JOURNAL OF EXPERIMENTAL SOCIAL PSYCHOLOGY 34, 530–561 (1998)
ARTICLE NO. JS981363

Self-Fulfilling Prophecies, Perceptual Biases, and Accuracy at the Individual and Group Levels

Alison E. Smith and Lee Jussim

Rutgers—The State University of New Jersey

Jacquelynne Eccles

University of Michigan

Michelle VanNoy

Mathematica Policy Research, Princeton, New Jersey

and

Stephanie Madon and Polly Palumbo

Rutgers—The State University of New Jersey

Received: June 20, 1997; revised: May 1, 1998; accepted: June 21, 1998

We examined ability grouping as a moderator of self-fulfilling prophecies in two ways. First, we examined ability group moderation for individual students with data from 1701 students and 97 teachers from sixth-grade math classes. Second, we examined ability group moderation at the class level with data from 108 sixth-grade math classes. Across both sets of analyses, results showed that: (1) teacher perceptions predicted student achievement primarily because those perceptions were accurate; (2) teacher perceptions biased marks

An earlier version of this paper was presented at the June 1996 Convention of the American Psychological Society. This research was supported by NICHD Grant 1 R29 HD28401-01A1 to the second author. Data collection was funded by grants to the third author from NIMH, NICHD, and NSF. We thank the personnel from all of the participating schools for their help with data collection. We also thank the following people for their aid in collecting and processing the data: Bonnie L. Barber, Christy Miller Buchanan, Harriet Feldlaufer, Connie Flanagan, Janis Jacobs, Dave Klingel, Doug MacIver, Carol Midgley, David Reuman, and Allan Wigfield.

Address correspondence and reprint requests to Alison Smith, Psychology Department, Tillett Hall, Rutgers University, New Brunswick, NJ 08903. E-mail: aesmith@rci.rutgers.edu.

0022-1031/98 $25.00

for students in heterogenous classes (consistent with perceptual bias); and (3) self-fulfilling prophecies were strongest when students were grouped within classes. We also found that teacher perceptions mediated the relationship between within-class ability grouping and student achievement. Implications for understanding accuracy, the power of self-fulfilling prophecies, and group level perceptions are discussed.

How is social perception related to social reality? Until recently, research in social psychology emphasized the power of false social beliefs to create reality and the tendency of perceivers to interpret reality in an expectancy confirming manner (e.g., Allport, 1954; Jones, 1986; Merton, 1948; Snyder, 1984). Research in the last 10 years, however, has consistently shown that the power of false social beliefs to influence both social reality and person perception is much smaller than once thought and that social perception is much more accurate than once believed (e.g., Funder, 1987, 1995; Jussim, 1991; Jussim, Eccles, & Madon, 1996; Kenny, 1994; Kenny & Albright, 1987; Kunda & Thagard, 1996). Consequently, much recent research on social perception has focused on identifying conditions under which phenomena such as self-fulfilling prophecies and cognitive biases are more likely to occur (e.g., Hilton & Darley, 1985; Jussim et al., 1996; Madon, Jussim, & Eccles, 1997; Neuberg, 1989; Snyder, 1992). This search has been fruitful. Demographics, goals, personality characteristics, power relationships, and situational factors all moderate self-fulfilling prophecies (e.g., Cooper & Hazelrigg, 1988; Harris, 1989; Jussim, et al., 1996; Madon, et al., 1997; Neuberg, 1989, 1994; Snyder, 1992).

One moderator of self-fulfilling prophecies that has not been examined is that of grouping targets. In this paper, we specifically examine whether grouping students by academic ability leads to more powerful self-fulfilling prophecies. The current research examined ability grouping as a moderator of self-fulfilling prophecies in two ways. It addressed: (1) whether self-fulfilling prophecies and perceptual biases were stronger among students grouped between-classes or within-classes than among students who were not grouped; and (2) whether self-fulfilling prophecies were stronger among students in high-ability groups, low-ability groups, or students who were not grouped.

Self-Fulfilling Prophecies, Perceptual Biases, and Accuracy

Targets may confirm perceivers' beliefs for any of three reasons—self-fulfilling prophecies, perceptual biases, or accuracy. A self-fulfilling prophecy occurs when a perceiver's originally false belief about a target is behaviorally confirmed by the target (Merton, 1948). For example, a self-fulfilling prophecy could occur if an employer falsely believes that an employee is inept and causes the employee to demonstrate incompetence.

Perceptual biases occur when a perceiver's beliefs influence their evaluation of target behavior. For example, a coach may believe that a particular athlete is especially talented. If the coach evaluates this athlete more favorably than another athlete of comparable talent, perceptual bias has occurred.

Perceivers' beliefs can also correspond to reality because they are accurate

(Jussim, 1991, 1993). Predictive accuracy refers to successfully predicting target behavior without influencing target behavior. Considerable evidence documents the accuracy of perceivers' beliefs and impressions of targets (Brophy, 1983; Funder, 1987, 1995; Jussim, 1989, 1991; Jussim & Eccles, 1992; Jussim et al., 1996; Kenny, 1994; Kenny & Albright, 1987).

Are Self-Fulfilling Prophecies Stronger When Students Are Grouped by Ability?

Types and levels of ability groups. Ability grouping refers to separating students into instructional groups based on their previous achievement. Students can be grouped by ability either within classes or between classes. For within-class groups, students are divided within the classroom into homogeneous instructional groups. Thus, in a within-grouped math class, there may be three math groups: high, average, and low ability. Each of these math groups may meet with the teacher separately to work on material geared to that group. For between-class groups, students of similar ability levels are divided into separate classrooms creating entire classes that are either high, average, or low ability. In this paper, we refer to classes as heterogenous when no ability grouping of any kind is used.

Ability grouping as a moderator of self-fulfilling prophecies. Alternative theoretical perspectives suggest that ability grouping might moderate the relationship between teacher perceptions and student achievement. We tested competing hypotheses generated by these alternative perspectives. One possibility is that ability grouping may lead to more powerful self-fulfilling prophecies because of labeling affects. Ability grouping represents institutional justification for believing that some students are smarter than others. Thus, it may lead to the type of rigid teacher perceptions that are most likely to evoke self-fulfilling prophecies and perceptual biases (Eccles & Wigfield, 1985; Jussim, 1986, 1990; Oakes, 1987). Consistent with this analysis, students in low-ability grouping often do not receive marks above a C, whereas students in high-ability grouping often do not receive marks below a B (Schaeffer & Olexa, 1971). Consequently, some researchers have argued that ability grouping creates a caste-like system wherein the academically rich get richer and the academically poor get poorer (Oakes, 1985, 1987; Rist, 1970). This perspective suggests that self-fulfilling prophecies and perceptual biases would be stronger for students in ability grouping.

Alternatively, ability grouping might increase the accuracy of teachers' perceptions. Between-class grouping reduces differences in student performance and ability within classes, thereby reducing the amount of information that teachers must consider. By simplifying the cognitive environment, between-class grouping may enhance the accuracy of teacher perceptions for two reasons: (1) it restricts the variability of student performance and talent within classes; and (2) it ensures that all students have prior records indicating that they fall above or below certain cutoffs. Teachers in these classes, therefore, may be less likely to develop wildly inaccurate perceptions. Therefore, grouping labels may help teachers form more accurate perceptions of their students.

In contrast, according to this perspective, teachers may be more likely to overestimate students (leading to frustration) or underestimate students (leading to boredom) in heterogeneous classrooms. Within heterogeneous classrooms, the wide range of ability may make it more difficult for teachers to develop accurate impressions of their students. This suggests that stronger self-fulfilling prophecies and perceptual biases would occur among students in heterogenous classes.

Do teacher perceptions mediate ability grouping effects? Ability grouping has often been criticized for increasing the achievement gap between students in high- and low-ability groups (Oakes, 1985, 1987; Rist, 1970). This could happen for at least one of two reasons: (1) ability grouping could hinder the achievement gains of students in low-ability groups; and (2) ability grouping could enhance the achievement gains of students in high-ability groups. Students in low-ability groups often fall increasingly further behind students in high-ability groups (Gamoran, 1992; Oakes, 1985, 1987). Furthermore, even if grouping programs do not limit the progress of students in low-ability groups, students in high-ability groups still benefit the most from grouping programs (Feldhusen & Moon, 1992; Kulik & Kulik, 1982, 1987, 1992).

One means by which ability grouping could increase differences in achievement among students in low- and high-ability groups is through teacher perceptions. Teachers may consistently overestimate students in high-ability groups and underestimate those in low-ability groups (Eiser, 1990; Krueger & Rothbart, 1990; Tajfel & Wilkes, 1963). If these perceptions lead to self-fulfilling prophecies, teachers could exacerbate preexisting differences between students in high- and low-ability groups. That is, teacher perceptions may mediate the relationship between level of ability grouping and student achievement. The mediation hypothesis states that effects of ability grouping on student achievement is at least in part due to teacher perceptions.

Are Self-Fulfilling Prophecies More Powerful for Whole Classes?

Perceptions of whole groups. Sometimes, perceptions of groups of people can lead to self-fulfilling prophecies. Group-level self-fulfilling prophecies are distinct from self-fulfilling prophecies for individuals and may apply to such issues as school tracking, collective bargaining, and team performance (at work, in sports, etc.). In this paper, when we discuss group-level self-fulfilling prophecies, we are not referring to demographic groups, but rather to groups such as classes, sports teams, or even work groups.

However, little research exists on self-fulfilling prophecies at the group level. With the exception of a single study (Eden, 1990), which examined the effect of positive expectations on platoon performance, all other studies have investigated self-fulfilling prophecies in dyadic interactions (see reviews by Darley & Fazio, 1980; Jussim, 1986; Jussim et al., 1996; Miller & Turnbull, 1986; Snyder, 1984, 1992).

Self-fulfilling prophecies may be more powerful among groups than among individuals (Brophy, 1983; Jussim & Fleming, 1996). Group-level self-fulfilling prophecies simply affect more people than do dyadic-level self-fulfilling prophe-

cies. Group-level self-fulfilling prophecies may also be more powerful because they may be supported by institutional policies (which perpetuate them and lend them credibility). For example, Merton (1948) discussed how the false belief that African-Americans were strikebreakers led to their exclusion from labor unions. Barred from this major avenue to employment, African-American workers had difficulty obtaining jobs. Consequently, when union workers went on strike, African-American workers readily accepted positions which had previously been unavailable to them. Thus, a false belief about a group that is supported by social or political institutions (such as labor unions) can have powerful effects on many individuals.

