Teacher Expectations II: Construction and Reflection of Student Achievement

Lee Jussim
Rutgers University

Jacquelynne S. Eccles
University of Colorado

Hypotheses regarding self-fulfilling prophecies, perceptual biases, and accuracy were tested using longitudinal data relating 98 6th-grade math teachers' expectations to 1,731 students' performance. Consistent with the self-fulfilling prophecy hypothesis, teacher expectations predicted changes in student achievement beyond effects accounted for by previous achievement and motivation. Consistent with the perceptual bias hypothesis, teacher expectations predicted their own evaluations of students' performance more strongly than they predicted standardized test scores. Consistent with the accuracy hypothesis, path coefficients relating teacher expectations to standardized achievement tests were about 80% lower than zero-order correlations, and the path coefficients relating teacher expectations to students' grades were 45% to 65% lower than the zero-order correlations. These results support a weak constructivist perspective.

How are social perception and social reality related? Strong social constructivist perspectives emphasize the power of beliefs to create reality (see Jussim, 1991, for a review). According to such perspectives, for example, children become what significant others, such as parents and teachers, expect them to become. But is social reality so malleable that, because of phenomena such as self-fulfilling prophecies and perceptual biases, it is readily transformed by erroneous social beliefs?

There are few contexts more important for investigating these basic issues than teachers' expectations for their students. Teachers develop expectations for students early in the year (Brophy, 1983; Rist, 1970), and students do indeed often confirm these expectations (Brophy & Good, 1974; Crano & Mellon, 1978; Humphreys & Stubbs, 1977; Williams, 1976). Experimental research has often examined how induction of erroneous teacher expectations influences teacher behavior and student performance (e.g., Rosenthal & Jacobson, 1968; Sutherland & Goldschmid, 1974; see Harris & Rosenthal, 1985; Raudenbush, 1984, for meta-analyses; Jussim, 1986, 1991, for reviews). This research has shown that teacher expectations, in principle, may influence students. Because such research was not even intended to address the accuracy of teacher expectations and because accuracy limits self-fulfilling prophecies (Jussim, 1989, 1990, 1991), such research provides little information regarding the extent to which naturally developed teacher expectations influence students' achievement.

Therefore, our research redressed this limitation by focusing on the extent to which naturally occurring teacher expectations influence student achievement. Among a sample of 98 teachers and over 1,700 students in sixth-grade math classes, we assessed the extent to which teacher expectations (a) were accurate, (b) changed students' achievement through self-fulfilling prophecies, and (c) biased teachers' evaluations of students' achievement.

Overview of a Controversy

How much does social perception construct or create social reality? This issue has generated considerable controversy. Experimental studies have demonstrated many errors and biases in social perception and have also documented a myriad of ways in which social perception constructs social reality (see reviews by Hamilton, Sherman, & Ruvolo, 1990; E. E. Jones, 1986; Snyder, 1984). Consequently, the experimental research has often been interpreted as consistent with a strong social constructivist perspective that paints a grim picture: People's expectations are allegedly highly inaccurate, likely to bias interpretations of others' actions, and likely to create self-fulfilling prophecies. As such, expectancies are often accused of playing a major role in the perpetuation of social injustices associated with stereotypes, prejudice, and inequalities in educational and occupational opportunity (e.g., Fiske & Taylor, 1991; Hamilton et al., 1990; Snyder, 1984).

This depiction of the strong constructivist perspective is no straw man; the strong constructivist perspective has a long history and has had widespread influence on many areas of psychology (see Jussim, 1991, for a review). Consider the following

This research was performed while Lee Jussim was supported by a National Academy of Education Spencer Foundation postdoctoral fellowship. It was also partially supported by a Rutgers Research Council grant. We gratefully acknowledge the insightful comments provided by the reviewers regarding a draft.

Correspondence concerning this article should be addressed to Lee Jussim, Department of Psychology, Tillett Hall, Rutgers University, New Brunswick, New Jersey 08903.

Journal of Personality and Social Psychology, 1992, Vol. 63, No. 6, 947–961

quotes: "Events in the social world may be as much *effects* of individuals' beliefs as they are *causes* of these beliefs" (Snyder, 1984, p. 294, emphasis in original) and "Teachers' expectancies influence students' academic performance to a greater degree than students' performance influences teachers' expectancies" (Miller & Turnbull, 1986, p. 236).

Many social–psychological perspectives claim or imply that the power of beliefs to create reality equals or exceeds the power of reality to influence people's beliefs (Darley, Fleming, Hilton, & Swann, 1988; Fiske & Taylor, 1991; Gage & Cronbach, 1955; Hamilton et al., 1990; Jones, 1986, 1990). Educational psychologists, in contrast, have reached diametrically opposed conclusions. Rather than being major agents in the construction of student achievement, teacher expectations accurately reflect student achievement, according to many educational psychologists (e.g., Brophy, 1983; Dusek, 1975; Meyer, 1985; Raudenbush, 1984; West & Anderson, 1976).

These perspectives—that teacher expectations are highly inaccurate and have powerfully self-fulfilling effects versus that teacher expectations are highly accurate and have highly limited self-fulfilling effects—are mutually exclusive. As broad generalizations, both cannot simultaneously be true. The current study, therefore, addresses the following question: How much do teacher expectations predict student achievement by creating self-fulfilling prophecies and perceptual biases rather than by being accurate?

How Much Do Teacher Expectations Create Versus Reflect Student Achievement?

Teachers' expectations may be confirmed because they lead to self-fulfilling prophecies or perceptual biases or because these expectations were accurate. Unfortunately, however, simple correlations among teacher expectations and student achievement provide no basis for distinguishing among these three sources of expectancy confirmation (Jussim, 1989, 1991). How, then, can they be distinguished under naturalistic conditions?

Our conceptual rationale and specific procedures for disentangling these three sources of expectancy confirmation under naturalistic conditions has been extensively described elsewhere (Jussim, 1989, 1991). Therefore, we only briefly summarize that rationale. Teachers' expectations must change students' performance for expectancy–behavior associations to be interpretable as evidence of self-fulfilling prophecies. The occurrence of a perceptual bias means that teachers view students as performing at levels more consistent with their expectations than is warranted on the basis of students' achievement. Assessing the accuracy of teacher's predictions involves determining the extent to which expectations predict, without causing, students' performance.

Jussim (1989) provided evidence for the occurrence of small self-fulfilling prophecies and perceptual bias effects and also showed that teachers' expectations predicted students' future achievement mainly because those expectations were accurate. Jussim's 1989 study remains the only study that has empirically distinguished and assessed self-fulfilling prophecies, perceptual biases, and accuracy under naturalistic conditions.[1] Consequently, the results should be viewed cautiously pending replication, especially because the evidence of high accuracy seems inconsistent with the strong constructivist perspective.

Overview of This Study

Purpose

The current study replicates and extends the Jussim (1989) study in several ways. First, our total sample includes a much larger number of students (1,731) and teachers (98) in sixth-grade math classes. Second, the current study includes data on students and teachers from nine different school districts in southeastern Michigan (rather than in the single district examined in Jussim, 1989). Third, the demographic characteristics of these districts are highly variable. The socioeconomic status of the districts ranges from working class to upper middle class, and the sample includes a substantial number of minority students (about 10% of the students). Because it includes a relatively large number of teachers who teach a wide variety of students in a variety of public school settings, we believe results from this study should be broadly generalizable.

Another purpose of the current study is to compare results obtained in Jussim (1989) with those obtained using a more diverse sample. Therefore, the current sample is divided into two groups: (a) the sample used in the Jussim (1989) study and (b) 1,288 students and 71 teachers in eight other school districts. Two-group LISREL analyses (Jöreskog & Sörbom, 1983) are used to compare the results obtained for these two samples.

A third purpose is to explore the impact of student sex on teachers' perceptions of students' talent and effort. Research consistently shows that (a) by junior high school, boys perform somewhat higher than girls on standardized math tests; (b) before junior high school, boys and girls perform similarly on standardized math tests; and (c) regardless of grade level, girls tend to receive higher math grades than boys (see Kimball, 1989, for a review). Some research suggests that student sex influences teacher perceptions and leads to differential treatment (e.g., Brophy & Good, 1974; Dweck, Davidson, Nelson, & Enna, 1978; Eccles & Blumenfeld, 1985; Eccles & Wigfield, 1985; Parsons, Kaczala, & Meece, 1982; Wilkinson & Marrett, 1985). Especially in math, teachers may believe that boys have more talent than girls, despite girls receiving higher grades in elementary and junior high school (Eccles & Blumenfeld, 1985; Parsons et al., 1982). Consistent with this perspective, teachers are more likely to attribute boys' success to ability and girls' success to effort (Dweck et al., 1978). This suggests that teachers would have biased perceptions of boys' and girls' talent and effort. Previous research, however, has not been focused on testing this hypothesis. Instead, most studies have investigated whether teachers treat boys differently from girls or make dif-

[1] There are some partial exceptions. One is a dissertation by Berman that disentangled predictive accuracy from self-fulfilling in therapist–patient relationships. West and Anderson (1976) also attempted to distinguish accuracy from self-fulfilling prophecy, but (a) did not address predictive accuracy (they only addressed accuracy in terms of the basis for teacher expectations) and (b) also ignored perceptual biases.

ferent attributions for the performance of boys and girls (e.g., Dweck et al., 1978; Parsons et al., 1982). Thus, there is little direct evidence indicating the extent to which teachers' perceptions of boys' and girls' effort and talent are biased. Providing such evidence, then, is another goal of this study. We predicted that teachers would rate girls as performing better and exerting more effort than boys and that teachers would rate boys as having more talent than girls.

