP results. When correlations were significant, teachers with more years of experience had higher accuracy rates than teachers with fewer years of experience.

When accuracy was examined for subgroups of students, there were essentially no differences in accuracy rates for minority or White students. Chi Square analyses were not significant. Teachers made accurate predictions for 63% of the top and 63% of the bottom White students on the DRP and 64% of the top- and 63% of the bottom-quartile minority students on the DRP. When making predictions for the top-quartile students on the ITBS Total Language, teachers

were accurate for 69% of the White and 69% of the minority students. Teachers were slightly less accurate for bottom-quartile White students (63%) than for bottom-quartile minority students (69%).

Differences, however, were found in accuracy rates for predictions about students who had been retained versus those who had never been retained. Accuracy patterns were similar for low-achieving retained students (75% on DRP, 83% on ITBS) and high-scoring nonretained students (66% on DRP, 68% on ITBS). However, accuracy rates dropped below 50% for top-quartile retained students and bottom-quartile nonretained students. This reversal in prediction accuracy patterns prompted an examination of the kinds of errors teachers were making.

Over-predictions and under-predictions of achievement were examined for students in the top-quartile, bottom-quartile and, by default, those who scored in the middle quartiles. An over-prediction (or optimistic error) resulted when the student scored in the bottom quartile but was predicted to score middle or high or when the student scored in the 26th to 75th percentile but was predicted to be top quartile. Likewise, under-predictions (or pessimistic errors) occurred when the student scored higher than predicted.

Teachers made more pessimistic errors for retained students than for nonretained students. Whereas teachers under-predicted the reading achievement of 45% to 57% of the retained students, the scores of 15% to 35% of the nonretained students were under-estimated. For students who scored in the top quartile on the DRP or the ITBS, teachers were more pessimistic about the achievement of retainees than nonretainees. At least half of the retainees who scored in the top quartile were not expected to score so well. By comparison the performance of about one-third of the nonretained group was underestimated. Even for students who scored in the bottom quartile, expectations were higher for nonretained students than retained students. Teachers made optimistic predictions for almost 60% of the nonretained students who scored in the bottom quartile on either test. Teachers made optimistic errors for 25% of the bottom-quartile retained students on the DRP and 17% of the bottom-quartile on the ITBS.

Teachers were asked to identify what information they considered when making their predictions. Teachers generated their comments rather than checked predefined options from a list. When comments were analyzed, the teachers most frequently considered classroom performance and test performance or level placement in the basal reading program when predicting top- and bottom-quartile students. Though less frequently identified, teachers also considered reading group placement, observations made during the standardized test administration, grades on class assignments, and study habits. Five of the 22 teachers mentioned identification as learning disabled and three mentioned retention status as pieces of information they considered when identifying low-scoring students.

Study 2. The Appendix shows the number of predictions within each range of scores for students who scored in the percentile bands of 1-15, 16-35, 36-50, and 51-99. The percent correct is based on the liberal application of the accuracy criteria. Predictions below the 50th percentile for students who scored anywhere below the 50th percentile were considered a "hit," and predictions over the 50th percentile were considered "misses." In general, teachers appear to be best at identifying those who scored in the lowest range as being below average. Even if they could not correctly identify how low such students performed, they could at least recognize that such students were below average. Chi Square analyses examining the frequency of any below-average prediction versus above-average predictions for students in each of the three below-average ranges were significant for grade 4 reading (χ^2 = 18.91, df=2, p<.0001) and mathematics (χ^2 =6.88, df=2, p<.05) and for grade 6 mathematics (χ^2 =9.85, df=2, p<.01), and approached significance for grade 2 mathematics. In each case teachers were more accurate making below-average predictions for students who scored within the 1st to 15th percentile range than for students who scored either 16th to 35th or 36th to 50th percentile.

Overall differences among grade levels were examined for hit and miss predictions of students who were at-risk (scored within the 1st to 35th percentile)

using the liberal application of the criteria for accuracy. A significant difference among grade levels was obtained for reading (x^2 = 14.07, df=2, p<.001) and mathematics (x^2 =8.32, df=2, p<.05). Sixth grade teachers were better than either second- or fourth-grade teachers at identifying students who scored below the 50th percentile. They made significantly more hits and fewer misses for at-risk students.

