

THE PENNSYLVANIA STATE UNIVERSITY  
SCHREYER HONORS COLLEGE

DEPARTMENT OF CURRICULUM AND INSTRUCTION

TEACHER JUDGMENTS OF STUDENT ACHIEVEMENT: INFLUENCE OF STUDENT  
RACE

KELLY N. NOWLIN  
Spring 2007

A thesis submitted in partial  
fulfillment of the requirements  
for a baccalaureate degree in  
Elementary Education with  
honors in Curriculum and Instruction

Reviewed and approved\* by the following

James C. DiPerna  
Assistant Professor of Education  
Thesis Supervisor

Debra Freedman  
Assistant Professor of Education  
Honors Adviser

\*Signatures are on file in the Schreyer Honors College.

Table of Contents

Abstract	2
Introduction	3
Method	11
Results	14
Discussion	17
References	23
Appendix	26

### Abstract

The purpose of this study was to explore the influence of student race on the formation of teachers' judgments of students' academic achievement. The study was conducted in a public school district in central Pennsylvania and included 5 teachers and 8 students. Each teacher was asked to complete a rating scale in both reading/language arts and mathematics (Academic Competence Evaluation Scales; DiPerna & Elliott, 2000) for two of their students' (one Caucasian, one minority). Teachers were also asked to predict students' proficiency scores on the Pennsylvania System of School Assessment (PSSA). These ratings and predictions were compared with the students' actual PSSA scores. The data revealed consistent judgments for both minority and Caucasian students.

### Teacher Judgments of Student Achievement: The Influence of Student Race

Teachers make judgments of students' academic achievement on a daily basis. Despite the fact that these judgments are used to guide instructional decisions and determine need for educational services, several researchers have hypothesized that judgments may be influenced by factors other than a students' academic skills (e.g., previous teachers' judgments, socioeconomic status, race, gender, and behavior; Alvidrez & Weinstein, 1999; Dougherty & Conolly, 1985). These judgments, accurate or not, lead to expectations for students' academic achievement. In their landmark study on teacher expectations, Rosenthal and Jacobson (1968) found that teachers' expectations varied across races, with the lowest expectations being for minority students. Furthermore, they found those students for whom lower expectations were held, did indeed show less academic gain during the school year. Other studies have confirmed the existence of expectancy effects as well (e.g., Alvidrez & Weinstein, 1999; Smith & Luginbuhl, 1976; St. George, 1983). While many studies have suggested evidence of expectancy effects in the classroom, the purpose of this study is to determine if the accuracy of teacher judgments, the building blocks for expectations, varies across race.

#### *Academic Achievement and Race*

Researchers have shown that there is a gap in academic achievement between African-American and Caucasian children, and, within the last decade, this research has been expanded to include other minorities as well (Currie & Thomas, 1996). Several studies have shown that minority students perform significantly lower than Caucasian students in all subject areas (e.g., Currie & Thomas, 1996; Phillips, Crouse, & Ralph 1998; Yando, Seitz, & Zigler, 1979). This achievement gap has been documented in standardized assessments of intelligence, assessments of specific academic content areas, and classroom performance (U.S. Department of Education,

2006). Although there is little agreement about the cause of the gap and how it progresses as children get older, there appears to be agreement that a gap, in some form, does exist.

Because teacher judgments have the potential to influence students' academic achievement through expectancy effects, they also have the potential, in the long run, to perpetuate (or help close) any achievement gap that already exists. Inaccurate teacher judgments that result in negative expectancy effects, however, can have adverse consequences for students, especially if those judgments tend to consistently underestimate the ability of minority students.

### *Teacher Judgments*

It is inevitable that teachers will make judgments about their students' academic ability. These judgments may be made consciously or unconsciously, but they are made as a way for teachers, like all other people in social situations, to organize their environment. Immediate perceptions and prior knowledge combine to form judgments about individuals we encounter (St. George, 1983). The same process occurs in classrooms across the country on the first day of school and throughout the school year. These judgments inform instructional practices and help determine eligibility for special services. These judgments continue to be made throughout the school year for a variety of reasons and purposes, and they are based on a variety of factors. Brophy (1983) conducted a review of the literature on teachers' judgments, expectations, and consequent expectancy effects. He concluded that a number of factors, including information about test performance, performance on assignments, group placement, classroom conduct, physical appearance, social class, ethnicity, sex, and diagnostic labels may affect the formation of teachers' judgments. These factors, however, had the greatest impact on teachers' judgments in studies where teachers were given carefully controlled information about fictitious students (e.g. Rosenthal & Jacobson, 1968). Studies of in-service teachers' expectations for their own

students, however, revealed that judgments of students' academic achievement are often accurate. Furthermore, any inaccurate judgments were corrected when new information became available (Borko, Cone, Russo, & Shavelson, 1979).