Hypotheses

Because this study uses naturalistic data, hypotheses are phrased in terms of some variables *predicting* rather than *causing* other variables. The self-fulfilling prophecy hypothesis suggests that teachers' expectations predict students' future achievement, even after controlling for students' prior achievement. The perceptual bias hypothesis suggests that teacher expectations predict their own judgments of students' achievement (i.e., grades) more than they predict independent assessments of students' achievement (standardized test scores). There are two sets of accuracy hypotheses: (a) Students' prior achievement and motivation predict teachers' expectations and (b) teachers' expectations correlate with students' future achievement at least partially because they are based on valid predictors of student achievement; controlling for those valid predictors substantially reduces the correlation of teacher expectations with student achievement.

Method

Sample

All analyses in this study are based on data obtained as part of a longitudinal study of adolescent development: the Michigan Study of Adolescents' Life Transitions (Eccles, 1988). The total sixth-grade sample included 108 teachers and 2,625 students in math classes in 11 school districts in southeastern Michigan. Analyses required, however, that students attend these school districts for 3 consecutive years —in fifth, sixth, and seventh grades. The 531 students who did not attend any of the schools in these districts during at least 1 of these 3 years were excluded from the analyses reported in this article.

In addition, some schools did not require students to take a standardized math test in late fifth grade or early sixth grade. The additional 283 students without such information regarding past levels of achievement also were excluded from the main analyses. Initial analyses showed that there were few differences between the included and excluded students on any of the teacher expectation, student motivation, or student performance variables. All analyses reported below were based on 27 teachers and 443 students in one school district (hence referred to as Sample 1) and 71 teachers and 1,288 students in eight other school districts (hence referred to as Sample 2).[2]

Questionnaires

Teachers evaluated each student in their class on a variety of dimensions early in the school year (generally around the first week of October). Included in the current study were teachers' assessments of each students' talent, effort, and performance in math—these are the three teacher expectation variables used in the current research. Questionnaires assessed students' beliefs, perceptions, and feelings in a variety of domains. This study included students' self-concept of ability in math, their self-reported effort in math, the time they spend on math homework, and the value they place on math.[3] All measures are valid and reliable (they are presented in the Appendix; for reliability and validity information, see Eccles (Parsons) et al., 1984; Jussim, 1987; Parsons, 1980).

Measures of Student Achievement

Several measures of student achievement were used. There were two measures of past achievement: final marks in fifth-grade math classes and scores on the math section of standardized tests taken in late fifth grade or early sixth grade. Included in each student's record was the percentile ranking of their performance on standardized tests, relative to national norms. These percentile rank scores were used to translate different standardized test scores onto the same scale.

Two measures of final achievement were used: final marks in sixth-grade math classes and scores on the math section of the seventh-grade Michigan Educational Assessment Program (MEAP). The MEAP has been widely used since the early 1970s and is administered in early October to all Michigan public school students in seventh grade. An overview of all measures used in this study is presented in Table 1.

Results and Discussion

Analysis Overview

All models were assessed using the LISREL VI program (Jöreskog & Sörbom, 1983), using covariance matrices as input. Correlations and standard deviations for all variables are presented in Table 2. All path coefficients may be interpreted identically to standardized betas from a regression equation.

Analyses are reported in three stages. In the first stage, results from estimating separate LISREL models for Sample 1 and for Sample 2 are reported. In the second stage, we used LISREL to assess directly the extent to which the Sample 2 results replicated those obtained for Sample 1. In the third stage, whether conceptualizing teacher expectations in different ways led to different patterns of results was explored.

Stage 1: Separate Model Assessment for the Two Samples

Model Assumptions

Analyses allowed (a) all student-background variables (previous grades and standardized test scores, self-concept of ability, self-perceptions of effort, time spent on homework, and gender) to predict the three teacher-perception variables and (b) all student background variables and teacher-perception

[2] The current study did not assess student motivational mediation of self-fulfilling prophecies, because little evidence of such mediation was found in Jussim (1989). Therefore, in contrast to Jussim (1989), the current study did not include student motivational variables assessed during the spring in any analyses. Because there were fewer variables, there were also fewer cases with missing data in Sample 1 than in the original Jussim (1989) study. Therefore, the total size for Sample 1 is 443, whereas it was 429 in Jussim (1989).

[3] Initial analyses revealed that neither intrinsic nor extrinsic value placed on math was involved in the teacher-expectation process in any way among either sample. Consequently, these variables are not discussed further.

Table 1
Conceptual Variables and Operational Measures

Past performance	Motivation	Teachers' expectations	Future achievement
Standardized math test scores[a]	Self-concept of math ability[b,c]	Perceptions of math performance[d]	MEAP math scores[e]
Final math marks in fifth grade	Effort in math[c]	Perceptions of math talent[d]	Final math marks in sixth grade
	Time spent on math homework[c]	Perceptions of effort in math[d]	
	Intrinsic and extrinsic value of math[c]		

[a] Different districts administered different tests in late fifth grade or early sixth grade. Percentile ranks were used to render the tests comparable. [b] Self-concept of ability was measured by a scale consisting of two items: how good students think they are at math and how good they think they are in comparison with other students. [c] The motivational variables were assessed in early October. [d] All teacher expectation variables were assessed in October, within a few days after the assessment of student motivation. [e] The MEAP (Michigan Educational Assessment Program) is a standardized test taken in October of seventh grade.

variables to predict students' final grades in sixth-grade math classes and MEAP scores.[4] In addition, the model assessed in Stage 1 assumed that teacher perceptions of student effort and talent could be based on teacher perceptions of student performance. Because they are difficult to observe directly, one person's judgments of another's effort and ability must depend on observable behaviors believed to be related to effort and ability. Performance is one such directly observable behavior (e.g., Heider, 1958; Nicholls, 1979; Weiner, 1979).

In this set of analyses, self-concept of ability was the only variable assessed with two measures: How good students felt they were at math and how they ranked themselves in comparison with other students in their math class. All analyses assume, therefore, that self-concept of ability is a latent variable with these two indicators. All other variables had only single indicators.[5] Therefore, measurement error is estimated and removed only for self-concept of ability. The use of single items for most variables represents an important limitation to this research: The extent to which measurement error affected the results is largely unknown.[6]

Organization and Presentation of Results

Because results involve 11 variables and 24 paths, results are depicted in two separate figures. However, this is to simplify presentation; results were obtained by assessing the full model (described in the previous section on Model Assumptions). Because of the large sample, some trivially small coefficients were statistically significant. Therefore, paths are shown only when the coefficient was at least .07 in at least one of the two samples. In the figures and in the subsequent text of this article, except where otherwise indicated, all reported paths are significant at $p < .05$. All figures display path coefficients and multiple correlation coefficients for both samples. Sample 1 coefficients are presented first, followed by a slash, followed by Sample 2 coefficients (e.g., in Figure 1, the .28/.17 coefficient relating standardized test scores to teacher perceptions of talent means that the path coefficient was .28 for Sample 1 and .17 for Sample 2). Similarly, in the text below, when we present pairs of coefficients, the first represents the result for Sample 1 and the second represents the result for Sample 2.

The results for Stage 1 are presented in four sections: (a) assessment of the bases of teacher expectations, (b) assessment of the self-fulfilling prophecy hypothesis; (c) assessment of the perceptual bias hypotheses; and (d) assessment of the extent to which teacher perceptions predicted, without influencing, students' grades and MEAP scores.

Bases of Teacher Expectations

Teacher perceptions of performance. Results assessing the bases of teacher perceptions are depicted in Figure 1. In both samples, results showed that teacher perceptions of student performance are largely based on appropriate factors. The main

[4] Because there are many students in each class, data points are not independent of one another. However, the path coefficients obtained in analyses controlling for teacher were virtually identical to those we present in this article.

[5] It might seem that students' self-perceptions of effort and time spent on homework could also be considered to be manifestations of an underlying "self-perceptions of effort" factor. However, because the correlations between these variables were quite low for both samples (see Table 2), we did not feel that it was appropriate to consider them as manifestations of a single underlying factor.

[6] However, although measurement error could not be estimated, we did test numerous models making differing assumptions regarding measurement error in the variables. For example, models assuming reliabilities of .7, .8, and .9 in each of the teacher-perception variables were tested. The model assuming reliabilities of .7 yielded some multiple correlation coefficients greater than 1.0, thereby indicating that .7 underestimated the reliability of the teacher-perception variables. Models assuming reliabilities of .8 or greater yielded results that were very similar to those reported in this article.