The patterns of accurate, over-, and under-predictions were examined for the total group of students at each grade level. Predictions in the diagonal of the tables (see Appendix) were summed to obtain the number of correct predictions. Inaccurate predictions above the diagonal were over-predictions (optimistic errors), and frequencies below the diagonal were under-predictions (pessimistic errors). The Chi Square values were significant for reading (x^2 =35.5, df=2, p<.0001) and for mathematics (x^2 =22.37, df=2, p<.0001). Sixth-grade teachers made fewer under-predictions and more accurate predictions than expected. Second-grade teachers made significantly more pessimistic errors for their students than expected.

Chi Square analyses were performed to examine prediction accuracy patterns for at-risk students within subgroups of students. When the hit versus miss frequencies were examined for low-SES versus high-SES students, the Chi Square values for grade 2 reading and grade 6 mathematics were significant (x^2 =6.65, df=1, p<.01; x^2 =9.8, df=1, p<.0001). Fewer misses than expected were made for low-SES students in second grade, and teachers missed more low-SES students than expected at grade 6. There was no significant association between patterns of hits and misses and ethnic group in either reading or mathematics at any grade level.

Patterns of optimistic and pessimistic errors for all students were examined to determine the direction of bias within socioeconomic and ethnic subgroups. Again, the frequencies in the diagonal of the table were summed for accurate predictions, and over- and under-prediction frequencies were summed. Significant Chi Square values were obtained for each analysis with the exception of low-SES versus high-SES at grade 2. Second-grade teachers made more pessimistic errors for minority students than for White students when predicting

reading (χ^2 =9.53, df=2, p<.01) and mathematics (χ^2 =10.77, df=2, p<.001) achievement. Sixth-grade teachers were also more likely to make pessimistic errors for minority students than White students in reading (χ^2 =18.74, df=2, p<.0001) or mathematics (χ^2 =13.73, df=2, p<.001). They were also more likely to under-estimate the performance of low-SES students than high SES students in reading and mathematics (χ^2 =17.52, df=2, p<.001; χ^2 =25.07, df=2, p<.0001, respectively). On the other hand, fourth-grade teachers, who had made predictions for their students in Study 1 were more likely to make optimistic errors for the low-SES students than high-SES students in reading (χ^2 =24.79, df=2, p<.0001) and mathematics (χ^2 =28.125, p<.0001). They also over-estimated the achievement of minority students in reading (χ^2 =19.40, df=2, p<.0001) and mathematics (χ^2 =13.202, df=2, p<.001).

Discussion

This study differs from traditional research on teachers' predictions of achievement. Whereas previous studies have focused on predictions over time, many using profiles of fictitious students, the studies presented here examined teachers' estimations of current levels of achievement to be validated by a standardized measure. Teachers had seven months of school for forming and revising their assessments and expectancies about their students. In essence, this study is about the nature of teachers' assessments about their students. Expectations cannot be suppressed or avoided because teachers build up expectations about their students simply from interacting with them. According to Good and Brophy (1973), the question of whether teachers solicit sources of information other than their classroom experiences with the students to form expectations is not as important as the question of how information is used. Information about students contributes to expectations about them, but information (regardless of the source) can be useful in planning individualized instruction to meet their specific needs.

The two studies presented here found that teachers make accurate

estimations of their students' level of achievement at a better than chance level. Overall, teachers were accurate at least 60 percent of the time. However, when the data were disaggregated, teachers were found to have differential accuracy rates for subgroups of students. Teachers were best at identifying students in the lowest percentile range. Most likely this is because students at the extreme lower end of the distribution have clearly identifiable deficits. An interview conducted with a teacher identified as being a good predictor in the first study characterized bottom quartile students as those who lack a framework for processing new information because of limited prior knowledge, require repetition, and cannot "sift" through material to evaluate the salience of pieces of information. The bottom quartile student differs from a student who has knowledge "gaps" and scores low. The rhythm of remediation is different for the latter student; less repetition is required, and improvement comes about more quickly.

Study 1 found retention status to be a biasing factor when estimating performance on standardized reading measures. Teachers apparently consider the retention status of a student when forming and communicating expectations about achievement; however, they may not be aware that this is a basis for their judgment because they rarely identified retention as a piece of information they considered. Underprediction of retainees has instructional implications. Pessimistic errors may point to lower expectations for these students which perhaps result in over-identification of retainees for remedial groups, further slowing the pacing of instruction for them, and an insidious lowering of their achievement over time.

