DOCUMENT RESUME

ED 348 350 SP 033 969

AUTHOR Mulholland, Lori A.; Berliner, David C.
TITLE Teacher Experience and the Estimation of Student Achievement.
PUB DATE Apr 92
NOTE 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 20-24, 1992).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Beginning Teachers; Comparative Analysis; Demography; Elementary Education; Evaluative Thinking; Mathematics Tests; *Predictive Validity; *Predictor Variables; Reading Tests; Scores; Teacher Expectations of Students; *Teaching Experience
IDENTIFIERS *Accuracy Measures; *Correlation Ratio; Iowa Tests of Basic Skills

ABSTRACT

Forty-two pairs of experienced and novice teachers predicted the rank order of their pupils' scores on the reading and mathematics portions of the Iowa Test of Basic Skills (ITBS). The pool of novice teachers were first semester students in the Arizona State University Professional Teacher Preparation Program (PTPP). The experienced teachers in this study were the placement teachers with whom the first semester PTPP students were placed. The correlation between perceived score and actual score on the ITBS was used as a measure of the accuracy of teachers' judgment of student achievement. The purpose of this study was to determine relationships between the accuracy of teachers' judgments of student achievement and the following variables: (1) years of teaching experience; (2) ethnic composition of classroom; (3) pupil gender; (4) class size; and (5) pupil ability as defined by scores on the ITBS. Correlations between the experienced and novice teachers' judgments were also obtained when both worked in the same classrooms. The experienced teachers were highly accurate in their predictions and significantly more accurate than novices; correlations varied widely within both groups of teachers. The relation between accuracy of predictions and years of teaching experience was negative but not substantial; there were no relationships between accuracy of predictions and classroom ethnicity, gender, and class size. Experienced teachers were more accurate in judging the performance of high scoring students than that of low scoring students, but not significantly so. Implications of these results and recommendations for further research are discussed. (Author/LL)

TEACHER EXPERIENCE AND THE ESTIMATION OF STUDENT ACHIEVEMENT

Lori A. Mulholland and David C. Berliner

Division of Psychology in Education

College of Education

Arizona State University

Tempe, AZ 85287

Paper presented at the meetings of the
American Educational Research Association
San Francisco, California
April, 1992

TM018552

ABSTRACT

Experienced and novice teachers predicted the rank order of their pupils' scores on the reading and mathematics portions of the Iowa Test of Basic Skills (ITBS). The correlation between perceived score and actual score on the ITBS was used as a measure of the accuracy of teachers' judgment of student achievement.

The purpose of this study was to determine relations between the accuracy of teachers' judgments of student achievement and the following variables: (a) years of teaching experience, (b) ethnic composition of classroom, (c) pupil gender, (d) class size, and (e) pupil ability (as defined by scores on the ITBS).

Correlations between the experienced and novice teachers' judgments were also obtained when both worked in the same classrooms. Experienced teachers were highly accurate in their predictions, with correlations between predicted and actual performance averaging .74 in reading and .73 in mathematics. Experienced teachers were significantly more accurate in their predictions than novices, who yielded average correlations of .51 in reading and .54 in mathematics. Correlations varied widely within both groups of teachers. The relation between accuracy of predictions and years of teaching experience was negative but not substantial (-.02 for reading, -.11 for math). There were no relations between accuracy of predictions and classroom ethnicity, gender and class size. Experienced teachers were more accurate in judging the performance of high scoring students than low scoring students, but not significantly so. Implications of these results and recommendations for further research are discussed.

Introduction

Research on teacher cognition suggests that teacher judgments of student comprehension are essential components in daily classroom decisions, in both the preinstructional and interactive phases of teaching (Borko, Cone, Russo & Shavelson, 1979). It follows then, that these regularly occurring judgments of student comprehension are used also in broader assessments of student achievement. Estimating the accuracy of such assessments has been the subject of a number of studies. What has not been explored as thoroughly is the role that teacher experience has in the accuracy of these judgments.

In a review of the literature on teacher-based judgments of student achievement, Hoge and Coladarci (1989) make reference to the general assumption held by many educational researchers and school psychologists, that these teacher perceptions of achievement are subject to bias and error. In their review of sixteen studies on teacher judgments of student achievement, Hoge and Coladarci concluded that, overall, teachers are actually very accurate in their perceptions of student achievement. Although the studies are varied in their methodological approaches to the question, they generally show moderate to high positive correlations between teacher judgments of student achievement and actual measurement of student achievement.

Nevertheless, several studies have revealed wide variations in the judgmental accuracy of individual teachers. For example, Helmke and Shrader (1987) found that correlations between estimates and criterion measures of achievement for a group of teachers ranged from .03 to .90. Hopkins, George, and Williams (1985) reported a range of correlations from .44 to .88 across individual teachers.