Table 2
Correlations Among the Variables

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1. Gender	0.500, 0.499	.088	.077	−.039	.005	−.083	−.027	.040	−.155	−.074	−.062	−.008
2. How good at math	.098	1.326, 1.450	.632	−.109	.258	.498	.452	.451	.395	.471	.399	.354
3. Rank self at math	.073	.750	1.231, 1.315	−.147	.270	.344	.336	.357	.286	.383	.295	.276
4. Time spent on homework	−.007	−.014	−.018	0.886, 0.866	.102	−.145	−.152	−.155	−.083	−.181	−.110	−.125
5. Self-perceptions of effort	.003	.204	.203	.161	1.320, 1.313	.192	.065	.122	.203	.170	.118	.088
6. Sixth-grade final marks	−.031	.459	.348	−.050	.075	2.151, 2.686	.632	.636	.632	.706	.674	.533
7. Standardized test scores	−.022	.420	.324	−.121	.019	.604	24.724, 24.848	.638	.452	.638	.585	.663
8. Teacher perceptions of talent	.037	.446	.350	−.112	.008	.586	.631	1.298, 1.327	.622	.799	.526	.493
9. Teacher perceptions of effort	−.180	.307	.271	−.089	.126	.514	.420	.494	1.272, 1.411	.732	.453	.362
10. Teacher perceptions of performance	−.044	.449	.402	−.027	.086	.559	.509	.732	.664	1.042, 1.143	.557	.538
11. Fifth-grade final marks	−.029	.415	.311	−.079	.127	.670	.673	.545	.433	.495	2.219, 2.536	.478
12. MEAP	−.035	.357	.282	−.072	.075	.573	.705	.571	.357	.460	.621	3.918, 4.841

Note. Correlations below the diagonal are for Sample 1; correlations above the diagonal are for Sample 2. The standard deviations are presented on the diagonal, with standard deviations for Sample 1 above those for Sample 2. Correlations above .09 and .06 for Samples 1 and 2, respectively, are significant at $p < .05$. MEAP = Michigan Educational Assessment Program.

predictors of teacher perceptions of students' performance are students' previous standardized test scores (βs = .24 and .39), final marks in fifth-grade math classes (βs = .21 and .21), and students' self-concept of math ability (βs = .29 and .23; see Figure 1).

As predicted, teachers perceived girls as performing better than boys; although this relation is not significant in Sample 1 ($\beta = -.06$), it is significant in Sample 2 ($\beta = -.10$, $p < .001$). When the two samples are combined, this relationship is significant ($\beta = -.08$, $p < .001$). This result obtained despite there being no differences between boys and girls on either their previous standard test scores or their MEAP scores (all *r*s ranged from −.04 to .03, *ns*). Thus, this result does not seem to represent accuracy; teachers perceived girls as performing more highly than boys despite the absence of objective evidence (in terms of standardized test scores) supporting this perception.

Teacher perceptions of talent. Teacher perceptions of talent are also largely based on appropriate factors (see Figure 1). The two largest predictors are teachers' own perceptions of students' performance (βs = .53 and .64) and previous standardized test scores (βs = .28 and .17). There also are small effects of previous grades (βs = .07 and .05, *p*s < .1 and .05, respectively). These results are consistent with much educational research suggesting that teacher expectations are usually based on valid information (Brophy & Good, 1974; West & Anderson, 1976; see reviews by Brophy, 1983; Cooper, 1979; Dusek, 1975).

As predicted, teachers believe boys have more talent than girls (βs = .06 and .08, both *p*s < .05). This result does not seem to represent accuracy because it is inconsistent with the lack of difference between boys' and girls' standardized test scores and teachers' own beliefs that girls perform more highly than boys.

Teacher perceptions of effort. Teacher perceptions of effort are only minimally based on the two student effort variables (for self-perceptions of effort, βs = .08 and .06; for time spent on homework, βs = −.07 and .04). The largest influence on perceptions of effort is teachers' own perceptions of students' performance (βs = .58 and .68, see Figure 1). To a small extent, teacher perceptions of effort also are related to final marks from fifth grade (βs = .10 and .06). It seems, therefore, that teachers assume that students who perform better are exerting more effort and that teachers were nearly oblivious to how hard students think they are trying.

Although it may be appropriate for teachers to base their perceptions of students' effort most heavily on their own perceptions of students' performance, our data and other research show that the relation between effort and performance often is weak or nonexistent (Schuman, Walsh, Olson, & Etheridge, 1985). Students' self-perceptions of effort correlated .07 (*ns*) and .09 with MEAP scores and .07 (*ns*) and .19 with final grades in math. Time reportedly spent on homework, however, correlated negatively with performance: −.07 (*ns*) and −.13 with MEAP scores and −.05 (*ns*) and −.15 with grades. Thus, the

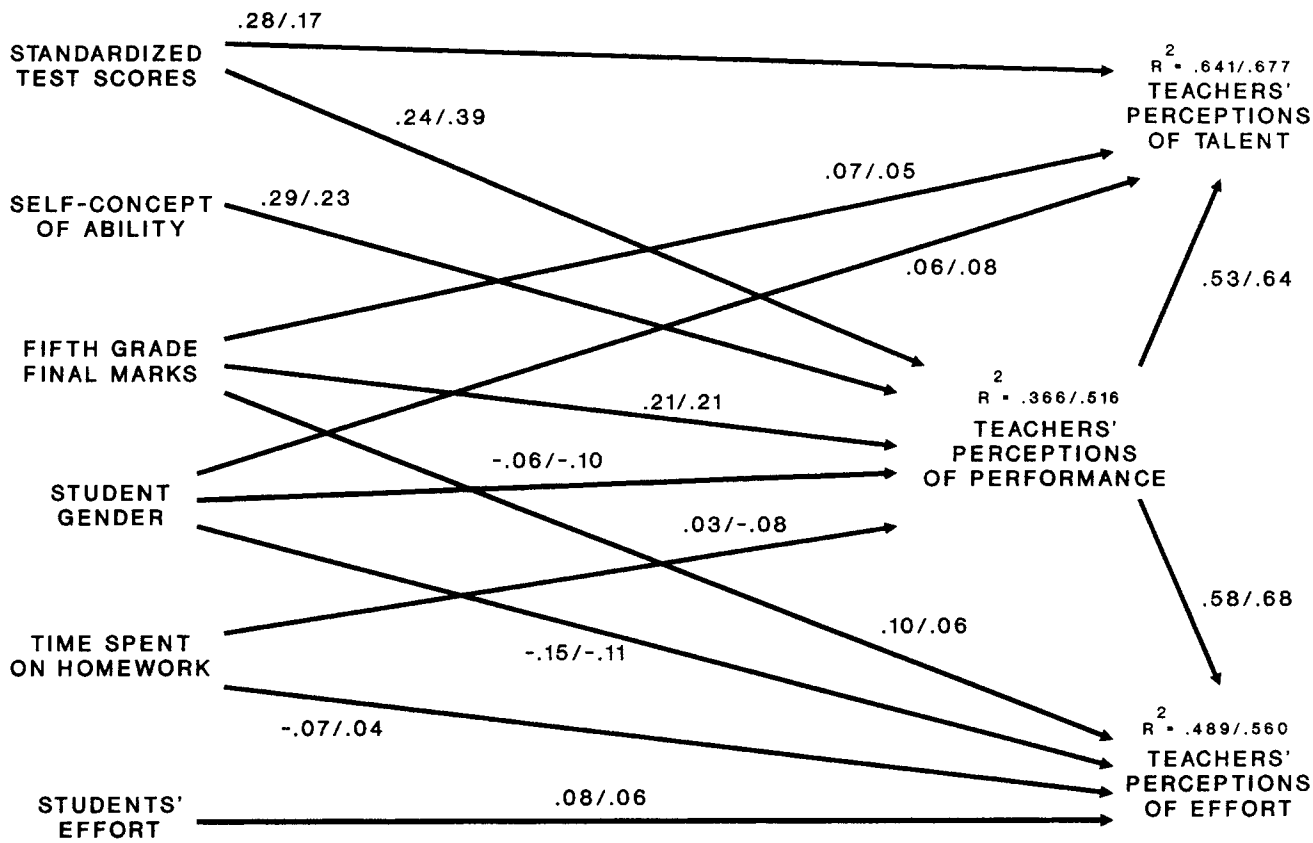

Figure 1. Bases of teacher perceptions. (Paths are shown only when the coefficient was at least .07 in at least one sample. For each path, two coefficients are shown: The first shows the relation for Sample 1 and the second shows the relation for Sample 2. All coefficients are standardized. All coefficients are significant at $p < .05$, except for the .07 coefficient for fifth-grade final marks, the −.06 coefficient for student gender, and the .03 coefficient for time spent on homework.)

relation between effort and performance is neither as powerful nor as direct as many teachers seem to believe.