The differentiating factor in the outcomes between these two types of studies is whether or not the teacher had the opportunity to interact with her students. Hoge and Coladarci (1989) completed a review of 16 empirical studies that focused on the accuracy of teacher judgments of their own students. Hoge and Coladarci defined accuracy as "the correspondence between the relative standing of two sets of values: (a) the teachers' judgments of their students and (b) the students' actual performance on a relevant standardized test" (p. 302). As a whole, the 16 studies reported a wide range of accuracy; however, the median correlation coefficient (.66) suggests a moderate correlation between teachers' judgments and actual student performance.

#### *Factors that Affect Teacher Judgments*

Several researchers have conducted studies to identify factors that may influence teachers' judgments. Bennet, Gottesman, Rock, and Cerullo (1993), for example, concluded that teachers' perceptions of their students' behavior has a significant effect on their judgments of academic skills. In this study, kindergarten, first, and second grade teachers in two urban school districts were asked to rate their students' academic skills and classroom behaviors. Those students whom teachers perceived to exhibit bad behavior received lower academic ratings regardless of actual academic skill.

Student gender also has been identified as a potential student characteristic that influences teacher judgments; however, findings regarding the relationship between student gender and judgments of academic achievement are inconsistent. Bennett et al. (1993) concluded that gender influenced teachers' judgments of academic achievement through differential

perceptions of behavior between boys and girls. They observed that gender had a direct effect on teachers' perceptions of behavior, which, in turn, was directly related to academic judgments. Doherty and Conolly (1985) noted similar results in their study of teachers' ability to predict students' standardized test scores. The results of this study indicated that there was an interaction between the students' gender and the teachers' perceptions of the students' tidiness. Judgments of academic achievement were more dependent on tidiness for boys. However, the teachers' evaluations of girls' academic achievement reflected actual achievement regardless of perceptions of tidiness. These findings, that gender is an influencing factor in the formation of teachers' judgments, are consistent with the earlier findings of McCandless, Roberts, and Starnes (1972), who observed that gender consistently affected teachers' judgments, in favor of the girls. In contrast, Hoge and Butcher (1984) observed that gender did not function as an influencing factor in teachers' judgments of academic achievement. Similarly, Prawat, and Jarvis (1980) concluded that teachers' ratings of their students academic achievement was not influenced by gender as much as other variables such as intelligence or reading achievement. These findings support the assertion that teachers utilize available academic information when making judgments about their students (Brophy, 1983).

In addition to student factors that may play a role in influencing teachers' judgments, certain teacher factors and classroom or curriculum factors may also influence judgments. For example, Hopkins et al. (1985) found that teachers made more accurate judgments of their students' language arts, reading, and mathematics achievement than their social studies and science achievement. In addition, individual differences in teachers and teaching style have been identified as moderators of judgment/achievement correlations. Hoge and Coladarci (1989) identified four studies (Hopkins et al., 1985; Helmke & Schrader, 1987; Hoge & Butcher, 1984;

Coladarci, 1986) that demonstrate the importance of considering individual teacher differences. These researchers found a wide range in the accuracy of teacher judgments attributable to individual differences among teachers. Teacher variables such as instructional practices (e.g., frequency of structuring cues and individual support for students), teacher confidence in making such judgments, years of teaching experience, and educational beliefs and practices were identified as factors that affected the formation of teachers' judgments. As such, Hoge and Coladarci (1989) concluded, "[Not] all teachers are equally adept at making these judgments" (p. 307).

### *Teacher Judgments and Race*

The primary focus of this study was to explore the relationship between the accuracy of teachers' judgments and student race. In light of the achievement gap between Caucasian and minority students (Currie & Thomas, 1996; Phillips, Crouse, & Ralph 1998; Yando, Seitz, & Zigler, 1979) and the potential for inaccurate judgments to cause negative expectancy effects, it is important to determine if race systematically influences the accuracy of teachers' judgments of achievement.