Clearly, teachers do vary in their ability to estimate student achievement. Thus, individual differences in the accuracy of teacher judgment should be further investigated to determine if differences among teachers are due to teacher characteristics, (e.g. experience, training, teaching philosophy) or if they are more a function of the composition of the class (e.g. ethnicity, ability, gender) or some other variable (e.g. type of measurement instrument used, assessment system of the school district)(Hoge & Coladarci, 1989).

The primary purpose of this study was to investigate the relationship of teacher experience to the accuracy of their judgments of student achievement. It was believed that individual differences in teacher accuracy might be explained primarily by teacher experience. Because new teachers are in the midst of trying to make sense of their experience and have not yet developed extensive case knowledge, it was expected that teachers with more experience would be better judges of student achievement than new teachers.

To determine whether there was a positive correlation between teachers' years of experience and accuracy of teacher judgments of student achievement, practicing teachers with varying years of experience (referred to as experienced teachers) and education students just beginning classroom field placements (referred to as novice teachers) were asked to make predictions of student achievement on a standardized test of achievement. These predictions were based on experienced teachers' and novice teachers' perceptions of student abilities.

Classroom ethnic composition was also examined in this study. Ethnic composition of the classroom had not been examined in relation to the accuracy of teacher judgments of student achievement and it was thought that it could explain, in part, individual differences in teachers' judgmental accuracy. Specifically, it was hypothesized that Anglo novice teachers would have more difficulty in assessing

comprehension and achievement among members of another ethnic group (Gage & Berliner, 1988). More experienced Anglo teachers, it was believed, would find this to be less of a problem, because of greater experience with children of various ethnic groups and ability levels.

Additional variables were also examined in this study. For example, the relations between accuracy of teachers' judgments of achievement and class size, gender, and ability were also explored. And the relations between novice and experienced teachers' judgments were examined for those cases where both novice and experienced teachers were working in the same classroom.

Method

Subjects

Participants in this study were recruited from three sources. One pool of subjects was first semester students in the Arizona State University Professional Teacher Preparation Program (PTPP). Thirty percent of the students in a required human development class chose to participate, and received extra credit.

All education students are assigned a field placement to provide them with the opportunity to interact with and learn from classroom teachers, as well as to see connections between theory and practice. As part of their coursework, students are placed in regular public school classrooms, where they observe the teacher, classroom environment, student behavior, and student-teacher interaction. The teacher with whom the PTPP student is placed is referred to as the placement teacher.

The design of the required child development class from which these students were recruited follows a combination lecture, discussion format with weekly observational projects which students complete in their field placement classroom. Through this class, students learn to observe, describe and explain child behavior.

Because the format of this class gives students experience in systematic observation and description of children's behavior and teacher's beliefs, practices and possible biases, their abilities in estimating student performance may not be typical of other beginning education students who are not involved in a similar class.

At the time this study was conducted, the state of Arizona required all students, grades 2 - 8, in regular classrooms (and not designated as learning disabled or using English as a second language), to take the annual Iowa Test of Basic Skills (ITBS) in April. Thus, only those novice teachers who were placed in classes grade 2 - 6, or in seventh or eighth grade mathematics or reading classes were eligible to participate in this study.

The second group of participants in this study was the placement teachers with whom the first semester PTPP students were placed. These teachers were recruited by their PTPP student. Again, like the PTPP students, the experienced teachers were eligible if they taught a regular class in second through sixth grade, or a seventh or eighth grade mathematics or reading class.

Forty-two pairs of teachers and PTPP students participated. No first or second term teachers were in this pool. To obtain some teachers with minimum experience, first and second year teachers in the geographic area were recruited through a mailing to former education students of the university. As incentive, these teachers were offered feedback regarding the accuracy of their judgments. Two first year teachers and three second year teachers agreed to participate in the study.

Materials and Measures

Iowa Test of Basic Skills. The Iowa Test of Basic Skills (ITBS), form J (Riverside Publishers, Chicago, 1990) was used as the criterion measure for determining the accuracy of teacher judgments of achievement. The ITBS was used because it is the only objective, reliable measure of student achievement that all Arizona students take every year. Teachers were also familiar with the administration of the test and students' performance on the measure. Finally, since it is given to all eligible students every year, it did not require any extra time on the part of the teachers or their students.

Ranking Form. Participants were asked to predict the performance of their students by rank ordering them according to their expected performance on the ITBS. Both experienced teachers and novice teachers used identical forms on which they placed the students in expected rank order for math and reading.

To ensure confidentiality of the teachers' judgments of students and the students' scores on the ITBS, code numbers were used to match names and rankings on the forms. These code numbers were a part of the ranking form. When ITBS results were returned to the school districts, name and code number were matched to the test results and then the student name files were destroyed.