In addition, as predicted, teachers rated girls as trying harder than boys (βs = −.15 and −.11). Because student gender was unrelated to either self-perceptions of effort or time spent on homework in either sample (*r*s ranged from −.04 to .01), the effect of student sex on teacher perceptions of effort again seems to represent a bias rather than accuracy.

Self-Fulfilling Prophecies

If teacher expectations lead to self-fulfilling prophecies, they should predict changes in student achievement. Figure 2 summarizes the results obtained regarding final marks and MEAP scores, respectively. Results consistent with the self-fulfilling prophecy hypothesis were obtained for both samples. In Sample 1, the clearest evidence of self-fulfilling prophecies was obtained regarding teacher perceptions of talent, which predicted both final marks (β = .13) and MEAP scores (β = .18). Although teacher perceptions of performance had no direct effects on either outcome, it had significant indirect effects (as mediated by teacher perceptions of talent and effort) on both grades (β = .17) and MEAP scores (β = .10).[7]

Results consistent with the self-fulfilling prophecy hypothesis were also obtained in Sample 2 (see Figure 2). Teacher per-

[7] Variables can have direct effects and indirect effects on each other. Direct effects refer to unmediated impact of one variable on another and are represented by path coefficients. Sometimes, however, all or part of one variable's impact on a second variable will be mediated by a third variable. We assumed that teachers' perceptions of students' effort and talent may be based on their own perceptions of students' performance (see the Conceptualizations of Teacher Expectations section). Therefore, in addition to its direct effects on grades and MEAP scores, teacher perceptions of performance may have indirect effects as mediated by teacher perceptions of effort and talent. In general, the indirect effect of variable *a* on variable *c* as mediated by variable *b* equals the path coefficient linking *a* to *b*, multiplied by the path coefficient linking *b* to *c*. The total effect of a variable is the sum of its direct and indirect effects. Relations among teacher perceptions of performance, talent, and effort are presented in the section on accuracy. For a more detailed discussion of direct, indirect, and total effects, see, for example, Alwin and Hauser (1975), Duncan (1975), or Pedhazur (1982).

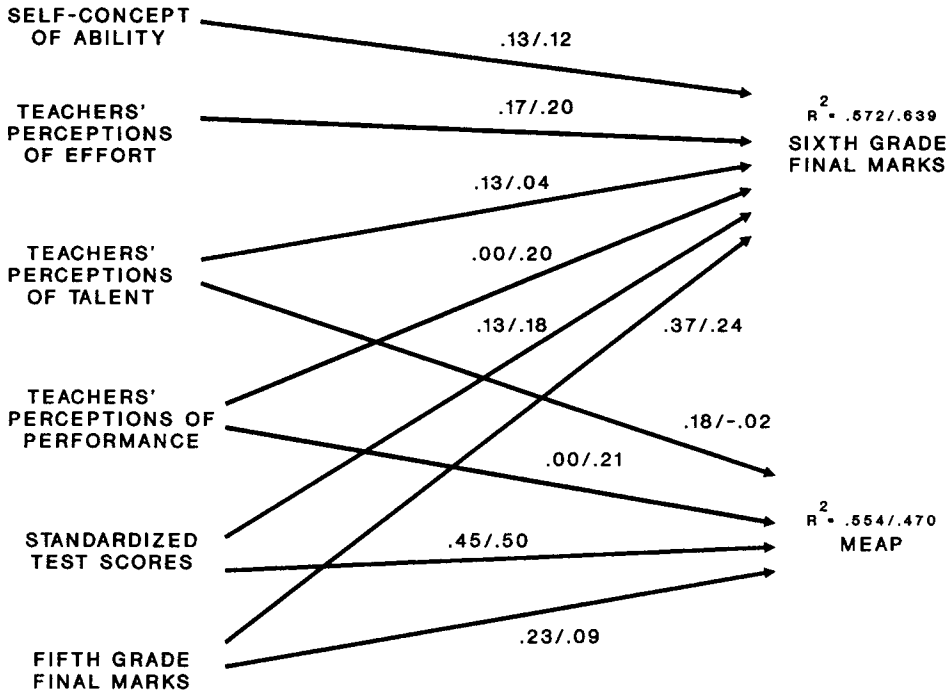

Figure 2. Predictors of students' final grades and MEAP scores. (Paths are shown only when the coefficient was at least .07 in at least one sample. For each path, two coefficients are shown: The first shows the relation for Sample 1 and the second shows the relation for Sample 2. MEAP refers to scores on the math section of the Michigan Educational Assessment Program. All coefficients are standardized. All coefficients greater than .08 are significant at $p < .05$.)

ceptions of performance significantly related directly both to final marks in sixth grade ($\beta = .21$) and MEAP scores ($\beta = .20$). In addition, teacher perceptions of performance indirectly related to final grades (through their relation to teacher perceptions of effort), so that the total effect of teacher perceptions of performance on grades was .34. Because teacher perceptions of talent ($\beta = -.02$, *ns*) and effort ($\beta = -.067$) had small negative relations with MEAP scores, the total effect of teacher perceptions of talent on MEAP scores was actually somewhat smaller than its direct effect ($\beta = .15$). Neither of the other teacher-perception variables significantly related to both grades and MEAP scores.

In both samples, therefore, results showed that the more favorably teachers viewed students in October, the more those students increased their grades and standardized test scores above the levels predicted by prior performance and motivation. The same pattern was also found for teacher perceptions of students' math talent in Sample 1.

Because the teacher expectation variables correlated substantially with one another (.6 to .8), these analyses have one potential limitation. Correlations among predictors may lead structural equation techniques to underestimate their independent effects (e.g., Gordon, 1968; Pedhazur, 1982). Consequently, a second set of analyses was performed to identify the extent to which the three teacher-expectation variables in combination predicted changes in student achievement. First, a model was estimated that included only students' past achievement and motivation as predictors of their future grades and MEAP scores. The multiple correlation coefficients in this analysis were then compared with the multiple correlation coefficients obtained in the first models that also included the teacher expectation variables as predictors. The relation of the three teacher-expectation variables to student achievement in this analysis is represented by the multiple semipartial correlation (Pedhazur, 1982), which is simply the square root of the increment in the multiple correlation (this renders its interpretation similar to that of the path coefficients obtained in the earlier analyses). In these analyses, these relations are .13 and .12 for MEAP scores; they are .23 and .29 for final grades. These results are similar to those obtained in models assessing the independent effects of each of the three teacher-expectation variables because (a) both showed significant effects of teacher expectations on student achievement, (b) the effect on MEAP scores was small, and (c) the effect on grades was greater than the effect on MEAP scores.

Self-Fulfilling Prophecies: Limitations

The correlational nature of the study leaves open alternative explanations. Although a reverse causal direction is not plausible (e.g., final marks in sixth grade did not cause teacher expectations at the beginning of sixth grade), accuracy cannot be eliminated conclusively as an alternative to the self-fulfilling prophecy interpretation. Perhaps teachers used some type or

types of information not included in this study that enabled them to predict student performance beyond levels accounted for by students' prior grades, standardized test scores, self-concept of ability, time spent on homework, effort, and intrinsic and extrinsic value placed on math.

Although the accuracy explanation cannot be conclusively eliminated, several factors limit its viability. First, these results are consistent with a long history of laboratory research, field experiments, and naturalistic studies of interpersonal expectations (e.g., Berman, 1979; Rosenthal & Jacobson, 1968; Snyder, Tanke, & Berscheid, 1977; West & Anderson, 1976; see reviews by Jussim, 1991; Miller & Turnbull, 1986). The modest effect sizes obtained are also consistent with those obtained in nearly all previous studies of the effects of naturally developed teacher expectations (Brattesani, Weinstein, & Marshall, 1984; West & Anderson, 1976; Williams, 1976).

Second, this study included controls that were more complete than those used in nearly all previous research on naturally developed teacher expectations. Few studies have used both past grades and standardized test scores as controls, few have used student motivation as a control, and none have included as broad a variety of student motivational factors as included here (with the exception of Jussim, 1989). A strength of this approach is that by virtue of controlling for students' previous performance (in terms of grades and standardized test scores), significant coefficients mean that teacher expectations predicted changes in student achievement. Furthermore, such predictive validity cannot have resulted from teachers accurately assessing students' motivation (at least in terms of any of the motivational constructs included in the analyses).

Furthermore, no matter how many potential sources of accuracy we assessed, if teacher perceptions predicted changes in student achievement, it would still be possible that "we just did not get the right ones." This is a fundamental limitation to this type of naturalistic study: The obtained path coefficients represent causal effects only if all relevant causes of outcome variables have been included in the model. And no matter how many potential sources of spurious relations one assessed, one can never know whether one assessed them all. Therefore, such research can never conclusively demonstrate causal relationships.

Nonetheless, an effort can be made to include as many potential sources of spurious relations as possible. Because of the large number of such variables included in the current study, we conclude that it provides some of the clearest evidence to date that naturally occurring teacher expectations led to self-fulfilling prophecies. Such a conclusion will warrant revision when future research demonstrates empirically that there are important sources of accuracy in teacher perceptions other than those we assessed.