Study 2 found that teachers were more likely to make under-predictions of performance for students in low-socioeconomic and minority subgroups than for their middle-class and White counterparts. This finding is consistent with much of the literature which has found that teachers generally hold lower expectations for these groups of students (Dusek & Joseph, 1985). The danger of differential expectations based on bias is the threat to instructional equity. Because teachers pursue differing instructional strategies and materials for groups of students who differ in actual or perceived achievement or need (Bawden, Burke, & Duffy, 1979; Brophy & Good, 1974; Metheny, 1980), modifications in the instructional design

precipitated by a source of error may be inappropriate for meeting the varying needs and learning styles of students. When socioeconomic status or ethnicity is a subconscious source of bias in the teachers' grouping or pacing decisions, incorrectly identified or grouped students may not profit from instruction and may indeed suffer.

Optimism may have as deleterious an effect on students as erroneously low expectations because students lacking the level of basic skills necessary to meet promotion standards or pass literacy or minimum competency tests go unnoticed in the classroom. In the second study many students in the high-SES or White subgroups who actually performed in the at-risk range (1st to 35th percentile) were not predicted to score low. Among the high-SES students in grades 2, 4, and 6, 19% were completely overlooked and expected to score above average; 21% of the White students who scored in the 1 to 35 range were not predicted to score low by their teachers. Optimistic errors were also found by Kottmeyer (1947) in an early study of predictions on Metropolitan Readiness Test scores and by Sullivan, Smith, and Lopez (1989) on teachers' predictions of scores on end-of-year writing competency tests. The reasons for optimistic errors have been attributed to halo effects, time of year bias, and self-enhancement and efficacy theories. In concurrent prediction studies when teachers are not forecasting performance at some later date, the reasons for over-estimating achievement still need to be examined.

That sixth-grade teachers had better accuracy rates than fourth-grade teachers who had done the prediction activity the previous year is an interesting finding. Several hypotheses may be proposed to account for this finding. First of all, with more years in school, variance in students' cumulative knowledge and variance in performance on assessment measures increases. With greater variation by sixth grade, students at the lower end of the distribution are more different from their peers who are on grade level than at earlier grade levels and, therefore, are more easily identifiable as in need of remediation. Promotion standards in the school system in which the studies were conducted permit students to be one

year below grade level and still meet minimum promotion to grade 6 or to grade 7. However, for entry to grade 2 students may not be more than one-fourth of a year below grade level or more than six months below grade level for promotion to grade 3. The stricter definition of "on grade level" or "at risk" at grade 2 may also explain the tendency for second-grade teachers to make more pessimistic estimations of their students' achievement. It may be that second grade teachers are not really worse predictors than teachers at other grade levels but rather that making fine discriminations among levels of achievement is a more difficult task when there is little variance. Teachers may view any deviation from the norm of expected achievement as a critical deviation.

Another hypothesis for sixth-grade teachers' better accuracy is their greater experience in using cumulative record information for making tracking and placement decisions when sending students on to seventh grade. Sixth-grade teachers in this school system make preliminary recommendations for seventh-grade course placements mid-year and then revise them in the spring when standardized test scores and other sources information are available. Teachers will also get feedback from the intermediate school about how their students are achieving in their courses. The importance of teachers' judgments, accountability for their recommendations, and feedback about their recommendations have perhaps created a feedback loop, which over time helps teachers improve and refine their cognitive processes for making decisions about students.

Something must be said against standardized achievement tests as a reliable indicator of current performance. The nature of standardized testing situations elicits students' frustration level performance and not their instructional level. From the written comments solicited in Study 1, teachers based their predictions on classroom performance—most likely based on students' instructional levels. Hence, the test situation builds in over-estimation of current instructional-level functioning. For some students, then, predictions appear to be low. Test scores, of course, also contain error of measurement which confounds the accuracy of predictions. If teachers were to predict scores on criterion-referenced tests or curriculum-based assessments which more narrowly test skills and which more

closely reflect the curriculum, accuracy rates would be expected to be higher than those obtained in studies which use standardized measures as the prediction criterion.

Directions for future research

Simple examinations of accuracy of teachers' estimations of current achievement on standardized measures or predictions of future performance mask important differences within subgroups of students. Study 2 clearly indicated that examination of accuracy patterns by socioeconomic or ethnic subgroup at the general level disguises patterns of over- and under-estimations. Bias in teachers' expectancies may not be revealed until data are given closer scrutiny. Further research should continue to disaggregate the data for subgroups of students, including subgroups based on level of achievement.