Confidence Ratings. A five point Likert scale was used to assess participants' confidence in the accuracy of their rankings. There were two confidence scales; one for reading and one for math. In addition, novice teachers were asked to explain why they felt that degree of confidence in their rankings. This question was asked only of the novice teachers because it helped to identify those among the novices who never interacted with the students in their placement classroom for either math or reading. The novice teachers were only placed in the classrooms for a minimum of four hours a week and had varying amount of interaction with children. This scale was designed to qualify their answers.

Information regarding the ethnic composition of the classroom was obtained from a standard form that all education students complete about their placement class.

Procedure

Students were instructed to take the ranking sheet and list on it, in alphabetical order, the names of all the students in class who were taking the ITBS. Then they made a copy of the ranking form for their placement teacher, before they filled out the rankings.

Next the directions called for the student to rank the children in the order in which they were expected to finish for both the total math and total reading portions of the ITBS, based on what he or she knew about each student's abilities. Students were instructed not to discuss their beliefs about student abilities with their placement teacher until after the rankings were complete. Instructions also stated, in bold print, not to refer to any grades or past ITBS results for this task.

The PTPP students were asked to complete the confidence rating form. They were to give the placement teacher a set of directions, a confidence rating form and the copy of the ranking form they made.

Design and Data Analysis

Accuracy of teacher judgments was assessed in terms of the correspondence between teacher judgments of students' achievement and the actual performance of

students on the ITBS. For each class, six correlations were calculated as follows:

1. Experienced teacher rankings of student achievement X ITBS rankings, in both reading and mathematics.
2. Novice teacher rankings of student achievement X ITBS rankings, in both reading and mathematics.
3. Experienced teacher rankings of student achievement X Novice teacher rankings of student achievement, in both reading and mathematics.

These correlations were repeated for the top and bottom scoring thirds of the class, to see if there was differential accuracy or more agreement between novice and experienced teachers' predictions of student achievement for either the high or low scoring group of students. Similar analyses were conducted for gender, inquiring if teachers' accuracy in estimating achievement is higher or lower for boys or for girls.

To determine the importance of teaching experience in the accuracy of teacher judgments of student achievement, correlations were calculated between teachers' years of experience and accuracy in predicting mathematics and reading achievement. In addition, the General linear model (GLM) procedure was used after the continuous variable of years of teaching experience was grouped, to test for differences in mean accuracy among the groups.

The classroom ethnic composition variable was analyzed by correlating the percentage of minority students in the classes of Anglo experienced and novice teachers with their accuracy. The GLM was used with percentage minority students grouped, to test for differences among means.

To determine how much teachers differentiate between student reading and math abilities, and how this is related to teachers' accuracy, the absolute difference between teachers' predicted rankings of each student's reading and math achievement was calculated. An average was taken for each class and these averages were correlated with the average reading and math accuracy of the teachers. Correlations were also calculated between class size and accuracy, and confidence ratings and accuracy.

Results

Correlations among teacher judgments and student performance on the ITBS: Teacher judgmental Accuracy

Spearman rank correlations were used to assess the relations among teachers' judgments of student performance on the ITBS and students' actual performance on the ITBS. The correlation coefficients for both novice and experienced teachers are shown in Tables 1 and 2.

Each correlation was transformed to a Fisher's Z coefficient. The mean Z was calculated and then transformed back to r. For all analyses in which average correlations were used, the Fisher's Z-transformation was used to more closely reflect normality (Helmstadter, 1970). This method of reporting mean correlations has been used by other researchers (Farr & Roelke, 1971; Coladarci, 1986).

As expected, there was wide variability in both experienced and novice teachers' judgments of student ITBS performance and students' actual ITBS performance, in both reading and mathematics. Experienced teachers' predictions of student reading achievement correlated positively with ITBS reading results. Correlations for individual teachers ranged from .48 to .95, with a mean of .74. Novice

teachers' predictions of student reading achievement also correlated with ITBS reading results, with the range of correlations between .21 and .74, and a mean of .51. The correlation coefficients for both novice and experienced teachers' judgmental accuracy in reading are shown in Table 1.

The accuracy of teacher judgments for math achievement was slightly lower than judgments of reading achievement in the experienced teacher group (.68 versus .72) and nearly identical to reading judgments in the novice teacher group (.51 versus .49). Correlations for math judgments are shown in Table 2. The range of correlations for experienced teachers' predictions of math performance and ITBS math results was between -.08 and .92, with a mean of .73. The range of correlations for novice teachers' predictions of math performance and ITBS math results was between -.06 and .83, with a mean of .54.

Correlations between experienced and novice teachers' judgments of reading and mathematics performance

As with the teacher judgments and ITBS results, there was also a wide range of correspondence between novice teachers' judgments of student performance and the judgments of the experienced teachers with whom they were placed. Correlations for reading ranged from .22 to .94 with a mean of .65. Correlations for mathematics ranged from .00 to .86, with a mean of .62. These correlations are shown in Table 3.