Perceptual Biases

If teacher expectations lead to perceptual biases, they should more strongly relate to grades than to standardized test scores (see Jussim, 1989, 1991, for more extended discussions of this issue). In both samples, the clearest evidence of a perceptual bias occurred for teacher perceptions of effort (see Figure 2). In both samples, teacher perceptions of effort predicted final grades (βs = .17 and .20) but had little relation to MEAP scores (βs = −.03 and −.067).

Results consistent with a perceptual bias also occurred for teacher perceptions of performance (total effects on final grades = .17 and .34; total effects on MEAP scores = .10 and .15), although this pattern was much stronger for Sample 2. In both samples, however, the direct effect of teacher perceptions of performance on grades (βs = .00 and .20) was similar to its effect on MEAP scores (βs = .00 and .21). The main reason for the increased total effect on grades was because there was also an indirect effect on grades, mediated by teacher perceptions of effort.[8]

Perceptual Biases: Limitations

Teacher perceptions of effort. To interpret these results as representing perceptual biases, one must rule out the accuracy explanation. Perhaps rather than biasing their evaluations of students' achievement, teachers simply rewarded students they accurately perceived as exerting strong effort by increasing their grades, and perhaps they punished students accurately perceived as lazy by lowering their grades. How accurate, then, were teacher perceptions of effort? Hardly accurate at all, according to the criteria available in our data. Teacher perceptions of effort were minimally related to students' global self-perceptions of effort, how much time they spent on homework, or their self-perceptions of math ability.

Strictly, however, the minimal association between teacher perceptions of effort and students' self-perceptions represents only a lack of agreement regarding students' level of effort. It represents inaccuracy on the part of teachers only if students' self-perceptions of effort were accurate. How good, then, are these criteria?

Additional analyses supported the validity of students' self-reported effort. The correlations of student effort with self-concept of ability were .22 in Sample 1 and .31 in Sample 2 (both $ps < .001$). This is precisely the type of relation one would

[8] Lower reliabilities for grades than for achievement test scores represents a potential alternative explanation of the stronger expectancy effects for grades (Cook & Campbell, 1979). The stabilities of grades were about .7 (see Table 2), which provides a lower bound estimate of their reliabilities (reliabilities much below .7 would yield stabilities, corrected for attenuation from measurement error, that were greater than 1.0—a mathematical impossibility). Therefore, the largest possible difference between their reliabilities assumes that the reliabilities for grades is .7 and the reliabilities for achievement test scores is 1.0 (it is unlikely that the reliability for test scores is 1.0; lowering this reliability, however, lowers the discrepancy between the reliabilities of grades versus test scores and therefore provides a weaker test of this alternative explanation). Therefore, we estimated another variation of the three separate variables model that assumed reliabilities of .7 for the grades and 1.0 for test scores. Changes in path coefficients relating teacher expectations to grades and test scores were minimal. No previously significant coefficients were reduced to nonsignificance, and the pattern of substantially stronger relations of teacher expectations to grades than test scores found in the main analyses remained essentially unchanged.

expect if, as many theorists have proposed, high self-concept of ability enhances motivation (e.g., Bandura, 1977; Dweck & Elliot, 1984; Eccles & Wigfield, 1985). Furthermore, self-perceptions of ability play a major role in motivating effort (e.g., Bandura, 1977; Dweck & Elliot, 1984; Eccles & Wigfield, 1985; Harter, 1984; Marsh, 1990; Weiner, 1979). Therefore, if teacher perceptions of students' effort were accurate, they should have been strongly related, at least, to students' self-concept of ability. This was not found.

Furthermore, we know of no evidence in the more general literature on attributions and person perception to suggest that observers are generally better judges of effort than actors. Both theory and evidence suggest that actors often generate more accurate attributions for their behavior than do observers (Monson & Snyder, 1977). Previous research has demonstrated that teachers' beliefs in strong associations between effort and intelligence are largely illusory (e.g., Barnard, Zimbardo, & Sarason, 1968; see Brophy & Good, 1974, for a review). Finally, our results showed that teachers inferred students' effort almost entirely from their performance—an inference that is not justified by evidence showing that students' self-perceptions of effort were only slightly positively related to grades and that the time they reported spending on homework was negatively related to MEAP scores.

We feel that these results, in conjunction with previous research documenting the illusory nature of teacher beliefs in strong associations between achievement and effort, are sufficient to warrant at least tentatively concluding that teacher perceptions of effort are largely inaccurate and thus produce a biasing effect on grades. This conclusion would warrant revision if future research, using demonstrably superior criteria, shows that teacher perceptions of effort predict grades more than standardized achievement test scores because these perceptions are accurate.

Teacher perceptions of performance. Teacher perceptions of performance predicted grades more strongly than MEAP scores primarily because of the biasing effects of teacher perceptions of effort. The direct effect of teacher perceptions of performance on grades and MEAP scores was similar in both samples. However, the model also assumed that teacher perceptions of performance influenced teacher perceptions of effort, and results showed that teacher perceptions of effort predicted grades more than MEAP scores. Therefore, teacher perceptions of performance also predicted grades more than MEAP scores because of the indirect effects mediated by teacher perceptions of effort.

This pattern challenges the plausibility of an accuracy interpretation. The accuracy interpretation proposes that teachers are particularly astute in their observation of the students in their classrooms and are better at predicting in-class performance than predicting standardized achievement test scores. Therefore, the accuracy explanation predicts that the direct relation of teacher perceptions of performance to grades should exceed the direct relation to MEAP scores (because they are better at predicting in-class performance). This pattern was not found. In contrast, it seems that teachers (a) simply assumed that higher achievers exerted more effort, (b) rewarded students who were already high achievers with even higher grades (or punished students who were low achievers with even lower grades), and (c) were largely oblivious to the actual effort students believed they exerted.

Why an effort-based bias? Several factors may underlie an effort-based bias. First, effort probably is difficult to observe directly. Therefore, it may be necessary for teachers to infer effort on the basis of some observable behavior. Performance is one directly observable behavior that seems especially likely to influence teachers' perceptions of students' effort. People often assume effort strongly influences performance (Covington & Omelich, 1979; Heider, 1958; Schuman et al., 1985; Weiner, 1979). Perhaps because they believe in the American work ethic (see, e.g., Schuman et al., 1985) or in a just world (Lerner, 1980), people assume that hard work pays off. According to this naive theory, it is reasonable to infer that high achievement generally reflects strong effort and therefore to reward those who are believed to try hard or to punish those seen as lazy. This effort-based bias, therefore, seems to operate in such a way that the academically rich get richer and the academically poor get poorer.

Because these perceptions influenced students' grades, they are of some practical importance. Decisions to place students in high or low tracks are often based at least partially on grades, and once placed, students rarely change tracks (Brophy & Good, 1974). Track placement, in turn, has implications for several future opportunities, including ease of getting into college.

Predictive Accuracy

Predictive accuracy refers to predictive validity without influence; the path coefficients obtained in the analyses described above are intended to represent influences.[9] Consequently, the difference between the zero-order correlations (overall predictive validities) and path coefficients (predictive validity due to influence) relating teacher expectations to student achievement is an index of the extent to which teacher expectations predicted, without influencing, student achievement (i.e., accuracy).

Results show considerable accuracy in teachers' expectations among both samples (these are summarized in Table 3). Zero-order correlations of the individual teacher-perception variables with grades range from .51 to .71; the structural equation analyses, which controlled for potential sources of accuracy, reduced these correlations to path coefficients of .04 to .34. Similarly, the multiple correlations of the three teacher-perception variables with grades are .64 and .73; the multiple semipartial correlations, obtained after controlling for potential sources of accuracy, are .23 and .29. On average, the zero-order correlations of teacher expectations with students' final grades were reduced by about 60% to 65% in analyses that controlled for

[9] As with all structural modeling of correlational data, path coefficients represent influences only if the model and measures are valid and if measurement error does not dramatically alter the results (see, e.g., Bollen, 1989; Kenny, 1979; Pedhazur, 1982, for extended discussions of these issues).

Table 3
Construction and Reflection of Student Achievement

	Teachers' perceptions of students' performance		Teachers' perceptions of students' talent		Teachers' perceptions of students' effort		All three		Latent teacher-perception model		Induced teacher-perception model	
Measure	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2
Correlation with grades	.56	.71	.59	.64	.51	.63	.64	.73	.68	.76	.64	.73
Effect on grades	.17	.34	.13	.04	.17	.20	.23	.29	.34	.46	.30	.41
Correlation with MEAP scores	.46	.54	.57	.49	.36	.36	.58	.55	.58	.56	.54	.50
Effect on MEAP scores	.10	.15	.18	−.02	−.03	−.07	.13	.12	.15	.16	.12	.10

Note. *Effects* refers to standardized total effects. The difference between correlations and path coefficients is an index of predictive accuracy (see text for explanation). The column titled *All three* reports the multiple correlation of all three teacher-perception variables with grades and MEAP scores in the correlation rows and reports the multiple semipartial correlation (controlling for the student background variables) with grades and MEAP scores in the effect rows. All coefficients greater than .03 are significant at $p < .05$. MEAP = Michigan Educational Assessment Program.

potential sources of accuracy in those expectations (previous achievement, motivation, etc.).