Group-level self-fulfilling prophecies may also be more powerful because erroneous perceptions regarding groups may be more difficult to disconfirm than erroneous perceptions of individuals. The stereotype literature attests to this—an atypical target may be perceived accurately, but may not change a perceiver's beliefs regarding the target's group (Brewer, 1988; Fiske & Neuberg, 1990). Finally, in the classroom, self-fulfilling prophecies may be more powerful for groups because teachers spend more time addressing their classes or ability groups as a whole than addressing individual students.

There are several ways to operationalize group-level expectations. One way is to directly measure a perceiver's expectations for a group of targets. Another way is to aggregate or average a perceiver's expectations for the individuals which compose a group of targets. In this paper, when we discuss teacher perceptions for whole classes, we are referring to the second method. That is, we aggregated teacher perceptions of the individual students in each class to yield teacher perceptions of each class as a whole (see Kenny, Kashy, & Bolger, in press, for a detailed discussion of separating out group-level and individual-level effects).

Whole-class teacher perceptions and ability grouping. We examined group-level self-fulfilling prophecies to determine whether these effects were more or less powerful than those among individual students. There is research to support both perspectives. First, self-fulfilling prophecies may be more powerful at the group level among classes that are grouped by ability. Teachers may hold rigid perceptions of students in classes that are labeled as high, average, or low ability because grouping labels may emphasize differences between the different ability-labeled classrooms. Research on categorization shows that placing stimuli into groups with different labels can lead perceivers to accentuate differences between stimuli from different groups (Eiser, 1971, 1990; Krueger & Rothbart, 1990; Tajfel & Wilkes, 1963). Furthermore, research has demonstrated that once perceivers are aware of categories, they are more likely to incorporate information that enhances intergroup differences into their category knowledge (Krueger & Rothbart, 1990; Krueger, Rothbart, & Sriram, 1989). Thus, grouping labels may lead teachers to overestimate differences between their ability-grouped classes.

Self-fulfilling prophecies may also be more powerful among ability-grouped classes because grouping labels may help to define upper and lower bounds for teachers' perceptions of the students in their classes. If teachers believe that a low-ability class cannot succeed with certain material, they may never present this

material to the class. Thus, teachers' perceptions may limit opportunities for ability-grouped classes to demonstrate that they can perform at higher levels (and disconfirm teachers' perceptions).

Alternatively, however, self-fulfilling prophecies and perceptual biases regarding whole classes may be small when students are grouped by ability. Although categorical accentuation of group differences sometimes occurs, often it does not (McCauley, 1995). Teachers spend a great deal of time with students and also have a great deal of information about them. When perceivers spend more time with targets, their perceptions are often more accurate and less likely to create self-fulfilling prophecies (Fiske & Neuberg, 1990; Raudenbush, 1984; Swann & Ely, 1984). In addition, individuating information, when available and especially when diagnostic, often eliminates the effects of stereotypes or group labels (Kunda & Thagard, 1996). Thus, teachers may be as accurate for their classes as they are for individual students because they have many opportunities to revise their perceptions with diagnostic information.

Research Overview

The present research is based on data from the Michigan Study of Adolescent Life Transitions. We used teacher perceptions of student achievement, measured early in the sixth-grade school year, to predict future student achievement (at the end of the sixth-grade school year). Previous work with this data set has documented the occurrence of self-fulfilling prophecies, perceptual biases, and predictive accuracy (Jussim, 1989; Jussim & Eccles, 1992; Jussim et al., 1996; Madon et al., 1997). The current research examined the role of ability grouping in moderating those effects by addressing the following questions: (1) Does type of ability grouping moderate relations between teacher perceptions and student achievement? That is, are perceptual biases and/or self-fulfilling prophecies stronger when students are grouped between-classes, within-classes, or not grouped? (2) Does ability group level moderate relations between teacher perceptions and student achievement? That is, are perceptual biases and/or self-fulfilling prophecies stronger among students in high-ability groups or in low-ability groups or among students who are not ability-grouped?

Ability grouping labels may also influence teacher perceptions and treatment of their students in varying ability groups. Thus, we also investigated whether teacher perceptions mediated the relationship between ability grouping and student achievement. We addressed the following questions: (1) Does level of ability grouping predict future student achievement? (2) Does level of ability grouping predict teacher perceptions over and above student achievement and motivation? And if so, (3) do teacher perceptions predict student achievement after controlling for student achievement, motivation, and level of ability grouping?

The current research also examined self-fulfilling prophecies for whole classes and moderation by ability grouping for whole classes. Recall that we averaged teacher perceptions of the individual students in each of their classes to yield an aggregate measure of teacher perceptions of their classes. Specifically, we addressed the

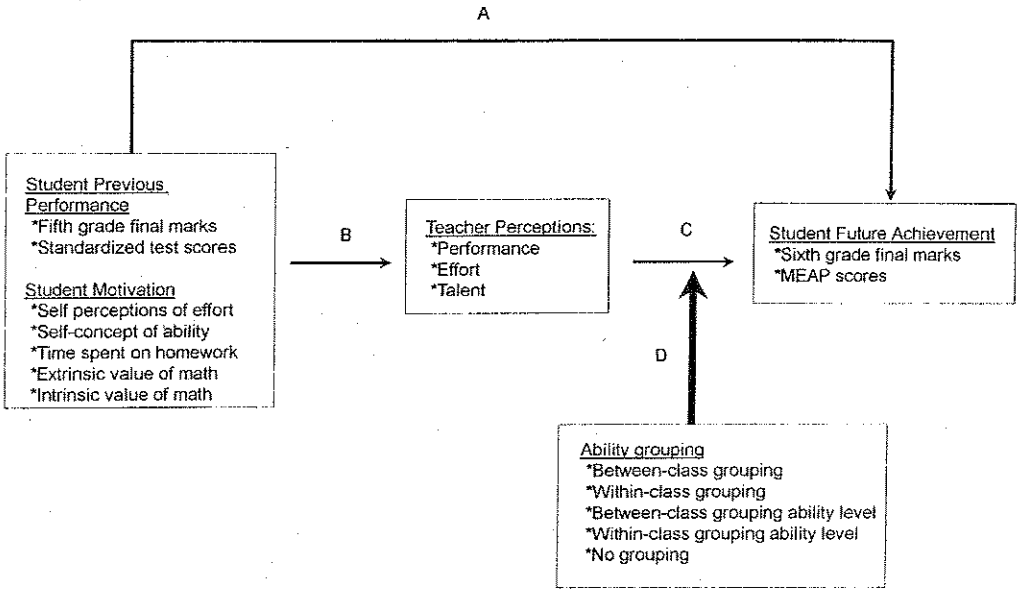

FIG. 1. Ability grouping as a moderator of the relationship between teacher perceptions and student achievement.

following questions: (1) Are self-fulfilling prophecies and perceptual biases stronger for whole classes than for individual students? (2) Does type of ability grouping moderate relations among teacher perceptions and class achievement?

Conceptual Model

This study examined the moderating effects of ability grouping on the relationship between teacher perceptions and future student achievement using a conceptual model developed by Jussim et al., 1996 (see Fig. 1). The model assumes that previous student achievement and student motivation influence future achievement and teacher perceptions (represented by paths A and B). The model also assumes that teacher perceptions can influence students' future achievement (represented by path C). Path C represents self-fulfilling prophecies—the extent to which teacher perceptions predict future student achievement beyond that predicted by previous achievement and student motivation.

Moderation by ability grouping is represented by path D. This path indicates that the self-fulfilling influence of teacher perceptions on student achievement may be more or less powerful depending on the type or level of ability grouping. The relationships presented in path D are the main focus of the current study.

METHOD

Subjects

Data from 1701 students (794 male, 907 female; 1536 white, 64 African-American, 43 listing other, and 58 missing race information) and 97 teachers from public school sixth-grade math classes were used from the Michigan Study of Adolescent Life Transitions (MSALT) (Eccles, 1988). There were 341 students in

between-grouping and 478 students in within-grouping. Among students in between-grouping, 129 were in high-ability groups, 73 were in low-ability groups, and 139 were in average-ability groups. Among students in within-grouping, 176 were in high-ability groups, 80 were in low-ability groups, and 219 were in average-ability groups.

To examine self-fulfilling prophecies for whole classes, we aggregated individual student data. We averaged data from students in each class for each variable. This resulted in new mean values for each classroom for each variable. Thus, class data are not based on teachers' overall assessments of their classes. Rather, we operationalized teacher perceptions of their classes as the averages of their ratings of each student in their classes.

There were 114 classes in total. Classes with less than 6 students with valid data on all variables were excluded from the analyses, leaving 109 classes. On average, 15.05 students in each class had valid data. We excluded one additional class from the analyses because it was missing data on ability grouping. Therefore, all analyses at the class level examined data from 108 classes. Twenty-one classes used between-class grouping and 31 classes used within-class grouping.

Questionnaires

In October of sixth grade, questionnaires assessed teacher perceptions of individual students' performance, talent, and effort in math. Teachers also indicated whether they used between-class or within-class grouping, and if so, which level students were in (high, average, or low). Questionnaires also assessed student motivation, including self-perceptions of effort in math, self-concept of math ability, time spent on math homework, extrinsic value of math, and intrinsic value of math (for validity and reliability information on these scales see Eccles, Wigfield, Flanagan, Miller, Reuman, & Yee, 1989; Eccles (Parsons) et al., 1983; Eccles-Parsons, Kaczala, & Meece, 1982).

Measures of Achievement

Fifth-grade final math marks and standardized math test scores (taken in fifth grade or at the beginning of sixth grade) served as measures of previous student achievement. Sixth-grade final math marks and math scores on the Michigan Educational Assessment Program (MEAP), a standardized test taken early in the seventh grade, served as measures of future student achievement.

RESULTS

Initial Analyses

Descriptives and correlations. Means and standard deviations for all variables (for students and for classes) are presented in Table 1, along with the range and scale for each variable. To control for classroom for all individual level analyses, we subtracted the corresponding class mean for each variable from every individual student's score. Correlations for individual and class data are presented in Table 2.