Teacher experience and accuracy of judgments

The prediction that years experience would be positively correlated with accuracy was not confirmed. The correlations computed for years of teaching experience and the accuracy of experienced teachers' judgments of reading and mathematics were -.02, $p<.88$, and -.11, $p<.48$ respectively. Although the relationship had been predicted to be positive, the correlations were not substantially different from zero.

The general linear model (GLM) was also used to test for differences between means associated with years of experience. The independent variable was years of teaching experience, which was separated into five groups as follows: (a) novice teachers (zero years experience); (b) one to five years experience; (c) six to ten years of experience; (d) eleven to fifteen years of experience; and (e) more than fifteen years of experience. The dependent variables were the teachers' judgmental accuracy for reading and mathematics. The overall effect for years of experience was significant, $F = 3.35$, $p<.014$ for math and $F = 10.79$, $p<.0001$ for reading. However, the only mean that was significantly different from others was that of the novice teacher group. Tables 4 and 5 show means, standard deviations and F values for this analysis. When a separate analysis with the GLM was done, without the novice group included, there were no significant differences between the means for years' experience. $F = .13$, $p<.94$, for reading and $F = .17$, $p<.91$ for mathematics.

Correlations between teacher judgmental accuracy and classroom ethnic composition

It was predicted that there would be a negative correlation between the judgmental accuracy of novice Anglo teachers and total percentage of minority students in the classroom. This was not confirmed. The correlations between novice teachers' judgmental accuracy and percent minority students in the classroom for

reading was .08. For mathematics the correlation between judgmental accuracy and percent minority enrollment was .09. It was also predicted that there would not be a relation between the judgmental accuracy of experienced teachers and percent minority students. This prediction was confirmed. The correlation between experienced teachers' judgmental accuracy and percent minority students in classroom for reading was -.25 and for math was -.04. None of the correlations were significantly different from zero.

The results of the GLM analysis are shown in Tables 6-9. Percent minority students in classrooms of Anglo novice and expert teachers were independent variables. Dependent variables were Anglo novice and experienced teachers judgmental accuracy in reading and math. The only effect that was significant at alpha .05 suggested that experienced teachers' predictions of math performance was affected by the percentage of minority students in their classrooms. Three of the comparisons were significantly different from each other.

Scatterplots of correlations between judgmental accuracy and percent minority were examined. The plot of experienced teachers' accuracy in math appeared to be curvilinear and calculation of Eta squared confirmed this. Although the plot for novice teachers' accuracy in math also looked curvilinear, the Eta squared coefficient was no different from r.

Teacher judgmental accuracy for top and bottom scoring thirds of class

Mean Fisher's Z values and corresponding correlations for students scoring in top and bottom thirds of classes are shown in Table 10. All means except one revealed that teachers, on average, are slightly more accurate in their judgments of top scoring students than in their judgments of low scoring students. There was more agreement in judgments between experienced and novice teachers on the top scoring students for reading, but more agreement in judgments between experienced and novice teachers on the low scoring thirds of classes for mathematics. However, z-tests between correlation coefficients used to determine the significance of the differences between top and bottom thirds for teachers' math and reading accuracy, failed to find a significant difference.

Correlations between teacher judgmental accuracy and student gender

Average correlations, and z-tests of teacher judgmental accuracy for boys and girls is reported in Table 11. There were no significant differences between teacher accuracy for girls versus boys. Although most every teacher showed some difference in judgmental accuracy of girls versus boys, most were not large and half the teachers were more accurate in their judgments of girls and half were more accurate in their judgments of boys.

Correlations between teacher judgmental accuracy and class size

The prediction that class size and judgmental accuracy would be negatively correlated was not confirmed. Although three of the four correlations between

teachers' judgmental accuracy and class size were negative, they were not significantly different from zero. The correlations between class size and experienced teachers' judgmental accuracy in reading was -.04 and was -.01 in mathematics. Correlations between class size and novice teachers' judgmental accuracy in reading was -.05. The only positive correlation was between class size and novice teacher judgmental accuracy in math, but again was so low (.13) that it was not significant. Correlations between class size and teacher judgmental accuracy are shown in Table 12.

Correlations between teacher confidence ratings and teacher judgmental accuracy

The novice and experienced teacher groups differed in ratings of confidence they awarded their own judgments. Table 13 shows the means, standard deviations and correlations between confidence and accuracy. Not surprisingly, experienced teachers were over one point higher than novices in confidence that their judgments were accurate in both reading and math. However, there was no relation between either groups' confidence in their rankings and the actual accuracy of those rankings.

Correlations between teacher differentiation of student math and reading ability and accuracy

The correlation between experienced teachers' judgmental accuracy (average of reading and math accuracy) and average differentiation between teacher judgments of student reading and math performance was -.42. This moderate negative relationship suggests that teachers who are highly accurate in their judgments of student achievement tend not to differentiate between student mathematics and reading ability in making their judgments.