Results provide even greater evidence of accuracy in predicting MEAP scores. Zero-order correlations of the individual teacher-perception variables with MEAP scores ranged from .36 to .57; path coefficients ranged from −.07 to .18. The multiple correlation of the three teacher-perception variables with grades were .58 and .55; the multiple semipartial correlations were .13 and .12. These results show that, on average, the zero-order correlations of teacher expectations with MEAP scores were reduced by about 80% in analyses that controlled for potential sources of accuracy in those expectations.

Stage 2: Direct Comparison of Sample 1 and Sample 2

The broad pattern of results across the two samples is quite similar: (a) Consistent with the self-fulfilling prophecy hypothesis, teacher expectations significantly predicted both final grades and MEAP scores; (b) consistent with the perceptual bias hypotheses, teacher expectations predicted final grades more strongly than MEAP scores; (c) teacher perceptions of performance and talent are largely based on appropriate factors; (d) teacher perceptions of effort are almost completely unrelated to students' perceptions of their effort; (e) students' gender, independent of their performance, related to all three teacher perceptions; and (f) zero-order correlations between teacher expectations and student achievement were reduced by 60% to 80% in the analyses that controlled for potential sources of accuracy.

Nonetheless, there were also some differences between Samples 1 and 2 (see Figures 1 and 2 and Table 3). Therefore, two-group LISREL analyses assessed the viability of assuming that all, some, or no identical relationships characterized both samples. The first two-group analysis assumed that the two models had the same structure (i.e., the same pattern of coefficients underlying the analyses presented thus far), but made no assumptions regarding the value any coefficient might take. This is the least restrictive model and provided a basis for comparing models that assume some or all paths are identical across the two models. The most restrictive model assumes that all coefficients are identical across the two samples.

A series of analyses assessed how well each of these models fit the data. Because different fit measures provide different information, three global measures of fit were used in the current analyses. The chi-square assesses the probability that a model fully accounts for all covariances among the observed variables. In essence, the chi-square answers the question "Is this model significantly different from a perfect model?" With large samples, virtually any deviation from perfection may pro-

Table 4
Summary of Fit Assessment for Two-Sample Comparisons

Model	χ^2	*df*	GFI	NFI
Null	9933.27	144		
No invariances	47.63	18	.995	.995
Only factor invariant	91.84	21	.993	.991
Whole invariant	323.74	87	.987	.967
Invariant phis	95.39	39	.994	.990
Invariant gammas	126.41	48	.992	.987
Invariant psis	122.02	25	.993	.988
Invariant betas	80.30	26	.994	.992

Note. This table displays the chi-squares obtained when assuming that various sets of coefficients are identical for the two samples. All *N*s = 1,731. *No invariances* means that no coefficients were assumed to be identical; *whole model invariant* means that all coefficients were assumed to be identical. Phis, gammas, psis, and betas all refer to subsets of LISREL coefficients (see Jöreskog & Sörbom, 1983). All chi-squares are significant at $p < .001$. GFI = goodness of fit index; NFI = normed fit index.

duce a statistically significant chi-square. Consequently, we also used two other global measures of fit that are independent of the number of subjects in a particular sample (Bollen, 1990). LISREL VI provides its own goodness-of-fit index (GFI), which reflects the relative amount of covariances accounted for by the model. We also report the normed fit index (NFI; Bentler & Bonett, 1980), which reflects the extent to which a model improves on a null model (as is most common, our null models assume complete independence among all variables). Both the GFI and NFI have upper limits of 1.0, and values exceeding .9 are generally assumed to reflect well-fitting models. Table 4 summarizes the results for all two-sample analyses.

Baseline Model

The baseline for comparison is the model making no assumptions regarding the values of any of the coefficients in the two models (which we refer to as the *no invariances* model; this model, which was described in the section titled Model Assumptions, simply combines Figures 1 and 2). The chi-square for this model is actually the sum of the chi-squares for the two separate models. Although this model significantly deviates from perfection, $\chi^2(18, N = 1{,}731) = 47.63$, $p < .001$, it accounted for nearly all of the covariances and variances among the variables (GFI and NFI = .995).

Is It Appropriate to Assume That All Path Coefficients for the Two Samples Are Identical?

The most stringent model requires that the path coefficients are the same in the two samples. We refer to this as the *completely invariant* model. In general, the coefficients generated by this model fall between those obtained when estimating the model separately for Sample 1 and Sample 2, although they also tended to be closer to the values obtained in Sample 2 because of its greater sample size.

This model significantly deviated from a perfect model, $\chi^2(87, N = 1{,}731) = 323.74$, $p < .001$, and significantly worsened the fit of the model, $\chi^2(69, N = 1{,}731) = 276.11$, $p < .001$. According to LISREL's GFI, however, these restrictions only slightly worsened the fit of the model (which decreased from .995 to .987). The NFI indicated a somewhat more substantial reduction in fit (from .995 to .967, although even .967 is generally considered very good fit; see Bentler & Bonett, 1980).

Is a Particular Subset of Coefficients Primarily Responsible for the Discrepancies Among the Two Samples?

Further analyses explored whether the discrepancies between the samples occurred primarily among a particular subset of coefficients. Consequently, models assuming identical coefficients only among the factor loadings and measurement errors, phi matrix (covariances of the exogenous variables), beta matrix (path coefficients relating the endogenous variables to one another), gamma matrix (path coefficients relating the exogenous variables to the endogenous variables) and psi matrix (variance–covariance matrix of the error terms of the endogenous variables) were assessed. These results are summarized in Table 4. Although all such restrictions produced significant worsening of fit according to the chi-square criterion (all *p*s < .001), the overall fit of all models as indicated by the GFI and NFI is quite good (around .99, see Table 4). These results further support the conclusion that the patterns of coefficients obtained for the two samples are similar. However, because of the large sample, even small differences are statistically significant.

Stage 3: Alternative Conceptualizations of Teacher Expectations

All results thus far reported allowed each of the three teacher-perception variables to have unique relations with other variables in the model. This "three separate variables" approach was previously used in Jussim (1989). There are, however, at least two other ways to conceptualize the teacher-perception variables: a "latent variable" model and an "induced variable" model. The latent teacher-perception model assumes that each of the three teacher-perception variables represent indicators of an underlying teacher-perception factor. This model estimates and removes measurement error from the underlying teacher-perception factor. It is identical to the three separate variables model, except that only the latent teacher-perception variable has direct relations with student background variables, future grades, and MEAP scores (the three measured teacher-perception variables are directly linked only to the latent teacher-perception variable).

A second alternative is that the three teacher perception variables caused an unmeasured teacher-expectation factor and that this factor influenced student achievement. This induced variable model (e.g., Bollen, 1989; Heise, 1972) assumes that the

Table 5
Fit Assessment of Different Conceptualizations of Teacher Expectations

Model	χ^2	*df*	GFI	NFI
	Sample 1 (N = 443)			
Null	2481.18**	66		
Three separate variable	8.39	9	.997	.997
Latent teacher perception	134.05**	25	.952	.946
Induced teacher perception	18.47*	11	.993	.993
	Sample 2 (N = 1,288)			
Null	7362.61**	66		
Three separate variable	39.24**	9	.995	.995
Latent teacher perception	244.41**	25	.969	.967
Induced teacher perception	60.25**	11	.992	.992

Note. GFI = goodness of fit index; NFI = normed fit index.
* $p < .10$. ** $p < .001$.

three measured teacher-perception variables are sources of teacher expectations. It specifies three stages of relations: (a) Student background variables influence the three teacher perception variables (it assumes no causal relations among teacher perceptions), (b) the three teacher perceptions influence an unmeasured global teacher-expectation variable, and (c) the unmeasured teacher expectation variable influences final grades and MEAP scores.

When predictors are correlated (as are the three teacher-perception variables), separate path estimates may be inappropriately small (Gordon, 1968; Pedhazur, 1982). By combining the teacher-perception variables into a single induced or latent variable, these models might provide a more accurate (and larger) estimate of relations of teacher expectations to student achievement than obtained using the three separate variables approach. Table 5 summarizes the fit assessment of these models. The latent teacher-perception model failed to fit the data for Sample 1, $\chi^2(25, N = 443) = 134.05$, $p < .001$, GFI = .952, NFI = .946, or Sample 2, $\chi^2(25, N = 1{,}288) = 244.41$, $p < .001$, GFI = .969, NFI = .967. Although the GFI and NFI indicate adequate fit, the highly significant chi-square for even the smaller Sample 1 strongly suggests that this model is not viable. The induced variable model fit the data almost, but not quite, as well as the three separate variable model. For Sample 1, $\chi^2(11, N = 443) = 18.47$, $p = .071$, GFI = .993, NFI = .993; for Sample 2, $\chi^2(11, N = 1{,}288) = 60.25$, $p < .001$, GFI = .992, NFI = .992. Table 3 shows that the general pattern of results representing reflection and construction of student achievement obtained for the latent variable model and induced variable model were similar to those obtained in the three separate variable model.