TABLE 1
DESCRIPTIVE STATISTICS

Variable	Min–max value	Mean	Standard deviation
Teacher perceptions			
Performance	−2.93 to 2.86 (1–5)	.00 (3.42)	1.04 (.43)
Talent	−4.50 to 3.33 (1–7)	.00 (4.85)	1.17 (.62)
Effort	−4.06 to 3.00 (1–7)	.00 (5.08)	1.22 (.67)
Students' previous achievement			
Percentile rank in 5th grade	−73.00 to 52.86[a] (1–99)	.00[a] (60.21)	21.61[a] (13.90)
Fifth grade math marks	−7.80 to 6.31[b] (3–16)	.00 (11.39)	2.14 (1.21)
Students' future achievement			
MEAP scores	−19.35 to 9.89[c] (1–28)	.00 (22.65)	4.09 (2.35)
Sixth-grade math marks	−8.20 to 6.39[b] (3–16)	.00 (11.48)	2.18 (1.33)
Student motivation			
Time spent on homework	−2.00 to 2.26[d] (1–4)	.00 (2.46)	.79 (.38)
Effort exerted in math	−5.13 to 2.85[e] (1–7)	.00 (5.74)	1.25 (.44)
Self-concept of math ability	−8.78 to 5.63[f] (2–14)	.00 (10.17)	2.34 (.81)
Extrinsic value of math	−13.57 to 6.00[g] (3–21)	.00 (18.32)	2.94 (1.12)
Intrinsic value of math	−10.55 to 7.14[h] (2–14)	.00 (9.48)	3.18 (1.36)

Note. Data for the individual students are presented first in all columns (n = 1701). All variables at the individual level of analysis are mean-deviated. The values in parentheses in the min–max column represent the minimum and maximum values of the variables before class means were subtracted from all individual scores. Aggregate class data (n = 108) are presented in the parentheses after means and standard deviations for student data in the third and fourth columns.

[a] Values reflect percentile ranks.
[b] Higher values reflect higher marks.
[c] Higher values reflect higher MEAP scores.
[d] Higher values reflect greater time spent on homework.
[e] Higher values reflect higher self-perception of effort.
[f] Higher values reflect more favorable self-conceptions.
[g] Higher values reflect greater extrinsic value of math.
[h] Higher values reflect greater intrinsic value of math.

Overview of Main Analyses

We performed three sets of analyses in this study. In the first set of analyses, we examined how ability grouping moderated the relationship between teacher perceptions and student achievement. We examined moderation by ability grouping for students in two ways: (1) we examined how *type* of grouping (between-class, within-class, no grouping) moderated the relationship between teacher perceptions and student achievement, and (2) we examined how *level* of grouping (high-ability, low-ability, or no grouping) moderated the relationship between teacher perceptions and student achievement.

In the second set of analyses, we examined whether or not teacher perceptions mediated the relationship between level of within-class ability grouping and student achievement. We examined mediation in three steps: (1) we used level of within-class grouping to predict future student achievement in the context of a

TABLE 2

CORRELATIONS FOR INDIVIDUAL STUDENT AND CLASS DATA

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Teacher perceptions																
1. Performance		.80	.73	.61	.54	.53	.71	.14	−.15	.48	.20	.27	.00	.00	.00	.74
2. Talent	.69		.57	.64	.52	.52	.63	.09	−.18	.47	.20	.24	.00	.00	.00	.65
3. Effort	.68	.58		.45	.45	.39	.62	.18	−.11	.38	.17	.26	.00	.00	.00	.52
Achievement																
4. Fifth-grade std. test scores	.59	.56	.41		.55	.65	.63	.05	−.16	.43	.17	.24	.00	.00	.00	.49
5. Fifth-grade math marks	.55	.52	.40	.71		.49	.65	.11	−.15	.39	.15	.25	.00	.00	.00	.39
6. MEAP scores	.44	.43	.25	.77	.56		.58	.07	−.13	.34	.17	.24	.00	.00	.00	.40
7. Sixth-grade math marks	.58	.56	.56	.63	.61	.47		.17	−.17	.49	.20	.31	.00	.00	.00	.56
Student motivation																
8. Effort in math	.15	.08	.22	.09	.15	.10	.14		.12	.25	.26	.28	.00	.00	.00	.08
9. Time spent on homework	−.16	−.08	−.03	−.12	−.03	−.09	−.05	.12		−.12	.06	−.05	.00	.00	.00	−.04
10. Self-concept of math ability	.43	.38	.31	.41	.37	.40	.36	.45	−.07		.24	.44	.00	.00	.00	.41
11. Extrinsic value of math	.11	.05	.16	.09	.06	.04	.04	.32	−.04	.30		.34	.00	.00	.00	.10
12. Intrinsic value of math	−.13	−.02	.04	−.21	−.16	−.12	−.10	.23	.09	.28	.42		.00	.00	.00	.16
Grouping variables																
13. Between-grouping	.11	.05	−.01	−.03	.12	−.02	−.01	.07	−.02	.19	−.04	−.11				
14. Within-grouping	.13	.08	−.06	.07	.08	.03	.02	−.16	−.10	−.01	−.11	−.18				
15. Level of between grouping																
16. Level of within grouping																

Note. The correlations presented to the right of the diagonal are based on the mean-deviated variables (controlling for class) and correlations $\geq |.05|$ were significant at $p \leq .05$ ($n = 1701$). The correlations presented to the left of the diagonal are based on the aggregate class-level data and for these, correlations $\geq |.19|$ were significant at $p \leq .05$ ($n = 108$). We do not present correlations at the class level for level of ability grouping because there were too few classes to examine the relationship between teacher perceptions and student achievement for high-, average-, and low-ability classrooms.

model controlling for previous achievement and motivation, (2) we used level of within-class grouping to predict teacher perceptions in the context of a model controlling for previous achievement and motivation, and (3) we used teacher perceptions to predict student achievement in the context of a model controlling for previous achievement, motivation, and level of within-class grouping.

Unfortunately, we could not examine whether teacher perceptions mediated relations of between-class grouping to achievement. Analyses at the individual level required removing between-class differences. However, after controlling for classroom, the correlation between level of ability grouping (which occurs between classes) and achievement is rendered zero. Consequently, it is only meaningful to examine whether teacher perceptions mediated the relations of between-class grouping to achievement for whole classes. However, there were only 21 classes that were grouped between. Thus, there were too few classes to examine the relationship between teacher perceptions and class achievement for low-, average-, and high-ability between-grouped classrooms. We did not examine mediation at the class level for within-class grouping because it is a

within-class rather than a between-class variable. Thus, we only examined mediation for levels of within-class grouping at the individual level.

In the third set of analyses, we examined self-fulfilling prophecies, perceptual bias, and accuracy at the level of whole classes. We also examined how *type* of grouping moderated the relationship between teacher perceptions and class achievement. We did not examine moderation by *level* of grouping for whole classes. Levels of within-class grouping could not be examined because they are a within-class, rather than a between-class, variable. In addition, we could not examine levels of between-class grouping because there were too few classes for each level (high, low, average/not grouped) to perform a meaningful analysis.

Testing Moderation by Ability Grouping

Base model for moderation analyses. A series of multiple regressions examined moderation by adding product terms to a base regression model that included teacher perceptions and control variables as predictors of student achievement (main effects for teacher perceptions and control variables have been presented elsewhere—see Jussim, 1989; Jussim & Eccles, 1992—and will not be presented in any detail in this paper). For all analyses testing moderation, the base model included previous achievement (fifth-grade final math marks and standardized math test scores), student motivation (self-perception of effort, self-concept of ability, time spent on homework, extrinsic value of math, and intrinsic value of math), and teacher perceptions of performance, talent, and effort.

Individual level analyses control for classroom. For analyses at the individual level, it was necessary to remove variation among classes. Thus, we controlled for between-class variation by subtracting the corresponding class mean from each student's scores. This altered the degrees of freedom in the analyses by the number of classes we controlled for in the analyses. Thus, it was necessary to recalculate the *t* values for each significant predictor based on the correct degrees of freedom. To recalculate *t* values, we first had to recalculate the standard error of the beta of each significant predictor based on the correct degrees of freedom (the total number of students minus the total number of classes: 1701 − 108; for the formula to calculate the standard error of a beta, see Cohen & Cohen, 1983, p. 109).

Assessing moderation. We added product terms to the base model to assess moderation. To assess moderation by type of ability grouping, we created product terms to compare students (or classes) in within-class and between-class grouping to students in heterogeneous classes. To assess moderation by levels of ability grouping, we created product terms to compare students in high-ability, low-ability, and average-ability groups or who were not grouped. For all results, we report both the unstandardized regression coefficient (*b*) and the standardized regression coefficient (β) for the base and final regression models.

First, we examined whether the presence or absence of ability grouping moderated self-fulfilling prophecies. To do so, we compared relations between teacher perceptions and student achievement: (1) among students in between-

class groups to students in heterogeneous classes and (2) among students in within-class groups to students in heterogeneous classes.

We created six product terms to examine whether the presence or absence of ability grouping moderated self-fulfilling prophecies and perceptual biases. To create the product terms we multiplied each of the three teacher perception variables by the grouping variables. Thus, there were three product terms for between-class grouping and three additional product terms for within-class grouping. We then added the products terms to the base model to predict final marks and MEAP scores.

We calculated predicted values for outcome measures to understand the nature of the interactions identified by the significant product terms. We followed the method proposed by Judd and McClelland (1989) (for more detail, a Technical Appendix is available from the authors; see also Jussim et al., 1996; Madon et al., 1997 for examples).

Collinearity. Because the product terms were highly correlated, simultaneously entering them into the regression equation may underestimate each term's ability to predict future student achievement.[1] The product terms were correlated because: (1) the teacher perception variables that were used to create the product terms are correlated (see Table 2); and (2) each of the teacher perception variables was multiplied by the same grouping variables, resulting in product terms that had an even higher degree of correlation. Data analytic procedures addressed this issue in the following manner.