Asking teachers to discriminate among degrees of below-average achievement as in Study 2 is a more difficult task than asking teachers to identify "at-risk" or bottom-quartile students. That teachers were best at identifying the lowest students indicates that teachers are fairly good at picking out who is really at risk. However, teachers may not always be able to identify the borderline student. Future studies might attempt to describe how teachers make cognitive distinctions among degrees of "at-riskness" to determine who receives remedial services and who does not. Accurately differentiating students with severe needs for remediation from those with moderate or mild needs has important implications for the determining the type of intervention for an individual and the quality and effectiveness of remediation strategies.

Along these lines, further research into the effects of context cues as mediating factors in correctly identifying at-risk students may be warranted. Kagan (1988) suggested that the variance of a given classroom serves as a context cue in biasing expectancies. Truncated distributions within a class or school (such as an alternative school) influence teachers to adjust their expectations. Experience with variance within a classroom or across years of experience likely helps teachers develop an appreciation for the notion of variance and extremes of the normal curve distribution. Informal conversations with teachers by the

researchers have revealed that teachers create a normal distribution in their mind for their class even if variance does not exist in fact. Their perceptions of high and low are relative to the distribution of the class. Therefore, a suggested line of research might examine the influence of context cues across classrooms in a school or across years on anchoring expectations for certain types or subgroups of students. Without schemas for distributions and variance, teachers may identify the lowest students in their classroom as being at risk, but the lowest students, especially when in a homogeneously average or high-ability classroom, may not be at probable risk for academic failure.

Exploring the use of inservice and feedback systems for increasing sensitivity to student characteristics which mediate academic performance may be profitable in terms of improving the efficiency of identifying at-risk students and the effectiveness of service delivery. Novice teachers may profit from expert teachers in learning how to use academic, behavioral, and background information as cues in assessing performance and identifying instructional needs. Educational systems cannot afford to wait for teachers to gain expertise from years of experience. Research into the bases of teacher judgment systems should examine whether teachers are able to differentiate skill levels and performance within the various content areas or whether they respond to the general "g" factor when appraising students' performance. In the first study, teachers' written comments suggested that they expected poor students to perform uniformly low regardless of the content or task. In both studies, if a student was predicted to perform low in one area, performance in the other area was also expected to be low. The studies suggested that teachers predict globally rather than differentiate predictions by content area or for discrete subskills.

In conclusion, research in this area has revealed that expectancies are inevitable and that they influence instructional decisions. However, additional research is still needed to clarify the bases and biases of teacher judgments. At the present time, the use of untrained teacher judgment as a sole criteria for entrance to or exit from remedial services is risky. However, in combination with inservice or feedback and other criteria of performance, consciousness about expectations and, hence, the reliability of teacher judgment might be improved.

References

Brophy J., & Evertson, C. M. (1981). Student characteristics and teaching. New York: Longman.

Brophy, J. E., & Good, T. L. (1974). Teacher-student relationships: Causes and consequences. New York: Holt, Rinehart & Winston.

Burnett, R. W. (1963). The diagnostic proficiency of teachers of reading. The reading teacher. 16, 229-234.

Calderhead, J. (1983, April). Research into teachers' and student teachers' cognitions: Exploring the nature of classroom practice. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Cooksey, R. W. & Freebody, P. (1985). Generalized multivariate lens model analysis for complex human inference tasks. Organizational behavior and human decision processes, 35, 46-72.

Cooper, H. M., Baron, R. M. & Lowe, C. A. (1975). The importance of race and social class information in the formation of expectancies about academic performance. Journal of educational psychology. 67, 312-319.

Cooper, H. M. & Burger, J. M. (1980). How teachers explain students' academic performance: A categorization of free response academic attributions. American educational research journal, 17, 95-109.

Doherty, J. & Conolly, M. (1985). How accurately can primary school teachers predict the scores of their pupils in standardized tests of attainment? A study of some non-cognitive factors that influence specific judgments. Educational studies, 11 (1), 41-60.

Doyle, W. (1979). Making managerial decisions in classrooms. In D. L. Duke (Ed.), Classroom management (The seventy-eighth yearbook of the National Society for the Study of Education) (pp. 41-74). Chicago: The University of Chicago Press.