The results of several studies have suggested that there may be differences in the accuracy of teachers' judgments for ethnic majority and minority students. St. George (1983) found that students in New Zealand belonging to the ethnic minority were judged less favorably by their teachers in terms of their academic achievement. Similarly, Rosenthal and Jacobson (1968) concluded that teachers' underestimated the achievement of minority students and consequently, developed low expectations. In a study of teachers' judgments of middle-school students, Hall, Howe, Merkel, and Lederman (1986) observed that teachers rated African

American students' achievement lower than their Caucasian counterparts with similar levels of academic achievement.

Findings that illustrate the influence of race on teachers' judgments have also been reported in a meta-analysis by Dusek and Joseph (1983). They conducted a meta-analysis of 24 studies that explored race as an influential factor for teachers' judgments. The studies included in the review were limited to those that sampled Caucasian and African American students. Dusek and Joseph found that race was a significant factor in the formation of teachers' judgments. Caucasian students were consistently predicted to perform better than African American students.

Potential consequences of inaccurate teacher judgments (i.e. the development of differential expectations) highlight the importance of determining the factors that affect such judgments. If race is one of them, negative judgments which may lead to lower expectations could prove harmful to the education of minority students as a whole, thereby perpetuating negative stereotypes and an endless cycle of low expectations and, consequently, low performance.

### *Expectancy Effects*

Once judgments are formed, there is the potential for expectancy effects to occur. Interest in the expectancy effects of teacher judgments has been generated by Rosenthal and Jacobson's (1968) *Pygmalion in the Classroom*. Although Rosenthal and Jacobson did not offer an explicit definition of teacher expectations, other authors since then have attempted to define and concretize the variable. Brophy and Good (1974), for example, defined teacher expectations as "inferences that teachers make about the present and future academic achievement and general classroom behavior of their students" (p. 32). Similarly, West and Anderson (1976) suggest the

following definition: “We have defined expectancy as a teacher’s attitude (behavioral predisposition) about a specific student (p. 616)”. This difficulty in conceptualizing and defining the key variable has led to questions about the reliability and validity of this line of research (Hoge, 1984). However, several researchers have asserted the existence of expectancy effects in the classroom (e.g., Alvidrez & Weinstein, 1999; Matthews, 1982; Rosenthal & Jacobson, 1968; Smith & Liginbuhl, 1976).

While there is agreement on the existence of expectancy effects, researchers have suggested many different factors can influence expectancy effects. Teacher-student interactions (quality and frequency), communication of differential judgments, and individual teacher and student characteristics are among the most commonly asserted moderators of teacher expectancy effects (Brophy, 1983; Brophy & Good, 1970). The confirmation of the existence of expectancy effects raises concerns about the judgments that precede them. The potential for negative expectations for minority students to adversely affect their education illustrates the importance of accurate judgments that may then lead to positive expectations.

### *Rationale and Purpose*

Classroom teachers are asked to make formal and informal judgments of their students’ academic achievement on a daily basis. These judgments are the basis for instructional decisions, such as creating instructional groups, communicating with parents, and identifying individuals who are eligible for special services. Therefore, it is extremely important for these judgments to be accurate. Inaccurate or inappropriately made teacher judgments, whether they are based on previous classroom performance, former teachers’ judgments, or some student characteristic, such as race, gender or behavior, may be equally detrimental to a student’s educational experience. In recognition of these potential negative implications of inaccurate teacher

judgments, the purpose of the current study is to explore the influence of student race on the accuracy of teachers' judgments of their students' academic achievement. Based on previous research (Dusek & Joseph, 1983; St. George, 1983), it was predicted that teachers will underestimate the achievement of minority students, while accurately predicting or overestimating the achievement of Caucasian students.

### *Research Questions*

What is the relationship between teachers' judgments of academic achievement and actual academic achievement as indicated by the PSSA?

Does this relationship vary by the race of the student? If so, how?