Table 1

Correlations Between Teacher Judgments of Student Reading Performance and ITBS Reading Results

Experienced Teachers	Class Size	Novice Teachers	Class Size
.95	24	.74	25
.93	22	.74	25
.89	25	.74	25
.88	23	.71	15
.87	25	.69	26
.86	19	.67	28
.85	21	.66	23
.82	23	.66	27
.81	22	.62	23
.80	21	.62	23
.75	28	.60	25
.75	25	.60	26
.75	19	.59	22
.74	26	.56	23
.73	24	.54	21
.73	26	.47	18
.73	23	.46	26
.73	21	.44	23
.72	15	.43	21
.71	23	.43	22
.70	24	.43	24
.70	27	.43	23
.68	22	.41	23
.68	30	.39	21
.68	26	.37	31
.68	31	.36 *	21
.62	23	.34 *	23
.62	25	.33 *	19
.62	26	.32 *	26
.59	23	.32	30
.59	23	.27 *	15
.58	21	.24 *	21
.55	24	.23 *	24
.48	25	.21 *	24
.48	21		

* Not significant at .05 alpha level. For class size < 30, significance determined by Critical values table of Spearman's rank correlation coefficient of r for Ho: p=0. Significance for class sizes > 30 was determined by Pearson critical values table of r for Ho: p=0. Tables J & K, Glass and Hopkins, (1984).

Table 2
Correlations Between Teacher Judgments of Student Math Performance and ITBS Math Results

Experienced Teachers	Class Size	Novice Teachers	Class Size
.92	22	.83	26
.90	22	.79	25
.90	20	.77	21
.89	31	.76	20
.89	23	.75	21
.88	26	.70	21
.88	25	.69	21
.87	19	.68	27
.86	25	.68	24
.86	24	.67	23
.86	21	.66	28
.83	27	.66	26
.82	21	.63	19
.81	23	.61	16
.79	23	.59	31
.77	16	.59	23
.76	23	.58	26
.75	19	.58	23
.75	24	.55	22
.73	25	.55	21
.73	21	.54	24
.71	26	.54	22
.70	28	.53	23
.70	26	.52	25
.70	25	.52	21
.70	21	.48	23
.69	24	.47	23
.68	23	.46	24
.66	29	.44	26
.66	21	.44	17
.63	26	.39	28
.62	21	.39	24
.56	24	.35	25
.54	23	.35	23
.54	22	.25 *	24
.54	22	.21 *	15
.46	16	.20 *	29
.43 *	15	.09 *	21
.39	24	.04 *	22
.38	24	-.06 *	19
.33 *	21		
.32 *	19		
-.08 *	28		

* Not significant at .05 alpha level.

Table 3
Correlations Between Novice and Experienced Teacher Judgments of Student Math and Reading Performance

Reading	Class Size	Math	Class Size
.94	25	.86	26
.92	21	.83	21
.90	24	.81	26
.87	26	.81	25
.83	28	.81	24
.83	23	.81	21
.83	23	.81	21
.81	15	.79	24
.80	21	.78	21
.74	26	.76	26
.74	23	.75	28
.69	25	.73	23
.69	24	.73	23
.68	25	.70	31
.68	25	.68	25
.67	27	.67	27
.64	26	.67	22
.62	23	.66	24
.62	22	.64	25
.61	30	.64	19
.61	23	.62	26
.61	22	.58	22
.56	31	.54	21
.55	26	.54	21
.51	21	.49	16
.46	19	.47	24
.45	24	.47	24
.43	21	.46	23
.26 *	23	.42	26
.25 *	23	.42	23
.22 *	21	.36	22
		.36 *	15
		.32 *	19
		.31 *	28
		.30 *	23
		.27 *	21
		.32 *	19
		.31 *	28
		.30 *	23
		.27 *	21
		.20 *	23
		.00 *	29

* Not significant at alpha .05 level.

Table 4
Means and Standard Deviations (SD) on Accuracy of Experienced Teachers' Predictions of Math Performance by Years of Experience

Years Experience	n	Mean	SD
1) 0 (novices)	39	.51	.210
2) 0 - 5	11	.71	.180
3) 6 - 10	7	.70	.134
4) 11 - 15	11	.67	.291
5) 16 +	13	.66	.191

GLM analysis has shown that differences can be inferred between means.

F (4/63) =10.79, $p<.0001$
Multiple t (LSD) differences: 1-2, 1-3, 1-4, 1-5

Table 5
Means and Standard Deviations (SD) on Accuracy of Experienced Teachers' Predictions of Reading Performance by Years of Experience

Teaching Experience	n	Mean	SD
1) 0 (novices)	33	.49	.164
2) 0 - 5	10	.73	.136
3) 6 - 10	7	.70	.125
4) 11 - 15	6	.73	.111
5) 16 +	12	.72	.116

GLM analysis has shown that differences can be inferred between means.