Dusek, J. B, & Joseph, G. (1985). The bases of teacher expectancies. In J. B. Dusek (ed.), Teacher expectancies. Hillsdale, N. J.: Lawrence Erlbaum Associates.

Ebbesen, J. A. (1968). Kindergarten teacher rankings as predictors of academic achievement in the primary grades. Journal of educational measurement, 5, 259-262.

Fogarty, J. L., Wang, M. C., & Creek, R. (1982, March). A descriptive study of experienced and novice teachers' interactive instructional decision processes. Paper presented at the annual meeting of the American Educational Research Association, New York City.

Foster, G. G., Schmidt, C. R., & Sabatino, D. (1976). Teacher expectancies and the label "learning disabilities." Journal of learning disabilities, 9, 58-66.

Foster, G. G. & Ysseldyke, J. (1976). Expectancy and halo effects as a result of artificially induced teacher bias. Contemporary Educational psychology, 1, 36-45.

Good, T. L. & Brophy, J. E. (1973). Looking in classrooms. (p. 81) New York: Harper & Row, Publishers.

Goodwin, W., & Sanders, J. (1969). An exploratory study of the effect of selected variables upon teacher expectation of pupil success. Paper presented at the annual meeting of the American Educational Research Association.

Guskey, T. R. (1975). The effects of change in instructional effectiveness on the relationship of teacher expectations and student achievement. Journal of educational research, 75, 345-349.

Hanes, B. F. (1979). Causal attributions by teacher-trainees for success and failure outcomes of elementary students labeled normal and gifted (Doctoral dissertation, Oklahoma State University, 1979). Dissertation abstracts international, 40, 3198-31995S. (University Microfilms No. 7928212)

Henig, M. S. (1949). Predictive value of a reading-readiness test and of teachers' forecasts. Elementary school journal, 50, 41-46.

Kagan, D. (1988). How do teachers define students at risk? The clearing house, 61 (7) 320-324.

Keogh, B. K. & Smith, C. E. (1970). Early identification of educationally high potential and high risk children. Journal of school psychology, 8 (4), 285-290.

Kermonian, S. B. (1962). Teacher appraisal of first grade readiness. Elementary English, 39, 196-201.

Kottmeyer, W. (1947, October). Readiness for reading. Elementary English, 360-366.

Metheny, W. (1980, January). The influences of grade and pupil ability levels on teachers' conceptions of reading (Research Series No. 69). East Lansing: Institute for Research on Teaching, Michigan State University.

O'Connell, E. J., Dusek, J. B., & Wheeler, R. J. (1974). A follow-up study of teacher expectancy effects. Journal of educational psychology, 66, 325-328.

Payne, B., Payne, D., Dagley, P. (1989). The ability of teachers to identify at-risk elementary students. Paper presented at the annual meeting of the American Educational Research Association.

Rist, R. (1970). Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. Harvard educational review, 40, 411-451.

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. New York: Holt, Rinehart & Winston.

Shavelson, R. J., Cadwell, J. & Izu, T. (1977). Teachers' sensitivity to the reliability of information in making pedagogical decisions. American educational research journal, 14, 83-97.

Stevenson, H. W. & Newman, R. S. (1986). Long-term prediction of achievement and attitudes in mathematics and reading. Child development, 57, 646-659.

Sullivan, H. J., Smith, M. & Lopez, C. (1989, March). Teacher bias in evaluating year long academic progress. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Tobiessen, J., Duckworth, B. & Conrad, G. (1971). Relationships between the Schenectady kindergarten rating scales and first grade achievement and adjustment. Psychology in the schools, 8, 29-36.

Wiley, M. G., & Eskilson, A. (1978). Why did you learn in school today? Teachers' perceptions of causality. Sociology of Education, 51, 261-269

Grade 2: All students

Actual Score Range	Predicted Range 1-15	16-35	36-50	51-99	N	% Correct
Reading:						
1-15	7	2	8	3	20	.85
16-35	6	26	14	22	68	.67
36-50	1	25	15	18	59	.69
51-99	2	43	84	250	379	.66
Total	**16**	**96**	**121**	**293**	**526**	
Math:						
1-15	6	11	6	6	29	.79
16-35	6	21	16	18	61	.70
36-50	2	26	13	30	71	.58
51-99	5	35	75	250	365	.68
Total	**19**	**93**	**110**	**304**	**526**	