The analyses involved several steps. We examined individual product terms further if the block of all product terms significantly increased R^2 at $p < .10$ or if one of the individual product terms was significant at $p \leq .05$. If either of these conditions was met, we followed a two-step procedure to examine individual product terms in more detail. First, we entered the most significant product term into the regression model along with the base model variables. Second, we entered the remaining product terms into the regression model as a block. If the R^2 increment for adding the remaining product terms was at least marginally significant at $p < .10$ or if any individual product term was significant at $p < .05$, we added the remaining most significant product term to the base model. We repeated this process until the R^2 increment or the individual product terms were no longer marginally significant at $p < .10$.

Distinguishing self-fulfilling prophecy from perceptual bias. Self-fulfilling prophecies occur when perceivers' beliefs influence targets' actual behavior. Perceptual biases occur when perceivers' beliefs influence *their judgments* of

[1] We considered eliminating this problem by simply summing the three teacher perception variables, thereby providing a single index of teacher perceptions. However, LISREL analyses employing this data set have shown that the assumption that the three teacher perception variables reflect a single, underlying factor is not viable. Models making this assumption have major theoretical weaknesses, are clearly misspecified, and do not fit the data (Jussim, 1989; Jussim & Eccles, 1992). Therefore, the analyses in the current paper do not combine the three teacher perception variables to form a single scale.

targets' behavior. Thus, if perceptual biases occur, teachers' perceptions should predict their judgments of student achievement more strongly than those perceptions predict actual student achievement.

Therefore, we tested the perceptual bias hypothesis by comparing the regression coefficient relating teacher perceptions to MEAP scores to the regression coefficient relating teacher perceptions to students' final marks. The perceptual bias hypothesis is that the coefficients relating teacher perceptions to final marks (teacher judgments) will be higher than the coefficients relating teacher perceptions to MEAP scores (objective achievement that is independent of teacher judgments—see Jussim, 1989, 1991; Jussim & Eccles, 1992, for a more detailed description).

Moderation by Type of Ability Grouping for Individual Students

We first examined whether within-class grouping moderated the relationship between teacher perceptions and student achievement. The product terms for within-class grouping did not significantly predict student MEAP scores (all p's $> .12$) or final math marks (all p's $> .09$). Thus, the relationship between teacher perceptions and student achievement outcomes was similar among students who were and were not grouped within classes.

Self-fulfilling prophecies. We next examined whether between-class grouping moderated the relationship between teacher perceptions and student MEAP scores. The base model and final model predicting MEAP scores are presented in Table 3. We obtained a final model after several steps. First, we added the six product terms (representing the relationship between teacher perceptions and type of ability grouping) as a block to the base model. The product term teacher perceptions of talent $\times$ between grouping significantly predicted MEAP scores ($b = .25$, $\beta = .03$, $p < .05$). Thus, we performed a second regression model, including teacher perceptions of talent $\times$ between grouping in the base model and adding the remaining five product terms as a block to the base model. Neither the block nor any single product term added at this step significantly predicted MEAP scores.

In the final model, the product term teacher perceptions of talent $\times$ between-class grouping significantly predicted student MEAP scores ($b = .46$, $\beta = .25$, $p < .02$). Students' scores on the MEAP increased as teacher perceptions of talent increased for students in heterogeneous classes ($b = .14$, $\beta = .04$). This result is consistent with a self-fulfilling prophecy, but it is an extremely small effect. For students in between-class grouping, as teachers' perceptions of students' talent increased, students' MEAP scores decreased ($b = -.22$, $\beta = -.09$). Although the relationship is negative, it is also very small. Thus, there was statistically significant moderation, but only because small relationships between teacher perceptions and MEAP scores in heterogeneous and between-grouped classes were in opposite directions.

Perceptual bias. We next examined whether between-class grouping moderated the relationship between teacher perceptions and student marks. The base

TABLE 3
DOES BETWEEN-CLASS GROUPING MODERATE THE RELATIONSHIP BETWEEN TEACHER PERCEPTIONS OF TALENT AND STUDENT MEAP SCORES?

Predictor variable	Base model		Full model	
	b	β	*b*	β
Teacher perceptions of performance	.66	.17***	.67	.17†
Teacher perceptions of talent	.08	.02	−.78	−.23*
Teacher perceptions of effort	−.07	−.02	−.06	−.02
Previous standardized test scores	.09	.47†	.09	.46†
Fifth-grade math marks	.25	.13†	.24	.13†
Intrinsic value of math	.06	.05*	.06	.05*
Extrinsic value of math	.03	.02	.03	.02
Time spent on math homework	−.02	.00	−.02	.00
Effort exerted in math	.00	.00	−.01	.00
Self-concept of ability	−.03	−.02	−.03	−.02
Between-grouping	.00	.00	.00	.00
Within-grouping	.00	.00	.00	.00
Talent × between grouping			.46	.25*

Note. $n = 1701$. $R^2 = .471$ for the base model; $R^2 = .472$ for full model. *b* refers to the unstandardized regression coefficient, and β refers to the standardized regression coefficient.

* $p \leq .05$.
*** $p \leq .001$.
† $p \leq .0001$.

model and final model for marks are presented in Table 4. We added the six product terms as a block to the base model and found that teacher perceptions of performance × between grouping significantly predicted student marks ($b = .17$, $\beta = .03$, $p < .05$). Subsequently, we included teacher perceptions of performance × between grouping in the base model and added the remaining five product terms as a block to the base model. Neither the block of five nor any single product term significantly predicted student marks.

In the final model, the product term teacher perceptions of performance × between-class grouping significantly predicted sixth-grade final marks ($b = .19$, $\beta = .17$, $p < .05$). Teacher perceptions of performance predicted final marks more strongly among students in heterogeneous classes ($b = .54$, $\beta = .26$) than among students in between-class grouping ($b = .35$, $\beta = .17$).

To determine if this result was consistent with perceptual bias (see Jussim, 1989; Jussim & Eccles, 1992, for more details), we compared the betas relating teacher perceptions of performance and final marks to the beta relating teacher perceptions of performance and MEAP scores ($\beta = .17$; see Table 3). There was no evidence of perceptual bias among students in between-class grouping. The beta relating teacher perceptions to marks was equal to the beta relating teacher perceptions to MEAP scores (both betas = .17). However, there was some evidence consistent with perceptual bias among students in heterogeneous classes.

TABLE 4
DOES BETWEEN-CLASS GROUPING MODERATE THE RELATIONSHIP BETWEEN TEACHERS' PERCEPTIONS OF PERFORMANCE AND STUDENT FINAL MARKS?

	Base model		Full model	
Predictor variable	*b*	β	*b*	β
Teacher perceptions of performance	.50	.24†	.16	.08
Teacher perceptions of talent	.05	.03	.05	.03
Teacher perceptions of effort	.29	.16†	.30	.17†
Previous standardized test scores	.02	.19†	.02	.18†
Fifth-grade math marks	.29	.28†	.29	.28†
Intrinsic value of math	.02	.03	.02	.03
Extrinsic value of math	.00	.00	.00	.00
Time spent on math homework	−.07	−.03	−.07	−.03
Effort exerted in math	.06	.03*	.05	.03
Self-concept of ability	.08	.08†	.08	.08†
Between-grouping	.00	.00	.00	.00
Within-grouping	.00	.00	.00	.00
Performance × between grouping			.19	.17*

Note. $n = 1701$. $R^2 = .651$ for the base model; $R^2 = .652$ for full model. *b* refers to the unstandardized regression coefficient, and β refers to the standardized regression coefficient.

* $p \leq .05$.

† $p \leq .0001$.

Among students in heterogeneous classes, teacher perceptions of performance predicted marks ($\beta = .26$) more strongly than they predicted MEAP scores ($\beta = .17$).

Accuracy. Predictive accuracy refers to predictive validity without influence. The zero-order correlations between teacher perceptions and student achievement reflect the overall predictive validity of teacher perceptions, including potential sources of accuracy and influence (i.e., self-fulfilling prophecy). The regression coefficients relating teacher perceptions to student achievement represent self-fulfilling prophecy—i.e., the extent to which teacher perceptions predict changes in student achievement (which are not based on previous student achievement or motivation). Consequently, the difference between the standardized regression coefficients and the zero-order correlations is an index of the extent to which teacher perceptions predicted student achievement without influencing it (see Jussim, 1989; Jussim & Eccles, 1992, for more details).

Accuracy results are presented when product terms significantly predicted achievement. In the absence of moderation (when product terms did not significantly predict achievement): (1) the pattern of accuracy was the same regardless of whether or not students were grouped by ability; and (2) this pattern was addressed in previous studies using the same data set (see Jussim, 1989; Jussim & Eccles, 1992).

We assessed accuracy separately among students who were and were not ability-grouped. The correlation between teacher perceptions of talent and MEAP

scores among students in heterogeneous classes was .55, and the standardized regression coefficient was .04. The difference between the correlation and the regression coefficient (.51) is an index of predictive accuracy (see Jussim, 1989; Jussim & Eccles, 1992). Therefore, among students in heterogeneous classes, teacher perceptions of talent predicted MEAP scores mainly because those perceptions were accurate.

This same pattern (moderate correlations and small regression coefficients) characterized relations between teacher perceptions and student achievement among students in between-grouping and in heterogeneous classes. Thus, teacher perceptions of performance and talent predicted student achievement mostly because those perceptions were accurate for students who were and were not grouped between classes.

Moderation by Level of Ability Grouping for Individual Students

Although analyses reported thus far addressed whether ability grouping moderated the relationship between teacher perceptions and student achievement, they provide no information regarding whether differing *levels* of ability groups moderate this relationship. Thus, we have not yet determined whether teacher perceptions more strongly predict achievement among low- versus high-ability groups. The next set of analyses, therefore, examined the relationship between teacher perceptions and student achievement for different *levels* of between-class and within-class grouping.

To perform these analyses, we created product terms for different ability levels using dummy codes.[2] Because there were three ability levels (high, average, and low), two dummy variables were required for each type of grouping (Pedhazer, 1982). Thus, we created four dummy variables, two for between-class grouping and two for within-class grouping.

These analyses were identical to the types of grouping analyses (comparing between-class grouping, within-class grouping, and no grouping), except that the four grouping level variables replaced the two types of grouping variables. We calculated 12 product terms that represented the interaction between group level and teacher perceptions by multiplying the three teacher perception variables (performance, talent, and effort) by each of the four dummy codes for ability group level (two dummy codes for within-class grouping and two for between-class grouping). The two dummy codes for within-grouping will be referred to as within-class grouping var. 1 and var. 2, and the two dummy codes for between-grouping will be referred to as between-class grouping var. 1 and var. 2.