F (4/76) =3.35, $p<.0140$
Multiple t (LSD) differences between 1 - 2, 2 - 4, 3 - 4

Table 6

Means and Standard Deviations (SD) on Accuracy of Experienced Teachers' Predictions of Reading Performance by Percent Minority in Class

% Minority	n	Mean	SD
1) 0 - 10%	8	+.76	.078
2) 11 - 20%	11	+.72	.130
3) 21 - 50%	9	+.73	.134
4) 51 -100%	7	+.68	.124

GLM analysis has shown that no differences can be inferred between means.

F (3/31) =.64, p<.594

Table 7

Means and Standard Deviations (SD) on Accuracy of Experienced Teachers' Predictions of Math Performance by Percent Minority in Class

% Minority	n	Mean	SD
1) 0 - 10%	13	+.60	.261
2) 11 - 20%	11	+.77	.146
3) 21 - 50%	11	+.76	.117
4) 51 -100%	8	+.58	.193

GLM analysis has shown that differences can be inferred between means.

F (3/39) =.64, p<.05

Multiple t (LSD) differences between 1 - 2, 2 - 4, 3 - 4

Table 8

Means and Standard Deviations (SD) on Accuracy of Novice Teachers' Predictions of Reading Performance by Percent Minority in Class

% Minority	n	Mean	SD
1) 0 - 10%	8	+.46	.194
2) 11 - 20%	9	+.47	.160
3) 21 - 50%	7	+.57	.171
4) 51 -100%	7	+.49	.134

GLM analysis has shown that no differences can be inferred between means.

F (3/27) =.63, p<.60

Table 9

Means and Standard Deviations (SD) on Accuracy of Novice Teachers' Predictions of Math Performance by Percent Minority in Class

% Minority	n	Mean	SD
1) 0 - 10%	13	+.42	.256
2) 11 - 20%	8	+.61	.112
3) 21 - 50%	9	+.61	.114
4) 51 -100%	8	+.47	.243

GLM analysis has shown that differences can be inferred between means.

F (3/34) =2.38, p<.08
Multiple t (LSD) differences between 1 - 2, 1 - 3

Table 10

Mean Correlations and z-tests of Teacher Judgmental Accuracy for Top Versus Bottom Scoring Students *

	Top	Bottom	z-test**
READING			
Experienced Teachers	.55	.48	.411
Novice Teachers	.30	.28	.076
Experienced/ Novice Teachers	.58	.64	-.359
MATH			
Experienced Teachers	.52	.44	.460
Novice Teachers	.44	.39	.414
Experienced/ Novice Teachers	.64	.41	1.280

* All mean correlations were obtained by Fisher's r to Z-transformations.

** The obtained z must exceed 1.96 to be considered significant at the .05 alpha level.

Table 11
Mean Correlations and z-tests for Teacher Judgmental Accuracy for Boys Versus Girls *

	Boys	Girls	z-test
READING			
Experienced Teachers	.71	.70	.144
Novice Teachers	.55	.55	-.031
MATH			
Experienced Teachers	.76	.75	.059
Novice Teachers	.53	.59	-.367

* All mean correlations were obtained by Fisher's r to Z-transformations.
** The obtained z must exceed 1.96 to be considered significant at the .05 alpha level.

Table 12
Correlations between teachers' judgmental accuracy and class size

Teacher Group	Reading	Math
Experienced Teachers	-.04 ($p<.80$)	-.01 ($p<.93$)
Novice Teachers	-.05 ($p<.79$)	+.13 ($p<.44$)

Table 13

Means, Standard Deviations and Correlations Between Teacher Confidence in Rankings and Judgmental Accuracy

Teacher group/subject	Mean	SD	r
Experienced teachers' confidence in judgments of MATH	3.44	.68	.02
Novice teachers' confidence in judgments of MATH	2.25	.81	.16
Experienced teachers' confidence in judgments of READING	3.67	1.02	.16
Novice teachers' confidence in judgments of READING	2.42	.94	.36

Discussion

The purpose of this study was to determine the relations between the accuracy of teachers' predictions of student achievement and students' actual performance, on a standardized measure of achievement. This relation was considered the cornerstone of the study, and was referred to as accuracy of teacher judgments. This measure of accuracy was then used as a variable to assess relations with other variables related to accuracy of teacher judgments of student achievement.