Conclusion

Teacher Expectations

This study provided several significant contributions. First, it showed that (a) the main reason teacher expectations predicted student achievement was because they were accurate; (b) there were small self-fulfilling prophecy effects on standardized math test performance; and (c) there were somewhat larger expectancy effects on grades. Path coefficients relating teacher expectations to standardized achievement tests were about 80% lower than zero-order correlations. This means that about 80% of the correlations between teacher expectations and standardized achievement test scores represents predictive validity without influence (accuracy), and about 20% represents self-fulfilling prophecy.

Path coefficients relating teacher expectations to students' grades were 45%–65% lower than the zero-order correlations. This means that about 45%–65% of the correlations between teacher expectations and grades represents predictive validity without influence (accuracy), and about 35%–55% represents expectancy effects. The finding of path coefficients of .3–.4 relating teacher perceptions of performance to grades represents one of the largest expectancy effects yet obtained in a naturalistic study that controlled for previous performance.

Because grades reflect teachers' judgments about students' performance, path coefficients relating teacher expectations to grades combine two different types of expectancy effects: self-fulfilling prophecies (teachers' expectations may influence students' actual performance) and perceptual biases (teachers' expectations may influence their own judgments of students' performance). A reasonable heuristic for approximating the self-fulfilling prophecy effects on grades is to assume they are similar to self-fulfilling prophecy effects on MEAP scores (which is indicated by the path coefficients relating teacher expectations to MEAP scores). On this basis, the last three columns of Table 3 (which show relations to grades and MEAP scores of all three teacher-perception variables together) suggest that about 25%–55% of the expectancy effect on grades is self-fulfilling prophecy and about 45%–75% is perceptual bias.[10]

A second contribution of this study is that it showed that these results replicated across two samples, including 98 teachers and over 1,700 students. Furthermore, they replicated across a variety of conceptualizations of the nature of teacher expectations. This is important because it attests to the robustness of findings that provide further evidence against the strong claims regarding the prevalence and power of expectancy effects that are frequently found in many social–psychological reviews (e.g., Hamilton et al., 1990; Jones, 1986; Miller & Turnbull, 1986; Snyder, 1984) and empirical articles (e.g., Frieze, Olson, & Russell, 1991; Skov & Sherman, 1986; Snyder et al., 1977), while supporting claims regarding the accuracy of teacher expectations and limited extent of self-fulfilling prophecies often found in the educational literature (Brophy, 1983; Dusek, 1975; Meyer, 1985; West & Anderson, 1976).

Gender Bias

The third contribution of the study was its exploration of the role of student sex in the teacher-expectation process. As predicted, student sex was related to teacher perceptions in a manner that suggests perceptual bias. Teachers rated boys as having more math talent than girls and rated girls as trying harder than boys. Neither of these differences appears to reflect accurate perceptions of gender differences. Furthermore, these results contribute to understanding why girls often receive higher math grades than boys (Kimball, 1989). Specifically, teachers erroneously assume that girls try harder than boys and then reward girls for their (misperceived) higher effort with higher grades.

Teacher perceptions were generally consistent with stereotypes of gender differences: Boys have more talent and girls compensate by working harder. These beliefs are quite common: Adolescents hold similar beliefs regarding their own tal-

[10] For example, coefficients for the induced teacher perception model for Sample 2 (shown in Table 3) are .10 for MEAP scores and .41 for grades. Our suggestion for disentangling self-fulfilling prophecy and perceptual bias effects on grades is to assume about .10 of the .41 effect on grades (or 24%) represents self-fulfilling prophecy and about .31 (or 76%) represents perceptual bias. Our range of 25%–55% perceptual bias and 45%–75% self-fulfilling prophecy derives from performing these calculations on the coefficients presented in each of the last three columns of Table 3.

ents and abilities (Eccles [Parsons], 1983; Parsons et al., 1982), and parents hold similar beliefs regarding the abilities and efforts of their own children (Yee & Eccles, 1988). Furthermore, this bias in parents' perceptions of their children's talent and effort in math appears to be linked to the parents' own category-based, gender role stereotypic beliefs regarding the general distribution of math talent between boys and girls (Eccles, Jacobs, & Harold, 1990). Future research should determine whether a similar mechanism underlies the apparent gender role stereotyped bias in teachers' perceptions of their students' talent and effort.

Strong and Weak Constructivist Perspectives

Results generally supported a weak version of social constructivism, which argues that although beliefs sometimes create social reality, people's perceptions usually accurately reflect social reality. These results do not support the strong social constructivist perspective, which assumes that social perception creates social reality as much or more than it reflects social reality. We do not claim that expectancies are always accurate or never have large self-fulfilling or biasing effects. Attributes with less objective criteria than math achievement, such as attitudes and personality dispositions, may be more subject to biases in social perception (e.g., Cronbach, 1955; Funder, 1987; Jussim & Osgood, 1989). One should not generalize from this study to different grade levels or to different types of social interaction (e.g., between friends). Nonetheless, results of the current study are consistent with the effect sizes of between .1 and .3 obtained in virtually all path-analytic and meta-analytic studies of expectancy effects (Cooper & Hazelrigg, 1988; Harris & Rosenthal, 1985; Jussim, 1989; Raudenbush, 1984; Rosenthal & Rubin, 1978; Smith, 1980; West & Anderson, 1976; Williams, 1976).

Accumulation of Expectancy Effects

Even small effects may lead to large differences if they accumulate over a sufficiently long time. Thus, part of the accuracy of sixth-grade teacher expectations may have been based on an awareness of the results of previous expectancy effects. The current study, however, did not attempt to assess whether expectancy effects occurred before sixth grade. Our purpose was to assess the extent of accuracy, self-fulfilling prophecy, and perceptual bias within a single social context—from a given starting point to a particular endpoint among a particular set of perceivers and targets (see Jussim, 1991, for a more extended discussion of "single social context"). Our approach is not useful for identifying the "ultimate" extent to which individual differences in students' achievement result from social constructive processes.

Furthermore, the little research on the accumulation of expectancy effects has yielded a mixed picture. Both Rosenthal and Jacobson's (1968) field experiment and West and Anderson's (1976) naturalistic study found that, rather than accumulating, teacher-expectation effects dissipated after the first year. In contrast, an observational study by Rist (1970) suggested that the accumulation of such effects may be quite powerful. However, because Rist provided no objective or quantitative assessment of the accumulation of teacher-expectation effects, identifying naturalistic conditions conducive to powerful expectancy effects remains a challenge for future research. For both the practical purpose of understanding how teachers enhance or undermine students' achievement and the theoretical goal of understanding how much social perception constructs and reflects social reality, empirical research on the accumulation of expectancy effects is sorely needed.

References

Alwin, D., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review, 40,* 37–47.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Bulletin, 84,* 191–215.

Barnard, J., Zimbardo, P., & Sarason, S. (1968). Teachers' ratings of student personality traits as they relate to IQ and social desirability. *Journal of Educational Psychology, 59,* 128–132.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606.

Berman, J. S. (1979). Social bases of psychotherapy: Expectancy, attraction, and the outcome of treatment (Doctoral dissertation, Harvard University). *Dissertation Abstracts International, 40,* 5800B.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley Interscience.

Bollen, K. A. (1990). Overall fit in covariance structure models: Two types of sample size effects. *Psychological Bulletin, 107,* 256–259.

Brattesani, K. A., Weinstein, R. S., & Marshall, H. H. (1984). Student perceptions of differential teacher treatment as moderators of teacher expectation effects. *Journal of Educational Psychology, 76,* 236–247.

Brophy, J. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology, 75,* 631–661.

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

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Boston: Houghton-Mifflin.

Cooper, H. (1979). Pygmalion grows up: A model for teacher expectation communication, and performance influence. *Review of Educational Research, 49,* 389–410.

Cooper, H., & Hazelrigg, P. (1988). Personality moderators of interpersonal expectancy effects: An integrative research review. *Journal of Personality and Social Psychology, 55,* 937–949.

Covington, M. V., & Omelich, C. L. (1979). Effort: The double-edged sword in school achievement. *Journal of Educational Psychology, 71,* 169–182.

Crano, W. D., & Mellon, P. M. (1978). Causal influence of teachers' expectations on children's academic performance: A cross-lagged panel analysis. *Journal of Educational Psychology, 70,* 39–49.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin, 52,* 177–193.

Darley, J. M., Fleming, J. H., Hilton, J. L., & Swann, W. B., Jr. (1988). Dispelling negative expectancies: The impact of interaction goals and target characteristics on the expectancy confirmation process. *Journal of Experimental Social Psychology, 24,* 19–36.

Duncan, O. D. (1975). *Introduction to structure equation models.* San Diego, CA: Academic Press.

Dusek, J. (1975). Do teachers bias children's learning? *Review of Educational Research, 45,* 661–684.