Grade 2: Low SES students

Actual Score Range	Predicted Range 1-15	16-35	36-50	51-99	N	% Correct
Reading:						
1-15	11	4	1	1	17	.94
16-35	5	13	7	3	28	.89
36-50	0	7	13	10	30	.67
51-99	0	11	21	40	72	.55
Total	**16**	**35**	**42**	**54**	**147**	
Math:						
1-15	11	7	1	1	20	.95
16-35	4	6	5	4	19	.79
36-50	1	8	7	12	28	.57
51-99	0	12	23	45	80	.56
Total	**16**	**33**	**36**	**62**	**147**	

Grade 2: High SES students

Actual Score Range	Predicted Range 1-15	16-35	36-50	51-99	N	% Correct
Reading:						
1-15	7	2	8	3	20	.85
16-35	6	26	14	22	68	.67
36-50	1	25	15	18	59	.695
51-99	2	43	84	250	379	.66
Total	**16**	**96**	**121**	**293**	**526**	
Math:						
1-15	6	11	6	6	29	.79
16-35	6	21	16	18	61	.705
36-50	2	26	13	30	71	.58
51-99	5	35	75	250	365	.68
Total	**19**	**93**	**110**	**304**	**526**	

Grade 2: Minority students

Actual Score Range	Predicted Range 1-15	16-35	36-50	51-99	N	% Correct
Reading:						
1-15	7	2	4	2	15	.86
16-35	6	14	6	8	34	.76
36-50	0	12	10	7	29	.76
51-99	1	8	23	47	79	.59
Total	**14**	**36**	**43**	**64**	**157**	
Math:						
1-15	9	2	4	1	16	.94
16-35	5	9	6	11	31	.645
36-50	2	16	7	11	36	.69
51-99	1	10	14	47	74	.68
Total	**17**	**37**	**31**	**70**	**157**	

Grade 2: White students

Actual Score Range	Predicted Range 1-15	16-35	36-50	51-99	N	% Correct
Reading:						
1-15	9	4	5	3	21	.857
16-35	5	12	15	17	49	.65
36-50	1	9	17	21	48	.56
51-99	1	27	55	244	327	.74
Total	**16**	**52**	**92**	**285**	**445**	
Math:						
1-15	8	9	3	6	26	.77
16-35	4	9	12	11	36	.69
36-50	1	12	9	31	53	.415
51-99	4	16	62	248	330	.75
Total	**17**	**46**	**86**	**296**	**445**	

Grade 4: All students

Actual Score Range	Predicted Range 1-15	16-35	36-50	51-99	N	% Correct
Reading:						
1-15	16	9	6	1	32	.97
16-35	6	28	36	26	96	.73
36-50	1	19	31	39	90	.56
51-99	1	22	74	335	432	.77
Total	**24**	**78**	**147**	**401**	**650**	
Math:						
1-15	12	12	12	7	43	.84
16-35	11	14	14	23	62	.63
36-50	4	15	35	33	87	.62
51-99	6	26	77	347	456	.76
Total	**33**	**67**	**138**	**410**	**648**	

Grade 4: Low SES students

Actual Score Range	Predicted Range 1-15	16-35	36-50	51-99	N	% Correct
Reading:						
1-15	6	3	1	0	10	1.00
16-35	1	7	14	11	33	.67
36-50	0	5	10	7	22	.68
51-99	1	4	14	27	46	.587
Total	**8**	**19**	**39**	**45**	**111**	
Math:						
1-15	3	4	4	0	11	1.00
16-35	5	5	7	5	22	.77
36-50	0	4	12	6	22	.73
51-99	2	5	17	30	54	.55
Total	**10**	**18**	**40**	**41**	**109**	

Grade 4: High SES students

Actual Score Range	Predicted Range 1-15	16-35	36-50	51-99	N	% Correct
Reading:						
1-15	9	5	5	0	19	1.00
16-35	5	19	20	15	59	.74
36-50	1	14	19	29	63	.54
51-99	0	17	56	298	371	.80
Total	**15**	**55**	**100**	**342**	**512**	
Math:						
1-15	9	8	8	6	31	.806
16-35	6	9	7	14	36	.61
36-50	4	11	23	22	60	.63
51-99	3	21	60	299	383	.78
Total	**22**	**49**	**98**	**341**	**510**	