We tested these paired product terms using the same procedure as with our

[2] Students in heterogeneous classes (not grouped) and students who were in average-ability groups were combined into one category for the moderation analyses because: (1) preliminary analyses showed that relations between teacher perceptions and student achievement did not differ between these two groups; and (2) ability grouping generally affects students who are in high- or low-ability levels the most (Hoffer, 1992; Kulik & Kulik, 1982, 1987, 1992; Slavin, 1988). Thus, we compared three levels of ability groups: (1) high ability; (2) average ability and not grouped; and (3) low ability.

TABLE 5
DOES LEVEL OF WITHIN-CLASS GROUPING MODERATE THE RELATIONSHIP BETWEEN TEACHERS' PERCEPTIONS OF EFFORT AND STUDENT MEAP SCORES?

Predictor variable	Base model		Full model	
	b	β	*b*	β
Teacher perceptions of performance	.67	.17†	.69	.18†
Teacher perceptions of talent	.09	.03	.26	.08
Teacher perceptions of effort	−.07	−.02	−.18	−.05
Previous standardized test scores	.09	.47†	.09	.47†
Fifth-grade math marks	.25	.13†	.25	.13†
Intrinsic value of math	.06	.05*	.06	.05*
Extrinsic value of math	.03	.02	.03	.02
Time spent on math homework	−.01	.00	−.01	.00
Effort exerted in math	.00	.00	.00	.00
Self-concept of ability	−.03	−.02	−.04	−.02
Between variable 1	.04	.00	.07	.01
Between variable 2	.02	.00	.08	.01
Within variable 1	−.16	−.02	−.31	−.03
Within variable 2	−.09	−.01	−.32	−.04
Talent × within grouping (var. 1)			−.48	−.06*
Talent × within grouping (var. 2)			−.26	−.07
Effort × within grouping (var. 1)			.34	.04
Effort × within grouping (var. 2)			.18	.05

Note. $n = 1701$. $R^2 = .471$ for the base model; $R^2 = .476$ for full model. *b* refers to the unstandardized regression coefficient, and β refers to the standardized regression coefficient.

* $p \leq .05$.

† $p \leq .0001$.

analyses examining types of grouping, with one difference. The paired product terms were tested together for this part of the study because they were necessary to identify moderation by ability levels. We included the pairs of product terms in the final regression equation if adding them as a block significantly increased the R^2 value ($p < .05$).

Self-fulfilling prophecies. First, we examined whether level of within-class grouping moderated the relationship between teacher perceptions and student MEAP scores. The base model and final model for MEAP are presented in Table 5. We obtained a final model after several steps. First, we added the 12 product terms (representing the relationship between teacher perceptions and level of ability grouping) as a block to the base model. The paired product terms for teacher perceptions of talent × within grouping (var. 1 and var. 2) significantly predicted MEAP scores (b's = −.06, −.09, β's = .24, .16, both p's < .05). Thus, we included this pair of product terms in the base model and then added the remaining block of 10 product terms to the base model. The paired product terms for teacher perceptions of effort × within grouping (var. 1 and var. 2) also significantly predicted MEAP scores (b's = .06, .08, β's = .15, .15, both p's < .05).

We performed a third regression analysis which included the paired product terms for teacher perceptions of talent × within grouping and teacher perceptions of effort × within grouping in the base model, and then added the remaining block of eight product terms. Neither the block of 8 nor any pair of product terms significantly predicted MEAP scores. Thus, the third regression analysis yielded the final model.

In the final model, the paired product terms for teacher perceptions of talent × within grouping (var. 1 and var. 2) and for teacher perceptions of effort × within grouping (var. 1 and var. 2) significantly predicted student MEAP scores (R^2 increase at $p < .01$ and $p < .05$, respectively). Teacher perceptions of talent predicted MEAP scores most strongly among students in low-ability within-class grouping ($b = 1.00$, $\beta = .21$). The results are consistent with a self-fulfilling prophecy among students in low-ability within-class grouping. Teacher perceptions of talent weakly predicted MEAP scores among students who were not grouped ($b = .00$, $\beta = .07$) and among students in high-ability within-class grouping ($b = -.22$, $\beta = -.06$).

Teacher perceptions of effort predicted MEAP scores most strongly among students in low-ability within-class grouping ($b = -.70$, $\beta = -.21$). However, this relationship was negative. Thus, as teacher perceptions of effort increased among students in low-ability groups, MEAP scores decreased. This result is not consistent with a self-fulfilling prophecy. Teacher perceptions of effort weakly predicted MEAP scores for students in high-ability within-class grouping, consistent with a self-fulfilling prophecy ($b = .16$, $\beta = .05$). Among these students, increased teacher perceptions were associated with a slight increase in MEAP scores. However, the effect size is practically zero. Finally, teacher perceptions of effort did not predict MEAP scores among students who were not grouped ($b = 0$, $\beta = 0$).

We also examined whether level of between-class grouping moderated the relationship between teacher perceptions and student MEAP scores. The product terms representing the relationship between teacher perceptions and levels of between-class grouping did not significantly predict MEAP scores (all p's $> .07$). Thus, the relationships between teacher perceptions and student MEAP scores were similar among different levels of between-class grouping.

Perceptual bias. We next examined whether level of within-class and between-class grouping moderated the relationship between teacher perceptions and student marks. None of the product terms including teacher perceptions and ability group level significantly predicted final marks in sixth grade. The relationship between teacher perceptions and student final marks was the same regardless of grouping practices (R^2 change adding the 12 product terms, $p > .81$).

Accuracy. To assess accuracy in this section, we examined the relationship between teacher perceptions and student achievement separately for each of the three levels of ability grouping (high, average/not grouped, low). The difference between the correlation and regression coefficient for teacher perceptions of talent

and MEAP scores was .30 among students in low-ability within-grouping, .17 among students who were not grouped, and .47 among students in high-ability within-grouping. This indicates that teachers were more accurate for students in high-ability within-grouping than they were for students in low-ability within-grouping or for students who were not grouped.

The relationship between teacher perceptions of effort and MEAP scores for student in high ability within-class grouping reflected mostly accuracy (.28) and very little self-fulfilling prophecy (.05). For the remaining groups, there was minimal evidence of predictive accuracy for teacher perceptions of effort.

Do Teacher Perceptions Mediate the Relationship between Grouping Level and Achievement?

One of the traditional concerns regarding ability grouping is that it widens the achievement gap between students in low- and high-ability groups (Gamoran, 1992; Oakes, 1985, 1987). We wondered if this occurred among the students in our sample. Thus, we examined the correlations among level of within-class grouping and student marks in math. We found that level of within-class grouping correlated more highly with final math marks in sixth grade ($r = .56$, $p < .01$) than with final math marks in fifth grade ($r = .39$, $p < .01$). These correlations suggest that the gap in achievement among students in low- and high-ability within-class groupings increased over the course of the year.

To further investigate if this were true, we performed a 2×2 repeated measures ANOVA. The between factor in this analysis was level of within-class grouping (low or high) and the within factor was final marks in math (in fifth grade and in sixth grade). There was a significant interaction between level of within-class grouping and student marks in math ($F(2, 472) = 4.35; p = .01$).

Students in high-ability within-grouping received higher marks, whereas students in low-ability within-grouping received lower marks in sixth-grade math than they had in fifth-grade math. Among students in low-ability within-grouping, the mean for final math marks in fifth-grade was 9.88 (equivalent to a C+) and the mean for final math marks in sixth grade was 9.34 (closer to a C). Among students in high-ability within-grouping, the mean for final math marks in fifth grade was 12.85 (almost a B+) and the mean for final math marks in sixth grade was 13.29 (slightly above a B+). These results demonstrates that the gap in achievement among students in low- and high-ability within-grouping did indeed widen (about one full point on the marks scale) over the course of the year.

We next examined whether this widening effect was due to teacher perceptions with a series of mediational analyses (see Baron & Kenny, 1986). All of the following analyses controlled for previous student achievement, motivation, and classroom. First, we examined whether level of within-grouping predicted final marks in sixth-grade math. It did ($b = .63$, $\beta = .23$, $p < .01$). In particular, students in high-ability within-class grouping received higher marks than students in low-ability within-grouping, despite controlling for previous achievement and student motivation. Table 6 presents these results.

TABLE 6
DOES LEVEL OF WITHIN-CLASS GROUPING PREDICT FUTURE STUDENT ACHIEVEMENT OR TEACHER PERCEPTIONS OF PERFORMANCE, TALENT, AND EFFORT?

	Future achievement				Teacher perceptions					
	6th final marks $R^2 = .567$		MEAP scores $R^2 = .433$		Performance $R^2 = .636$		Talent $R^2 = .563$		Effort $R^2 = .335$	
Predictor variable	*b*	β	*b*	β	*b*	β	*b*	β	*b*	β
Previous standardized test scores	.03	.33†	.09	.51†	.01	.26†	.01	.29†	.01	.11*
Fifth-grade math marks	.24	.27†	.23	.13**	.03	.06	.05	.09*	.07	.13**
Intrinsic value of math	.02	.03	.01	.01	.00	−.01	−.01	−.03	.01	.02
Extrinsic value of math	−.01	−.01	.03	.02	−.01	−.02	.02	.06	−.01	−.03
Time spent on homework	−.07	−.03	−.01	−.00	−.03	−.02	−.10	−.07*	−.13	−.08*
Effort exerted in math	.02	.01	−.09	−.03	.01	.02	−.06	−.06	.10	.10*
Self-concept of ability	.12	.14***	.09	.05	.05	.13***	.07	.15***	.03	.07
Level of grouping	.63	.23†	.45	.08	.73	.54†	.70	.42†	.64	.38†

Note. $n = 475$. *b* refers to the unstandardized regression coefficient and β refers to the standardized regression coefficient.

* $p \leq .05$.
** $p \leq .01$.
*** $p \leq .001$.
† $p \leq .0001$.

Second, we examined whether level of within-grouping predicted teacher perceptions of student performance, talent, and effort. It did (*b*'s = .73, .70, and .65; β's = .54, .42, and .38 respectively; all *p*'s < .01). This result demonstrates that teachers gave better ratings to students in high-ability within-grouping than to students in low-ability within-grouping, even after controlling for students' achievement and motivation (see Table 6).