The relations between teacher judgments of student achievement and actual student performance on the standardized achievement test were, as predicted, generally positive and wide in variability. Experienced teachers were significantly more accurate in their judgments than novice teachers. Nevertheless, although one-fourth of the novice teacher predictions were not significantly different from zero, the mean of the correlations demonstrated that novice teachers were remarkably accurate considering the fact that before the ITBS was administered, most of the novice teachers had between 16 to 20 hours of experience with the students they were judging. Novice teacher accuracy could be due, in part, to discussions between novice and experienced teachers, about their students. The degree to which this is responsible for the relatively high degree of accuracy of novice teachers, as opposed to novice teachers' own independent thoughts about students, could not be ascertained in this study.

It is also possible that the novice teachers were as accurate as they were after such a short period of time, because they were knowledgeable about observational and descriptive methods learned in their human development class. This class also made it necessary for the novice teachers to become familiar with their placement teachers' beliefs about students, education and learning, which also may have influenced their judgments.

As other researchers have discovered, accuracy in judgments of reading performance were slightly higher than judgments for mathematics performance. There was also a greater range of correlations for the subject of mathematics than reading among both experienced and novice teacher groups. Although the means don't reflect large differences in accuracy, it appears that both experienced and novice teachers are less adept at judging math performance than reading performance in their students. It may be that judgments about reading are easier for teachers to form than judgments of math because teachers commonly give reading instruction to students grouped by reading ability, whereas math instruction is not generally taught this way.

Teacher Experience

It was expected that there would be a positive relationship between years of experience and accuracy of teacher judgments. The results indicated that among the practicing teachers, there was no relationship between years of experience and judgmental accuracy. The only significant difference based on years of experience was between the experienced and novice teachers, which was expected. Within the group of experienced teachers, it was surprising that some of the very best judges of student achievement were in the early years of their careers. Although the numbers involved were too few to draw any firm conclusions, it is possible that the relationship between years of experience and accuracy is not positive and linear. Further research

is needed to reveal the relations between experience and judgmental accuracy. Specifically, future research could address the issue of experience better, by including a larger sample of beginning teachers. The high accuracy displayed by the beginning teachers in this study may have been a result of a combination of the incentive of feedback and the element of self-selection. Experienced teachers were recruited by the novice teacher placed in their class. Their incentive was to help the novice teacher gain extra credit for a class. Beginning teachers chose to participate so they could find out just how accurate their judgments were. It would be advisable for any future research to ensure that incentives are identical and salient to all participants.

Classroom Ethnic Composition

There was no significant correlation between judgmental accuracy of either Anglo novice or experienced teachers and the percent of minority students in their class. However, using the GLM, a significant difference appeared between the mean accuracy scores of experienced teachers in the topic of mathematics. The only other condition that came close to being significant was the novice teachers' judgments of math. Interestingly, in both cases, the means for classes with 11-20% minority students, and 21-50% minority students were significantly (or almost significantly, in the case of novice teachers) higher than the two extremes of 0-10% and over 50% minority students.

The reason that teachers were more accurate with the middle two groups might be a result of the amount of information available to them about students. Specifically, ethnicity may provide teachers with additional information about a student's achievement. The fact that many of the minority students in grade schools are speaking English as a second language and have historically achieved lower scores on the standardized achievement tests, may provide teachers with extra information which if correct, would increase judgmental accuracy.

This information, useful in increasing overall accuracy for classes with 11-50 % minority students, may be less useful at the extreme ends. If accuracy is increased by teachers knowing the ethnicity of their students, then in an ethnically homogenous classroom (10% minority students or less) such information would not sizably increase the overall accuracy. Likewise, in an ethnically diverse class (in this study classes with over 50% minority students), the power of the information would be diminished because knowing ethnicity would not be unique enough to help the teacher differentiate performance among the students' achievement. On the other hand, in the middle ranges (11-50% minority students) the teachers may be able to use the information about ethnicity and its relation to achievement in such a way that it would raise their overall accuracy of prediction.

Why higher accuracy is shown in classes with between 11% and 50% minority students in math and not reading is not clear either. It is possible that teachers may include ethnicity in their thoughts used to judge achievement in math more so than reading because they have less information to use about math achievement than reading (as evidenced by the slightly lower judgmental accuracy in math than reading). So any additional information about students that is known might be used. Also, if many of the minority students are not English proficient, it may be more difficult for teachers to convey and monitor the comprehension of mathematical concepts.

The main limitation of this analysis is in the distribution of the sample of minority representation in classrooms. The sample was heavily skewed, which is why

the percent minority grouping number 4 ranged from 51-100%. There simply were not enough classes, in this sample, that were high in ethnic diversity to assess teacher judgmental accuracy reliably. Also for each analysis, each group was small (between 6 to 13 students). Thus, the means would be easily altered by one or two unusual classes.

Student Ability

The average correlations for teacher accuracy for the top and bottom scoring thirds of class seemed to demonstrate that teachers are more accurate in judging performance for their top scoring students. The z-tests that were conducted in this study failed to show that the differences in teacher accuracy for top versus bottom scoring students was significant. It is not surprising that significant differences were not found because of the small sample sizes on which the correlations for this analysis were based. However, these results were consistent with those obtained by other researchers, who have investigated the issue of differences in teacher judgmental accuracy according to ability of students.