### *Inquiry v. Project*

Teacher inquiry is defined as the "systematic, intentional study of one's own professional practice" (Dana & Yendol-Silva, 2003). The current study reflects this philosophy in that it stemmed from my own observations of classrooms in which I have somehow been involved in the planning and/or implementing of instruction. As a pre-service teacher, it is important to learn how to turn observations and reflections from one's own teaching practice into meaningful research that can provide me with information and be useful in a specific context. In having had experiences in many types of educational settings over the past few years, and now spending a full year in a classroom, I have noticed the frequency with which and the necessity of teacher judgments of students' academic achievement, formal and informal. While I acknowledge the importance of such judgments in running day-to-day activities, I also wonder what factors mediate the judgments that teachers form. This is a question that led to the research questions of this study. After collecting and analyzing my own data, in a context specific to my student teaching experience, I expect that this inquiry will inspire a change in my own behaviors as a

teacher. Armed with knowledge of the research that already exists about teachers' perceptions of their students' academic achievement and the mediating factors, I intend to explore this relationship and its implications in my own teaching.

## Method

### *Participants*

Participants in this study included teachers ( $n = 5$ ) and their students ( $n = 8$ ) from one elementary school in a public school district in central Pennsylvania. The participants were selected from intermediate elementary classrooms, with two students being selected from each classroom. Of the five teacher participants, one was male and four were female, and all were Caucasian. These teachers had a range of teaching experience that ranges from 3 to 28 years. From each classroom, pairs of students were selected based on the similarity of their state achievement test scores, but each student in the pair was of a different race: one Caucasian and one minority.

*Triad A.* Triad A consisted of Teacher A and Students A1 and A2. Teacher A was a female, fourth-grade teacher who had been teaching for 6 years. Student A1 was a Caucasian female, and Student A2 was an Asian male. Student A2's native language was Chinese, and he received ESL instruction until a few months before data collection. Both students were 9-years-old at the time of data collection. Both students were in Teacher A's classroom for both language arts and mathematics instruction.

*Triad B.* Triad B consisted of Teacher B and Students B1 and B2. Teacher B was a female, fourth-grade teacher who had been teaching for 3 years. Both students from this classroom were female and were 10-years-old. Student B1 was Caucasian, and Student

B2 was African American. Both students are native English speakers and received language arts and mathematics instruction in Teacher B's classroom.

*Triad C.* Triad C consisted of Teacher C and Students C1 and C2. Teacher C was a female, fourth-grade teacher who had been teaching for 10 years. Both students from this classroom were female, and both students received language arts and mathematics instruction in Teacher C's classroom. Student C1 was 9-years-old, and Student C2 was 8-years old at the time of data collection. Student C1 was Caucasian, and Student C2 was African American. Both were native English speakers; however, Student C2 also spoke her family's native African language.

*Triads D and E.* Triad D consisted of Teacher D and Students D1 and D2. Triad E consisted of Teacher E and Students D1 and D2. These students received language arts instruction in Teacher D's classroom but received advanced mathematics instruction in Teacher E's classroom. Teacher D was a male, fifth-grade teacher who had been teaching for 28 years. Teacher E was a female, fifth-grade teacher who had been teaching for 21 years. Student D1 was a male, Caucasian student. Student D2 was a female, African American student. Both students are native English speakers. Student D2, however, also spoke her family's native African language. Both students were 10-years-old at the time of data collection.

### *Materials*

*Teacher Judgment Measures.* Teachers' indirect judgments of their students' academic achievement was measured using the academic skills portion of the Academic Competence Evaluation Scales (ACES; DiPerna & Elliott, 2000). Teachers were asked to rate their students' skill proficiency in reading and mathematics as well as the importance of these skills in their classroom. Ratings of the students' skills ranged from 1 (*far below grade level*) to 5 (*far above*

*grade level*). The rating of the importance of these skills ranged from 1 (*not important*) to 3 (*critical*). (See Appendix for examples of items.) Reliability coefficients for this measure are high, ranging from .88 to .97 in a test-retest reliability analysis. The academic skills portion of the ACES was compared with the Iowa Test of Basic Skills and the California Test of Basic Skills. Both comparisons yielded high correlations: .87 and .78 respectively (DiPerna & Elliott, 2000). In addition, teachers were asked to predict their students' proficiency level on the state standardized test from the previous year (i.e., *Below Basic*, *Basic*, *Proficient*, *Advanced*) as a direct measure of teacher judgments.