The results thus far have demonstrated that: (1) teachers perceived greater differences in achievement between students in high- and low-ability groups than really existed (prior to sixth grade); and (2) the difference in marks between students in high- and low-ability groups actually increased from fifth to sixth grade. The final set of analyses, therefore, examined the extent to which teacher perceptions accounted for (via mediation) the increasing differences between students in high- and low-ability groups.

To demonstrate mediation, teacher perceptions should predict student achievement beyond that predicted by previous achievement, student motivation, and level of within-grouping. Therefore, we performed a regression analysis in which student achievement and motivation, level of within-class grouping, and teacher

TABLE 7
DO TEACHER PERCEPTIONS MEDIATE THE RELATIONSHIP BETWEEN LEVEL OF WITHIN-CLASS GROUPING AND FUTURE STUDENT ACHIEVEVMENT?

	Sixth-grade final marks $R^2 = .631$	
Predictor variable	*b*	β
Teacher perceptions of performance	.28	.14*
Teacher perceptions of talent	.23	.14**
Teacher perceptions of effort	.30	.19†
Previous standardized test scores	.02	.23†
Fifth-grade math marks	.20	.22†
Intrinsic value of math	.02	.04
Extrinsic value of math	−.01	−.01
Time spent on math homework	.00	.00
Effort exerted in math	.00	.00
Self-concept of ability	.08	.09**
Level of within-class grouping	.07	.02

Note. $n = 475$. *b* refers to the unstandardized regression coefficient and β refers to the standardized regression coefficient.

* $p \leq .05$.

** $p \leq .01$.

† $p \leq .0001$.

perceptions predicted final marks in math. The results we found were consistent with mediation and are presented in Table 7.

In this analysis, teacher perceptions of performance, talent, and effort significantly predicted student final math marks in sixth-grade (*b*'s = .28, .23, and .30; β's = .14, .14, and .19, respectively; all *p*'s < .05). In addition, the coefficient relating level of within-class grouping to final grades was not only a nonsignificant predictor of final marks, but was trivially small ($b = .07$, $\beta = .02$, n.s.). These results are consistent with the conclusion that teacher perceptions completely mediated the relationship between level of within-grouping and student marks. We do not present results for MEAP as an achievement outcome because level of within-grouping did not significantly predict student MEAP scores.

Relationship between Teacher Perceptions and Whole-Class Achievement

The results from the previous analyses demonstrated moderation by both types and levels of ability grouping for individual students. To examine the relationship between teacher perceptions and individual student achievement within classes, we removed variation due to classroom. However, differences in teacher perceptions of whole classes are also a potentially important source of self-fulfilling prophecies. The next set of analyses examined the relationship between teacher perceptions and class achievement. We aggregated individual student data within classes to obtain average scores by classroom on each variable (see Kenny et al., in press). Thus, the unit of analysis in the next set of analyses was the class, rather

TABLE 8
WHAT IS THE RELATIONSHIP BETWEEN TEACHERS' PERCEPTIONS AND CLASS ACHIEVEMENT?

	Dependent variable			
	Sixth-grade math marks		MEAP scores	
Predictor variable	*b*	β	*b*	β
Teacher perceptions of performance	.14	.05	.03	.01
Teacher perceptions of talent	.18	.08	.06	.02
Teacher perceptions of effort	.55	.27**	−.37	−.11
Previous standardized test scores	.03	.27*	.13	.75†
Fifth-grade math marks	.25	.23*	.06	.03
Time spent on math homework	.03	.01	.02	.00
Effort exerted in math	.03	.01	.11	.02
Self-concept of ability	.08	.05	.32	.11
Extrinsic value of math	−.08	−.07	−.16	−.07
Intrinsic value of math	.00	.00	.07	.04

Note. $n = 108$. $R^2 = .549$ for marks; $R^2 = .611$ for MEAP scores. *b* refers to the unstandardized regression coefficient and β refers to the standardized regression coefficient.

* $p \leq .05$.

** $p \leq .01$.

† $p \leq .0001$.

than the individual student. The base model (see Table 8) for these analyses was identical to the base model for the student analyses with the exception that all variables are now in terms of class means and not individual scores.

Self-fulfilling prophecies. Teacher perceptions of performance, talent, and effort did not significantly predict mean class MEAP scores (all p's $> .25$). These results offer no evidence of self-fulfilling prophecies at the class level (see Table 8).

Perceptual biases. Teacher perceptions of class effort predicted final marks in sixth-grade math ($\beta = .27, p < .01$). This result is consistent with the perceptual bias hypothesis because teachers' perceptions predicted the marks teachers assigned but did not predict actual class achievement on the MEAP (see Table 8).

Accuracy. The zero-order correlations between teacher perceptions of talent and performance and marks and standardized test scores ranged from about .4 to .6. The nonsignificant path coefficients ranged from .01 to .08. This means that teacher perceptions of talent and performance predicted whole-class achievement almost entirely because those perceptions were accurate.

Moderation by Types of Ability Grouping for Whole Classes

Although we found no evidence of self-fulfilling prophecies at the class level, it was still possible that class-level self-fulfilling prophecies occurred among some types of classes. Next, therefore, a series of regression analyses examined whether self-fulfilling prophecies occurred when classes were or were not grouped.

TABLE 9
DOES WITHIN-CLASS GROUPING MODERATE THE RELATIONSHIP BETWEEN TEACHERS' PERCEPTIONS OF EFFORT AND CLASS MEAP SCORES?

	Base model		Full model	
Predictor variable	*b*	β	*b*	β
Teacher perceptions of performance	.11	.02	−.11	−.02
Teacher perceptions of talent	.07	.02	−.03	−.01
Teacher perceptions of effort	−.42	−.12	1.60	.46
Previous standardized test scores	.12	.73†	.13	.74†
Fifth-grade math marks	.08	.04	.14	.07
Intrinsic value of math	.05	.03	.05	.03
Extrinsic value of math	−.16	−.07	−.22	−.10
Time spent on math homework	.01	.00	−.15	−.02
Effort exerted in math	.07	.01	.07	.01
Self-concept of ability	.36	.12	.43	.15
Between grouping	−.21	−.04	−.31	−.05
Within grouping	−.21	−.04	5.53	1.07
Effort × within grouping			−1.12	−1.22*

Note. $n = 108$. $R^2 = .613$ for the base model; $R^2 = .630$ for full model. *b* refers to the unstandardized regression coefficient and β refers to the standardized regression coefficient.

* $p \leq .05$.

† $p \leq .0001$.

The first set of analyses examined whether within-class grouping moderated the relationship between teacher perceptions and student achievement at the class level. The second set of analyses examined whether between-class grouping moderated the relationship between teacher perceptions and student achievement at the class level. All analyses were identical to those used in the moderation analyses for individual students, with the following exceptions: (1) all variables in the whole class analyses were class means, instead of individual scores, (2) we examined the relationship between teacher perceptions and class achievement in the presence or absence of ability grouping, and (3) we created six new product terms to represent the relationship between class level teacher perceptions and type of grouping.

Moderation by within-grouping. First, we examined whether within-grouping moderated the relationship between teacher perceptions and class MEAP scores. Results are presented in Table 9. We obtained a final model after several steps. First, we added the six product terms to the base model. The product term teacher perceptions of effort × within grouping significantly predicted class MEAP scores ($b = -1.22$, $\beta = .01$, $p < .05$). Second, we performed a regression analysis including teacher perceptions of effort × within grouping in the base model and adding the remaining block of five product terms to the base model. Neither the block nor any single product term significantly predicted MEAP scores.

TABLE 10
DOES BETWEEN-CLASS GROUPING MODERATE THE RELATIONSHIP BETWEEN TEACHERS' PERCEPTIONS OF PERFORMANCE AND CLASS FINAL MARKS?

	Base model		Full model	
Predictor variable	*b*	β	*b*	β
Teacher perceptions of performance	.18	.06	−1.69	−.55
Teacher perceptions of talent	.19	.09	.23	.11
Teacher perceptions of effort	.52	.26*	.50	.25*
Previous standardized test scores	.02	.24*	.02	.24*
Fifth-grade math marks	.27	.24*	.33	.30**
Intrinsic value of math	−.01	−.01	−.04	−.04
Extrinsic value of math	−.08	−.07	−.09	−.08
Time spent on math homework	.03	.01	.10	.03
Effort exerted in math	.02	.01	−.01	.00
Self-concept of ability	.11	.07	.15	.09
Between grouping	−.19	−.06	−3.94	−1.17*
Within grouping	−.06	−.02	−.08	−.03
Performance × between grouping			1.10	1.31*

Note. $n = 108$. $R^2 = .552$ for the base model; $R^2 = .578$ for full model. *b* refers to the unstandardized regression coefficient and β refers to the standardized regression coefficient.

* $p \leq .05$.

** $p \leq .01$.

In the final model, the product term for teacher perceptions of effort × within grouping significantly predicted MEAP scores ($b = -1.12, \beta = -1.22, p < .05$). More specifically, teacher perceptions of effort predicted average class MEAP scores among within-groups ($b = .48, \beta = .14$). Among within-groups, as teacher perceptions of effort increased, MEAP scores also increased. This is consistent with a self-fulfilling prophecy. There was a negative relationship between teacher perceptions of effort and MEAP scores among classes that did not use within-grouping ($b = -.64, \beta = -.18$). Thus, as teacher perceptions of effort increased, MEAP scores decreased. This is not consistent with a self-fulfilling prophecy.

Moderation by between-grouping. Next we assessed whether between-grouping moderated the relationship between teacher perceptions and final marks in sixth-grade math. Results are presented in Table 10. We first added the block of six product terms to the base model. Teacher perceptions of performance × between grouping significantly predicted class marks ($b = 1.31, \beta = .02, p < .05$). We then performed a regression analysis including teacher perceptions of performance × between grouping in the base model and adding the remaining block of five product terms. Neither the block nor any single product term significantly predicted class marks.

Thus, in the final model, the product term teacher perceptions of performance × between grouping significantly predicted classes' final marks in math ($b = 1.10, \beta = 1.31, p < .02$). Among heterogeneous classes, teacher percep-

tions of performance predicted classes' final marks in sixth grade ($b = .51$, $\beta = .16$). Class final marks increased as teacher perceptions of performance increased for heterogeneous classes. This is consistent with perceptual bias because teacher perceptions of performance predicted final marks, but not MEAP scores. Among between-grouped classes, as teacher perceptions of performance increased, class marks decreased ($b = -.59$, $\beta = -.19$). This result is not consistent with perceptual bias.