Student Gender

Average correlations of teacher accuracy for boys and girls were quite similar. Tests of significance failed to indicate that teachers were differentially accurate for boys and girls. As with other studies that have examined the role of student gender in the accuracy of teacher judgments, it may be concluded that gender does not play a significant role in teachers' judgmental accuracy.

Class Size

The prediction that class size would be negatively correlated with teacher judgmental accuracy was not confirmed. This was surprising since we hypothesized that the larger the class size, the more difficult it would be for teachers to learn about individual student ability. But it is possible that the correlations obtained were low because the range of class was somewhat restricted. About 70 percent of all classes had between 20 and 25 students. There were very few classes on the extreme ends, and this could have underestimated the correlations.

Teacher Confidence in Judgments

There was no relation between ratings of confidence and the accuracy of novice and experienced teachers' rankings. Hoge & Butcher (1984) found a similar pattern in their research on teacher judgments of student achievement. Surprisingly, the strength of the relationship between confidence and accuracy was stronger for the novice teachers than for the experienced teachers. Experienced teachers, on average, rated their confidence one point higher than the novice teachers, but both groups tended to keep their ratings toward the middle of the scale and avoid the extreme ends, as typically happens with Likert-type scales.

Teacher Differentiation of Math and Reading Ability

The moderate negative correlation between the average experienced teacher accuracy and the average differentiation between teacher judgments of student reading and math performance was not expected. Teachers who were highly accurate in their judgments did not differentiate between student math and reading ability to the same extent that less accurate teachers did. It appears that the accurate teachers were using general student ability to make their judgments.

It should be noted that this correlation only addresses the question of the differences teachers perceive in their students' reading and math ability. It does not take into account the actual differences found in the students' ITBS rankings, or the direction of that difference.

Limitations of Study

Aside from the problematic issues that have been raised above, there was another factor which may have contributed to the results that were obtained. This factor was the small number of subjects involved. Although the total number of subjects seemed to be just enough to do the study, when subjects were categorized by certain variables, the numbers dropped enough to minimize the possibility of significant results.

Implications of Study and Further Research

No evidence was found in this study to indicate that experience mediates teachers' judgmental accuracy. The wide range of accuracy of teacher judgments found in this study is consistent with much other research on this topic. These results say something about what can be expected from teachers and what kind of judgmental accuracy appears to be common.

As with any skill, wide variation exists in judgmental accuracy among teachers. It is encouraging that approximately two-thirds of the experienced teachers' judgments of student achievement correlated with actual student performance at .70 or higher. However, the remaining third were not very accurate. What qualities separate the very accurate teacher from the very inaccurate teacher is certainly a question worthy of further exploration. Teachers who cannot, somewhat accurately, judge student class ranking in a subject or level of mastery over a content area, are more likely to make erroneous instructional decisions for students.

Future research could focus on the beliefs held by accurate and inaccurate judges of student achievement, such as teachers' beliefs about student performance, ability and motivation, beliefs about what achievement tests measure and other teacher characteristics and variables within the classroom.

REFERENCES

Borko, H., Cone, R., Russo, N.A., & Shavelson, R.J. (1979). Teachers' decision making. In P.L. Peterson & H.J. Walberg (Eds.), Research on teaching: Concepts, findings, and implications (pp. 136-160). Berkeley, CA: McCutchan.

Coladarci, T. (1986). Accuracy of teacher judgments of student responses to standardized test items. Journal of Educational Psychology, 78, 141-146.

Farr, R., & Roelke, P. (1971). Measuring subskills of reading: Intercorrelations between standardized reading tests, teachers' ratings, and reading specialists' ratings. Journal of Educational Measurement, 8, 27-32.

Gage, N.L., & Berliner, D.C. (1988). Educational psychology. Boston: Houghton Mifflin.

Glass, G.V. & Hopkins, K.D. (1984). Statistical methods in education and psychology. New Jersey: Prentice-Hall, Inc.

Helmke, A., & Schrader, F-W. (1987). Interactional effects of instructional quality and teacher judgmental accuracy on achievement. Teaching and Teacher Education, 3, 91-98.

Helmstadter, G.C. (1970). Research concepts in human behavior: Education, psychology, sociology. New York: Meredith.

Hoge, R., & Butcher, R. (1984). Analysis of teacher judgments of pupil achievement levels. Journal of Educational Psychology, 76, 777-781.

Hoge, R., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. Review of Educational Research, 59, 297-313.

Hopkins, K.D., George, C.A., & Williams, D.D. (1985). The concurrent validity of standardized achievement tests by content area using teachers' ratings as criteria. Journal of Educational Measurement, 22, 177-182.