*Pennsylvania System of School Assessment (PSSA)*. PSSA scores from the Spring 2006 administration were used as an indicator of actual student achievement. The purpose of the PSSA is to assess students' attainment of Pennsylvania State Standards. Reliability of this assessment was calculated in two ways: internal consistency for all multiple choice items and inter-rater reliability for writing samples and open-ended questions. Internal consistency coefficients for reading range from .91 to .92. For mathematics, reliability ranges from .91 to .93. Inter-rater reliability was calculated as a kappa coefficient. Weighted kappa coefficients for reading ranged from .63 to .80, and for math, coefficients ranged from .81 to .96. (CTB McGraw Hill, 2006)

### *Procedure*

Consent was obtained from all of the teachers, and two students from each class were selected as potential participants. The students were selected, with the help of the principals of the schools, based on the similarity of their PSSA scores (to control for differences in actual achievement) and their races. When possible, the two students in each class were also the same sex to control for the possibility of gender differences influencing teacher judgments. After potential student participants were identified, consent forms were sent home to parents to get

permission for teachers to complete the ratings of achievement and to release their 2006 PSSA scores for the purposes of this project.

After obtaining consent from parents, the teachers were asked to complete the Reading/Language Arts and Mathematics subsections of the ACES for each of the student participants from their classroom. They also were asked to predict the students' 2006 PSSA score proficiency level (*Basic, Below Basic, Proficient, Advanced*). Students' 2006 PSSA scores for reading and math were collected from school district records.

### *Design and Analysis*

The design of this study was non-experimental and qualitative. Each teacher-student triad was considered an individual case, but data also were analyzed across all five triads in both reading and mathematics to identify the presence of patterns in the accuracy of teacher judgments. Due to the smaller than expected sample size, the results of this study must be viewed as exploratory and interpreted with caution.

### Results

Table 1 displays the students' actual PSSA proficiency ratings and the teachers' predicted proficiency ratings and the discrepancy between them. Four of the five teachers predicted both the minority students' and the Caucasian students' scores to be within a single proficiency level of their actual performance on the PSSA. Individually, teachers who either over or underestimated their students' scores, did so consistently across both students from their classroom. For example, Teacher D predicted that both students from his class achieved a *Proficient* score on the PSSA Reading test; however, both students actually achieved a score of *Advanced*. Although Teacher D predicted scores one proficiency level lower than the actual

score, these predictions were consistent across both students. For the most part, this is true for all teacher-student triads.

Table 1

*Actual vs. Predicted PSSA Achievement Ratings*

	Reading/Language Arts			Mathematics		
	Actual Proficiency	Predicted Proficiency	Accurate?	Actual Proficiency	Predicted Proficiency	Accurate?
Teacher A						
Student A1	Proficient	Proficient	Yes	Advanced	Proficient	Somewhat (-)
Student A2	Proficient	Proficient	Yes	Advanced	Advanced	Yes
Teacher B						
Student B1	Proficient	Proficient	Yes	Advanced	Proficient	Somewhat (-)
Student B2	Proficient	Below Basic	No (- -)	Basic	Below Basic	Somewhat (-)
Teacher C						
Student C1	Advanced	Advanced	Yes	Advanced	Proficient	Somewhat (-)
Student C2	Advanced	Advanced	Yes	Advanced	Proficient	Somewhat (-)
Teacher D/E						
Student D1	Advanced	Proficient	Somewhat (-)	Advanced	Advanced	Yes
Student D2	Advanced	Proficient	Somewhat (-)	Advanced	Advanced	Yes

*Note:* Yes = agreement between predicted and actual proficiency rating. Somewhat = prediction within one level of actual proficiency rating. No = 2 or more levels of discrepancy between predicted and actual proficiency rating. + and - denote the magnitude and direction of disagreement.

The only large and inconsistent discrepancy was that of Teacher B’s prediction for Student B2’s level of reading proficiency on the PSSA. When compared with this teachers’ prediction of Student B1’s level of reading proficiency, the former represents an underestimation of the minority students’ achievement. This discrepancy, however, is not present in the predictions for mathematics achievement.