Accuracy. We assessed accuracy separately for grouped and heterogeneous classes. Teacher perceptions of performance predicted marks primarily because those perceptions were accurate. When classes were between-grouped, the correlation relating teacher perceptions of performance to marks was .76; the regression coefficient was −.19. This means that teachers were even more accurate than indicated by the already-high correlation. For heterogeneous classes, the correlation with marks was .53; the regression coefficient was .16. In these classes, accuracy accounted for about two-thirds of the relation between teacher perceptions of performance and final marks and perceptual bias accounted for about one-third. Teacher perceptions of effort were not particularly accurate predictors of marks regardless of grouping status.

DISCUSSION

Our research sought to identify conditions under which self-fulfilling prophecies and perceptual biases might be more powerful. We found that both type and level of ability grouping moderated the relationship between teacher perceptions and student achievement. We found: (1) results consistent with perceptual biases for students in heterogeneous classes and for heterogeneous classes; and (2) results consistent with self-fulfilling prophecies among students in low-ability within-grouping and among classes that used within-grouping. In this study we examined the institutional labeling of certain students as high ability and others as low ability—but we did not find evidence of powerful self-fulfilling prophecies or perceptual biases. These results are consistent with a mounting body of literature demonstrating that self-fulfilling prophecies are rarely very powerful (Cooper & Hazelrigg, 1988; Jussim & Eccles, 1995; Raudenbush, 1984; Rosenthal & Rubin, 1978).

At the individual level, teacher perceptions mediated the relationship between level of ability grouping and student achievement. The achievement gap among students in low- and high-ability within-grouping increased over the course of the sixth-grade year, and teacher perceptions almost completely accounted for this pattern of increasing differences.

Moderation by Ability Grouping for Individual Students and Whole Classes

A traditional concern regarding between-grouping is that it may lead teachers to develop inaccurate perceptions (Brophy & Good, 1974; Oakes, 1985, 1987; Schaeffer & Olexa, 1971). However, our results indicated that the pattern of small

self-fulfilling prophecies and high accuracy was no different between students in heterogeneous classes and those in between-grouping. Furthermore, among students in between-grouping and among between-grouped classes, there was no evidence of perceptual bias, whereas there was evidence of perceptual bias among students in heterogeneous classes. Thus, despite its reputation (Oakes, 1985, 1987), between-grouping may actually help reduce self-fulfilling prophecies.

Between-grouping may reduce perceptual bias for two reasons: (1) there is less variability in the achievement histories of students in grouped classes; and (2) the groups are clearly delineated as including students with high-, average-, or low-achieving histories. These aspects simplify the cognitive environment, reducing the amount of information that teachers have to process. Thus it should be much more difficult for teachers to develop wildly inaccurate perceptions of students in between-grouping.

We found results consistent with a self-fulfilling prophecy among students in low-ability within-grouping and among classes that used within-grouping. Teacher perceptions were less accurate for students in low-ability within-grouping than among students in high-ability within-grouping. If teachers spend disproportionate time with highs, it will take away from the time they spend with lows, thereby decreasing the accuracy of teacher perceptions of students in low-ability groups. This analysis is consistent with research demonstrating that teachers interact more with highs and provide them with greater opportunities to demonstrate their knowledge than lows (Brophy & Good, 1974; Brophy, 1983; Slavin, 1988). Furthermore, the more time perceivers spend with targets, the more accurate they become and the less likely they are to create self-fulfilling prophecies (Fiske & Neuberg, 1990; Raudenbush, 1984; Swann & Ely, 1984).

Self-fulfilling prophecies may also have been stronger for students in low-ability within-grouping because within-grouping may promote a differential learning environment. We found that differences in marks among students in low- and high-ability within-grouping increased over the course of the year almost entirely because of teacher perceptions. Within-grouping may reflect a classroom structure in which teachers treat high- and low-expectancy students quite differently. It is precisely these types of classrooms in which self-fulfilling prophecies are strongest (Brattesani, Weinstein, & Marshall, 1984; Jussim, 1986).

Limitations

This study has several important limitations. First, although the data set includes 12 different school districts, it is specific to one area. Further research is necessary to assess if our results generalize to other grade levels, geographic areas, or grouping practices.

Omitted variable problem. Another limitation involves the possibility of omitted variables. Regression coefficients can be interpreted as representing causal effects only when the model includes all relevant variables. If a third variable is responsible for the link between teacher perceptions and student/class

achievement and it is omitted from the model, then the regression coefficients relating teacher perceptions to student/class achievement are inflated.

However, few naturalistic studies have used both achievement test scores and final marks as measures of previous achievement. Furthermore, few studies have used more than one measure of student motivation—we used five (Jussim, 1989; Jussim et al., 1996). The use of these extensive controls greatly reduces the chance that we omitted a relevant variable from the analyses.

Of course, it is still possible that we omitted an important predictor of teacher perceptions and student achievement. If so, our analyses may have overestimated the relationship between teacher perceptions and student achievement. Perhaps teachers were more accurate and self-fulfilling prophecies were smaller than even we have concluded!

Time frame of assessment. It is important to consider the time frame in which teacher perceptions and student achievement were assessed. Perceivers' judgments and decisions may change, depending upon whether they are judging or deciding about events in the near future versus the distant future (Gilovich, Kerr, & Medvec, 1993; Liberman & Trope, in press). Had teacher perceptions been assessed earlier or later in the year, the relationship between teacher perceptions and student achievement might have been different from what we found. In fact, experimental manipulations of teacher perceptions late in the school year rarely produce self-fulfilling prophecies (Raudenbush, 1984). Thus, our results are specific to relations among teacher perceptions assessed early in the school year and student achievement assessed at the end of the school year or at the beginning of the subsequent school year. In interpreting our results, one should not conclude that the relations among teacher perceptions and student achievement would be the same if teacher perceptions or student achievement were assessed at times of year different from those we studied.

Teacher perceptions. In this study, we operationalized predictive accuracy as the extent to which teacher perceptions early in the year predicted, without causing, student achievement at the end of the year. However, we did not ask teachers to explicitly predict future student achievement. Rather, we asked them to evaluate current student performance, talent, and effort in math. Thus, our results are best interpreted as showing that perceptions generally have small self-fulfilling effects on target behavior and are usually quite accurate. This would leave open the possibility that predictions would relate to target behavior in different ways (e.g., stronger, weaker) or for different reasons (e.g., stronger or weaker self-fulfilling prophecies, greater or lesser accuracy).

We are aware of only one study, however, that directly compared perceivers' evaluations and predictions of targets (Kahneman & Tversky, 1973). Perceivers received information describing students as "intelligent, self-confident, well-read, hardworking, and inquisitive" and then evaluated their ability and predicted their college GPA. Thus, much like teachers in elementary school, Kahneman and Tversky's (1973) perceivers had the opportunity to base their evaluations and predictions on students' prior accomplishments. Their results showed that the

predictions and evaluations were virtually identical. Although this is only a single study conducted over 25 years ago, it suggests that there are unlikely to be dramatic differences between teacher perceptions of students and teacher predictions for students. At minimum, it provides no basis for expecting dramatic differences between teacher perceptions and teacher predictions.

In addition, research on self-fulfilling prophecies has often examined perceptions rather than predictions (e.g., Hilton & Darley, 1985; Neuberg, 1989; Snyder, Tanke, & Berscheid, 1977). For example, Merton's (1948) classic example of a self-fulfilling prophecy involved a bank failure. Rumors that the town bank was insolvent (not that it would be insolvent sometime in the future) spurred depositors to rush to the bank to withdraw their money. The false belief that the bank was insolvent led to its insolvency. In addition, even the titles of major reviews of self-fulfilling prophecies clearly convey that researchers construe the issue as going well beyond simple predictions. For example, the title of Snyder's (1984) classic review of self-fulfilling prophecies was "When *Belief* Creates Reality" (emphasis added); and the title of Jussim's (1991) review was "Social *Perception* and Social Reality" (emphasis added). These examples demonstrate that, since Merton (1948) first coined the term "self-fulfilling prophecy," it has been common for work in this area to examine perceivers' current beliefs and perceptions.

In addition, the pattern of results we found is similar to that found in both experimental and naturalistic studies that have examined the relationship between teachers' expectations and students' future performance (see reviews by Jussim & Eccles, 1995; see meta-analyses by Raudenbush, 1984; Rosenthal & Rubin, 1978). Indeed our results suggesting the occurrence of small self-fulfilling prophecies are also consistent with the results of most experimental research outside the teacher–student context (see Jussim, 1991; Jussim et al., 1996, for reviews). Thus, existing empirical research on expectancies and self-fulfilling prophecies provides no basis for believing that there are important differences between perceptions of targets and expectations for targets or for believing that the pattern of results we found greatly depends on the type of judgement assessed. However, this conclusion will warrant revision if future research empirically demonstrates important differences between perceptions and predictions.

We also did not directly assess teachers' overall perceptions of their classes in this study. Rather, the analyses examining whole class effects of teacher perceptions are based on aggregate individual data. It may be that group impressions differ from simply averaging individual impressions because the processes underlying impression formation for individuals and groups sometimes differ (Hamilton & Sherman, 1996). Therefore, future research on group-level self-fulfilling prophecy effects should directly assess group-level perceptions.