Grade 4: Minority students

Actual Score Range	Predicted Range 1-15	16-35	36-50	51-99	N	% Correct
Reading:						
1-15	4	3	2	0	9	1.00
16-35	2	7	15	11	35	.68
36-50	0	5	11	7	23	.69
51-99	0	5	18	39	62	.63
Total	**6**	**20**	**46**	**57**	**129**	
Math:						
1-15	7	5	6	2	20	.90
16-35	1	6	4	6	17	.647
36-50	1	5	15	8	29	.72
51-99	2	6	15	39	62	.63
Total	**11**	**22**	**40**	**55**	**128**	

Grade 4: White students

Actual Score Range	Predicted Range 1-15	16-35	36-50	51-99	N	% Correct
Reading:						
1-15	12	6	4	1	23	.956
16-35	4	20	21	15	60	.75
36-50	1	14	20	32	67	.52
51-99	1	17	55	294	367	.80
Total	**18**	**57**	**100**	**342**	**517**	
Math:						
1-15	5	7	6	5	23	.78
16-35	10	9	12	13	44	.70
36-50	3	12	23	20	58	.655
51-99	4	21	65	303	393	.77
Total	**22**	**49**	**106**	**341**	**518**	

Grade 6: All students

Actual Score Range	Predicted Range 1-15	16-35	36-50	51-99	N	% Correct
Reading:						
1-15	23	11	12	4	50	.92
16-35	8	30	49	13	100	.87
36-50	2	14	38	21	75	.72
51-99	2	19	89	398	508	.78
Total	**35**	**74**	**188**	**436**	**733**	
Math:						
1-15	19	19	11	7	56	.875
16-35	11	37	30	16	94	.83
36-50	2	15	31	34	82	.846
51-99	1	27	73	397	498	.797
Total	**33**	**98**	**145**	**454**	**730**	

Grade 6: Low SES students

Actual Score Range	Predicted Range 1-15	16-35	36-50	51-99	N	% Correct
Reading:						
1-15	8	4	4	1	17	.94
16-35	3	8	15	2	28	.82
36-50	1	4	15	1	21	.95
51-99	1	9	21	37	69	.55
Total	**13**	**25**	**55**	**41**	**135**	
Math:						
1-15	9	4	1	1	15	.93
16-35	6	2	10	18	36	.50
36-50	1	5	9	5	20	.75
51-99	0	11	15	45	71	.63
Total	**16**	**32**	**35**	**69**	**142**	

Grade 6: High SES students

Actual Score Range	Predicted Range 1-15	16-35	36-50	51-99	N	% Correct
Reading:						
1-15	14	7	8	3	32	.906
16-35	5	22	34	9	70	.87
36-50	1	9	22	13	45	.71
51-99	1	10	67	326	404	.807
Total	**21**	**48**	**131**	**351**	**551**	
Math:						
1-15	10	15	10	6	41	.85
16-35	5	25	19	8	57	.86
36-50	1	10	22	29	62	.53
51-99	1	16	54	320	391	.82
Total	**17**	**66**	**105**	**363**	**551**	

Grade 6: Minority students

Actual Score Range	Predicted Range 1-15	16-35	36-50	51-99	N	% Correct
Reading:						
1-15	13	4	5	2	24	.916
16-35	3	13	12	4	32	.875
36-50	1	5	16	5	27	.81
51-99	2	9	27	55	93	.59
Total	**19**	**31**	**60**	**66**	**176**	
Math:						
1-15	10	5	2	2	19	.89
16-35	5	17	11	7	40	.825
36-50	1	9	5	6	21	.71
51-99	0	14	15	65	94	.69
Total	**16**	**45**	**33**	**80**	**174**	

Grade 6: White students

Actual Score Range	Predicted Range 1-15	16-35	36-50	51-99	N	% Correct
Reading:						
1-15	9	7	5	2	23	.91
16-35	5	17	31	8	61	.87
36-50	1	9	20	13	43	.697
51-99	0	10	56	340	406	.837
Total	**15**	**43**	**112**	**363**	**533**	
Math:						
1-15	8	14	8	4	34	.88
16-35	6	20	15	9	50	.82
36-50	1	6	25	24	56	.57
51-99	1	12	56	323	392	.82
Total	**16**	**52**	**104**	**360**	**532**	