Dweck, C. S., Davidson, W., Nelson, S., & Enna, B. (1978). Sex differences in learned helplessness: II. The contingencies of evaluative feedback in the classroom: III. An experimental analysis. *Developmental Psychology, 14,* 268–276.

Dweck, C. S., & Elliott, E. S. (1984). Achievement motivation. In P. H. Mussen (Ed.), *Handbook of child psychology* (Vol. 4). New York: Wiley.

Eccles, J. (1988). Achievement beliefs and environment. Final report to the National Institute of Child Health and Development.

Eccles, J. S., & Blumenfeld, P. (1985). Classroom experiences and student gender. In L. C. Wilkinson & C. B. Marrett (Eds.), *Gender influences in classroom interaction* (pp. 79–114). San Diego, CA: Academic Press.

Eccles, J., Jacobs, J. E., & Harold, R. D. (1990). Gender-role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of Social Issues, 46,* 183–201.

Eccles, J., & Wigfield, A. (1985). Teacher expectations and student motivation. In J. Dusek (Ed.), *Teacher expectancies.* Hillsdale, NJ: Erlbaum.

Eccles (Parsons), J. (1983). Attributional processes as mediators of sex differences in achievement. *Journal of Educational Equity and Leadership, 3,* 19–27.

Eccles (Parsons), J., Adler, T., & Meece, J. L. (1984). Sex differences in achievement: A test of alternate theories. *Journal of Personality and Social Psychology, 46,* 26–43.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). Reading, MA: Addison Wesley.

Frieze, I. H., Olson, J. E., & Russell, J. (1991). Attractiveness and income for men and women in management. *Journal of Applied Social Psychology, 21,* 1039–1057.

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin, 101,* 75–90.

Gage, N. L., & Cronbach, L. J. (1955). Conceptual and methodological problems in interpersonal perception. *Psychological Review, 62,* 411–422.

Gordon, R. A. (1968). Issues in multiple regression. *American Journal of Sociology, 73,* 592–616.

Hamilton, D. L., Sherman, S. J., & Ruvolo, C. M. (1990). Stereotype-based expectancies: Effects on information processing and social behavior. *Journal of Social Issues, 46*(2), 35–60.

Harris, M. J., & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological Bulletin, 97,* 363–386.

Harter, S. (1984). Developmental perspectives on the self-system. In P. H. Mussen (Ed.), *Handbook of child psychology* (Vol. 4, pp. 276–384). New York: Wiley.

Heider, F. (1958). *The psychology of interpersonal relations.* New York: Wiley.

Heise, D. R. (1972). Employing nominal variables, induced variables, and block variables in path analysis. *Sociological Methods and Research, 1,* 147–173.

Humphreys, L. G., & Stubbs, J. (1977). A longitudinal analysis of teacher expectation, student expectation, and student achievement. *Journal of Educational Measurement, 14,* 261–270.

Jones, E. E. (1986). Interpreting interpersonal behavior: The effects of expectancies. *Science, 234,* 41–46.

Jones, E. E. (1990). *Interpersonal perception.* New York: Freeman.

Jöreskog, K. G., & Sörbom, D. (1983). *LISREL VI user's guide.* Chicago: International Educational Services.

Jussim, L. (1986). Self-fulfilling prophecies: A theoretical and integrative review. *Psychological Review, 93,* 429–445.

Jussim, L. (1987). Interpersonal expectations in social interaction: Self-fulfilling prophecies, confirmatory biases, and accuracy (Doctoral dissertation, University of Michigan, 1987). *Dissertation Abstracts International, 48,* 1845B.

Jussim, L. (1989). Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology, 57,* 469–480.

Jussim, L. (1990). Social reality and social problems: The role of expectancies. *Journal of Social Issues, 46,* 9–34.

Jussim, L. (1991). Social perception and social reality: A reflection-construction model. *Psychological Review, 98,* 54–73.

Jussim, L., & Osgood, D. W. (1989). Influence and similarity among friends: An integrative model applied to incarcerated adolescents. *Social Psychology Quarterly, 52,* 98–112.

Kenny, D. A. (1979). *Correlation and causality.* New York: Wiley Interscience.

Kimball, M. M. (1989). A new perspective on women's math achievement. *Psychological Bulletin, 105,* 198–214.

Lerner, M. (1980). *The belief in a just world.* New York: Plenum Press.

Marsh, H. W. (1990). The causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal panel analysis. *Journal of Educational Psychology, 82,* 646–656.

Meyer, W. J. (1985). Summary, integration, and prospective. In J. Dusek (Ed.), *Teacher expectancies* (pp. 353–370). Hillsdale, NJ: Erlbaum.

Miller, D. T., & Turnbull, W. (1986). Expectancies and interpersonal processes. *Annual review of psychology, 37,* 233–256.

Monson, T. C., & Snyder, M. (1977). Actors, observers, and the attribution process: Toward a reconceptualization. *Journal of Experimental Social Psychology, 13,* 89–111.

Nicholls, J. G. (1979). Quality and equality in intellectual development: The role of motivation in education. *American Psychologist, 34,* 1071–1084.

Parsons, J. E. (1980). *Final report to the National Institute of Education, Washington, DC.* (ERIC Document Reproduction Service No. ED 186 477)

Parsons, J. E., Kaczala, C. M., & Meece, J. L. (1982). Socialization of achievement attitudes and beliefs: Classroom influences. *Child Development, 53,* 322–339.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research: Explanation and prediction.* (2nd ed.). New York: Holt, Rinehart & Winston.

Raudenbush, S. W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy inductions: A synthesis of findings from 18 experiments. *Journal of Educational Psychology, 76,* 85–97.

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: Teacher expectation and student intellectual development.* New York: Holt, Rinehart & Winston.

Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *The Behavioral and Brain Sciences, 3,* 377–386.

Schuman, H., Walsh, E., Olson, C., & Etheridge, B. (1985). Effort and reward: The assumption that college grades are affected by quantity of study. *Social Forces, 63,* 945–966.

Skov, R. B., & Sherman, S. J. (1986). Information-gathering processes: Diagnosticity, hypothesis-confirmatory strategies, and perceived hypothesis confirmation. *Journal of Experimental Social Psychology, 22,* 93–121.

Smith, M. L. (1980). Teacher expectations. *Evaluation in Education, 4,* 53–55.

Snyder, M. (1984). When belief creates reality. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 18, pp. 247–305). San Diego, CA: Academic Press.

Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and

interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology, 35,* 656–666.
Sutherland, A., & Goldschmid, M. L. (1974). Negative teacher expectation and IQ change in children with superior intellectual potential. *Child Development, 45,* 852–856.
Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of Educational Psychology, 71,* 3–25.
West, C., & Anderson, T. (1976). The question of preponderant causation in teacher expectancy research. *Review of Educational Research, 46,* 613–630.
Wilkinson, L. C., & Marrett, C. B. (1985). *Gender influences in classroom interaction.* San Diego, CA: Academic Press.
Williams, T. (1976). Teacher prophecies and the inheritance of inequality. *Sociology of Education, 49,* 223–236.
Yee, D., & Eccles, J. S. (1988). Parent perceptions and attributions for children's math achievement. *Sex Roles, 19,* 317–333.

Appendix

Measures

Teacher Expectations

1. How much natural mathematical talent does this student have?

very little math talent 1 2 3 4 5 6 7 a lot of math talent

2. How hard does this student try in math?

does not try at all 1 2 3 4 5 6 7 tries very hard

3. Compared to other students in this class, how well is this student performing in math?

1	2	3	4	5
near the bottom of the class	below the middle of the class	in the middle of the class	above the middle of the class	one of the best in the class

Students' Self-Perceptions of Effort and Time on Homework

1. How hard do you work in math?

a little 1 2 3 4 5 6 7 a lot

2. How much time do you spend on math homework? (Check one answer.)

_____ 1) less than 15 minutes a day
_____ 2) 15 to 30 minutes a day
_____ 3) 30 minutes to an hour a day
_____ 4) an hour or more a day

Students' Self-Concept of Math Ability

1. How good at math are you?

not at all good 1 2 3 4 5 6 7 very good

2. If you were to rank all the students in your math class from the worst to the best in math, where would you put yourself?

the worst 1 2 3 4 5 6 7 the best

Received February 11, 1991
Revision received July 8, 1992
Accepted July 10, 1992 ■

Carr Appointed Editor of the *Journal of Experimental Psychology: Human Perception and Performance*, 1994–1999

The Publications and Communications Board of the American Psychological Association announces the appointment of Thomas H. Carr, PhD, Michigan State University, as editor of the *Journal of Experimental Psychology: Human Perception and Performance* for a 6-year term beginning in 1994. As of December 15, 1992, manuscripts should be directed to

Thomas H. Carr, PhD
Department of Psychology
Michigan State University
East Lansing, Michigan 48824

Manuscript submission patterns for *JEP: Human Perception and Performance* make the precise date of completion of the 1993 volume uncertain. The current editor, James E. Cutting, PhD, will receive and consider manuscripts until December 14, 1992. Should the 1993 volume be completed before that date, manuscripts will be redirected to Dr. Carr for consideration in the 1994 volume.