CONCLUSION

Self-fulfilling prophecies were once believed to exert a powerful and pervasive influence on social reality, and bias and inaccuracy were believed to dominate

social perception (e.g., Devine, 1995; Fiske & Taylor, 1984; Gilbert, 1995; Jones, 1986, 1990; Snyder, 1984). Unfortunately, existing empirical evidence does not support such a conclusion. Instead, existing evidence (e.g., Funder, 1995; Jussim, 1991; Jussim et al., 1996; Kenny, 1994; Madon et al., 1997; Raudenbush, 1984; Rosenthal & Rubin, 1978; Smith, Jussim, & Eccles, 1998) does support the following conclusions: (1) Self-fulfilling prophecies are typically small, although there are some conditions under which they are powerful. (2) Perceivers' beliefs generally predict targets' behavior far more because perceivers are accurate than because of self-fulfilling prophecies. (3) Even teacher perceptions of students from different demographic groups tend to be quite accurate (i.e., in classrooms, there is little evidence of rampant biases produced by stereotypes). (4) Self-fulfilling prophecies may be more likely to help than to harm students. (5) The effects of self-fulfilling prophecies are more likely to dissipate over time than to accumulate over time. To this list, the current research can add the following: (6) Although ability grouping does moderate self-fulfilling prophecies, even the largest effects are not all that powerful. (7) Self-fulfilling prophecies do seem to increase differences between students in high- and low-ability within-grouping. (8) Self-fulfilling prophecies for whole classes are not any larger than self-fulfilling prophecies for individuals (they are typically small).

Social and cognitive psychologists have identified a slew of errors and biases in social perception and social judgment. And at least sometimes, those errors may lead to profound and damaging consequences. In contexts such as the classroom, however, where perceivers have a wealth of contact with targets, and rich, highly diagnostic, objective information about them, accuracy, far more than error, bias, and self-fulfilling prophecy, seems to characterize social perception. This leaves open the possibility that error, bias, and self-fulfilling prophecy *do* characterize social interaction in naturally occurring contexts in which perceivers do not have such a wealth of information about targets. Of course, the possibility that such contexts exist provides no evidence that they *actually* exist. Identifying conditions under which naturally occurring social interaction is characterized by widespread error and bias, and by powerful self-fulfilling prophecies, remains a challenge for future research.

REFERENCES

Allport, G. (1954). *The nature of prejudice.* Cambridge, MA: Addison–Wesley.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology,* **51,** 1173–1182.

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.

Brewer, M. B. (1988). A dual-process model of impression formation. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Advances in social cognition* (pp. 1–36). Hillsdale, NJ: Erlbaum.

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

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

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

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

Darley, J., & Fazio, R. (1980). Expectancy confirmation processes arising in the social interaction sequence. *American Psychologist,* **35,** 867–881.

Devine, P. G. (1995). Prejudice and out-group perception. In A. Tesser (Ed.), *Advanced social psychology* (pp. 467–524). New York: McGraw–Hill.

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

Eccles (Parsons), J., Adler, T., Futterman, R., Goff, S. B., Kaczala, C., Meece, J., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. Spence (Ed.), *Achievement and achievement motivation* (pp. 75–146). San Francisco, CA: Freeman.

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

Eccles, J., & Wigfield, A. (1985). Teacher perceptions and student motivation. In J. Dusek (Ed.), *Teacher expectancies* (pp. 185–226). Hillsdale, NJ: Erlbaum.

Eccles, J., Wigfield, A., Flanagan, C., Miller, C., Reuman, D., & Yee, D. (1989). Self-concepts, domain values, and self-esteem: Relations and changes at early adolescence. *Journal of Personality,* **57,** 283–310.

Eden, D. (1990). Pygmalion without interpersonal contrast effects: Whole groups gain from raising manager perceptions. *Journal of Applied Psychology,* **75,** 394–398.

Eiser, J. R. (1971). Enhancement of contrast in the absolute judgement of attitude statements. *Journal of Personality and Social Psychology,* **17,** 1–10.

Eiser, J. R. (1990). *Social judgement.* Pacific Grove, CA: Brooks/Cole.

Feldhusen, J., and Moon, S. (1992). Grouping gifted students: Issues and concerns. *Gifted Child Quarterly,* **36,** 63–67.

Fiske, S. T., & Neuberg, S. (1990). A continuum of impression formation, from category based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Psychology,* **23,** 1–74.

Fiske, S. T., & Taylor, S. E. (1984). *Social Cognition.* Reading, MA: Addison–Wesley.

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

Funder, D. C. (1995). On the accuracy of personality judgement: A realistic approach. *Psychological Review,* **102,** 652–670.

Gamoran, A. (1992). Is ability grouping equitable? *Educational Leadership,* **50,** October, 11–17.

Gilbert, D. T. (1995). Attribution and interpersonal perception. In A. Tesser (Ed.), *Advanced social psychology* (pp. 99–147). New York: McGraw–Hill.

Gilovich, T., Kerr, M., & Medvec, V. H. (1993). Effects of temporal perspective on subjective confidence. *Journal of Personality and Social Psychology,* **64,** 552–560.

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review,* **103,** 336–355.

Harris, M. J. (1989). Personality moderators of expectancy effects: Replication of Harris and Rosenthal (1986). *Journal of Research in Personality,* **23,** 381–387.

Hilton, J., & Darley, J. M. (1985). Constructing other persons: A limit on the effect. *Journal of Experimental Social Psychology,* **21,** 1–18.

Hoffer, T. (1992). Middle school ability grouping and student achievement in science and mathematics. *Educational Evaluation and Policy Analysis,* **14,** 205–227.

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

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

Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model-comparison approach.* Orlando, FL: Harcourt Brace Jovanovich.

Jussim, L. (1986). Self-fulfilling prophecies: A theoretical and integrative review. *Psychological Review,* **93,** 1–17.

Jussim, L. (1989). Teacher perceptions: 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. (1993). Accuracy in interpersonal perceptions: A reflection-construction analysis of current and classic research. *Journal of Personality,* **61,** 637–668.

Jussim, L., & Eccles, J. (1992). Teacher perceptions. II. Construction and reflection of student achievement. *Journal of Personality and Social Psychology,* **63,** 947–961.

Jussim, L., & Eccles, J. (1995). Naturally occurring interpersonal expectancies. *Review of Personality and Social Psychology,* **15,** 74–108.

Jussim, L., Eccles, J., and Madon, S. (1996). Social perception, social stereotypes, and teacher perceptions: Accuracy and the search for the powerful self-fulfilling prophecy. In M. P. Zanna (Ed.), *Advances in experimental social psychology,* (pp. 281–388). New York: Academic Press.

Jussim, L., and Fleming, C. (1996). Self-fulfilling prophecies and the maintenance of social stereotypes: The role of dyadic interactions and social forces. In N. Macrae, C. Stangor, and M. Hewstone (Eds.), *The foundations of stereotypes and stereotyping* (pp. 161–192). New York: Guilford.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review,* **80,** 237–251.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis.* New York: Guilford.

Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin,* **102,** 390–402.

Kenny, D. A., Kashy, D., & Bolger, N. (in press). Data analysis in social psychology. In D. Gilbert, S. T. Fiske, & J. Lindzey (Eds.), *Handbook of social psychology* (4th ed.). New York: McGraw–Hill.

Kulik, C., and Kulik, J. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal,* **19,** 415–428.

Kulik, J., and Kulik, C. (1987). Effects of ability grouping on student achievement. *Equity and Excellence,* **23,** 23–30.

Kulik, J., and Kulik, C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly,* **36,** 73–77.

Krueger, J., & Rothbart, M. (1990). Contrast and accentuation effects in category learning. *Journal of Personality and Social Psychology,* **59,** 651–663.

Krueger, J., Rothbart, M., & Sriram, N. (1989). Category learning and change: Differences in sensitivity to information that enhances or reduces intercategory distinctions. *Journal of Personality and Social Psychology,* **56,** 866–875.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review,* **103,** 284–308.

Liberman, N., & Trope, Y. (1998). The role of feasibility and desirability considerations in near and distant future decisions: A test of temporal construct theory. *Journal of Personality and Social Psychology,* **75,** 5–18.

Madon, S., Jussim, L., & Eccles, J. (1997). In search of the powerful self-fulfilling prophecy. *Journal of Personality and Social Psychology,* **72,** 791–809.

McCauley, C. (1995). Are stereotypes exaggerated? Looking for perceptual contrast and cognitive miser in beliefs about group differences. In Y. T. Lee, L. Jussim, & C. R. McCauley (Eds.), *Stereotype accuracy: Toward appreciating group differences* (pp. 215–243). Washington, D.C.: Amer. Psychol. Assoc.

Merton, R. (1948). The self-fulfilling prophecy. *Antioch Review,* **8,** 193–210.

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

Neuberg, S. L. (1989). The goal of forming accurate impressions during social interactions: Attending to the impact of negative expectancies. *Journal of Personality and Social Psychology,* **56,** 374–386.

Neuberg, S. L. (1994). Expectancy-confirmation processes in stereotype-tinged social encounters: The moderating role of social goals. In M. P. Zanna & J. M. Olson (Eds.), *The psychology of prejudice: The Ontario symposium* (Vol. 7). Hillsdale, NJ: Erlbaum.

Oakes, J. (1985). *Keeping track: How schools structure inequality,* New Haven: Yale Univ. Press.

Oakes, J. (1987). Curriculum inequality and school reform. *Equity and Excellence,* **23,** 8–14.

Pedhazer, E. (1982). *Multiple regression in behavioral research: Explanation and prediction,* 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 induction: A synthesis of findings from 18 experiments. *Journal of Educational Psychology,* **76,** 85–97.

Rist, R. (1970). Student social class and teacher perceptions: The self-fulfilling prophecy in ghetto education. *Harvard Educational Review,* **40,** 411–451.

Rosenthal, R., and Rubin, D. (1978). Interpersonal expectancy effects: The first 345 studies. *The Behavioral and Brain Sciences,* **3,** 377–415.

Schaeffer, W., and Olexa, C. (1971). *Tracking and opportunity: The locking-out process and beyond.* Scranton, PA: Chandler.

Slavin, R. (1988). Synthesis of research on grouping in elementary and secondary schools. *Educational Leadership,* **46,** 67–77.

Smith, A., Jussim, L., & Eccles, J. (1998). *Do self-fulfilling prophecies accumulate, dissipate, or remain stable over time?* Manuscript submitted for publication.

Snyder, M. (1984). When belief creates reality. *Advances in Experimental Social Psychology,* **18,** 247–305.

Snyder, M. (1992). Motivational foundations of behavioral confirmation. *Advances in Experimental Social Psychology,* **25,** 67–114.

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.

Swann, W. B., & Ely, R. J. (1984). A battle of wills: Self-verification versus behavioral confirmation. *Journal of Personality and Social Psychology,* **46,** 1287–1302.

Tajfel, H., & Wilkes, A. L. (1963). Classification and quantitative judgement. *British Journal of Psychology,* **54,** 101–113.