Table 2 displays data comparing the students' actual PSSA proficiency level and the teachers' ratings of academic skills. ACES proficiency levels were computed by summing the teacher's ratings for each skill domain. Based on this number, each student was assigned a proficiency rating of *Developing*, *Competent*, or *Advanced*. These levels of proficiency were aligned with the PSSA levels of proficiency as follows: *Developing* corresponds with *Below Basic* and *Basic*; *Competent* corresponds with *Proficient*; and *Advanced* corresponds with *Advanced*. The results of this comparison show the same consistency as seen in Table 1. However, there was less agreement between the teachers' ACES ratings and actual PSSA proficiency levels than there was between predicted and actual PSSA proficiency levels. In this case, ACES ratings were consistently lower than PSSA proficiency levels. Again, no difference in the accuracy of the ratings exists between minority and Caucasian students.

Table 2

*Actual vs. Teacher Ratings (ACES)*

	Reading/Language Arts			Mathematics		
	Actual Proficiency	ACES Proficiency Rating	Accuracy?	Actual Proficiency	ACES Proficiency Rating	Accuracy?
Teacher A						
Student A1	Proficient	Developing	Somewhat (-)	Advanced	Developing	No (- -)
Student A2	Proficient	Developing	Somewhat (-)	Advanced	Competent	Somewhat (-)
Teacher B						
Student B1	Proficient	Developing	Somewhat (-)	Advanced	Developing	No (- -)
Student B2	Proficient	Developing	Somewhat (-)	Basic	Developing	Yes
Teacher C						
Student C1	Advanced	Competent	Somewhat (-)	Advanced	Developing	No (- -)
Student C2	Advanced	Competent	Somewhat (-)	Advanced	Competent	Somewhat (-)
Teacher D/E						
Student D1	Advanced	Developing	No (- -)	Advanced	Competent	Somewhat (-)
Student D2	Advanced	Developing	No (- -)	Advanced	Competent	Somewhat (-)

*Note:* Yes = agreement between predicted and actual proficiency rating. Somewhat = prediction within one level of actual proficiency rating. No = 2 or more levels of discrepancy between predicted and actual proficiency rating. + and – denote the magnitude and direction of disagreement.

### Discussion

The previous finding that teachers' judgments of students' academic achievement tend to be moderately accurate is supported by the data from the current study. Although the sample was small, all but one of the direct ratings fell within one proficiency rating of the students' actual level of achievement. This finding is consistent with the outcomes of Brophy's (1983) review. Brophy maintained that teachers tend to form accurate judgments of achievement when the students they are asked to judge are from their own classroom. The data presented here show that

teachers, when asked to directly predict their students score on the PSSA, were able to predict with a fair degree of accuracy. It is interesting to note, though, that although the discrepancies were relatively minor and consistent across both students from their classroom, all of the discrepant predictions erred on the side of an underestimation. That is, all discrepant predictions were underestimations of the students' achievement in comparison with the level of achievement indicated by PSSA scores. This finding was even more prevalent in the case of ACES ratings, which reflect the teachers' judgments of students' specific academic skills in the classroom. The ACES ratings were consistently one to two proficiency levels lower than the students' actual academic achievement. One potential explanation for this finding is that the PSSA assessment is not optimally aligned with the specific academic skills for which they were asked to make judgments using the ACES. Also, teachers may have been able to use previous standardized test data to inform their PSSA predictions; whereas the ACES ratings could only be informed by information obtained in the classroom. In addition, it is important to note that the sample was drawn from a high-achieving school district relative to the state population. As such, teachers' "grade-level expectations" (which are used to complete the ACES ratings) may be higher than the proficiency standards identified for the statewide student population.

In the one case where the actual versus predicted PSSA discrepancy is greater than one proficiency level, there may be a number of possible factors that influenced this judgment, the first being that Teacher B is the least experienced of the five teachers that participated in the study. Hoge and Butcher (1984) found that an experienced group of teachers' judgments were consistent with standardized test scores, and they suggested that this may be due, in part, to the teachers' level of experience. They postulated that the results would have been different if the teachers had been less experienced. Each teacher in that study had at least 6 years of experience.

Teacher B's relative lack of experience may be one explanation for the discrepancy between predicted and actual proficiency ratings. In the case of the ACES ratings, the mathematics predictions were inconsistent. Ratings for the minority student were accurate, while ratings for the Caucasian student were two proficiency levels below the students' actual score. Teacher B's overall inconsistency provides support for lack of experience as a possible explanation, because the ability to accurately, or at least consistently, rate students' level of proficiency may be a factor of years of teaching experience.

Another possible explanation for the discrepancy lies in the actual skills of Student B2. This particular student received widely discrepant scores on the PSSA. While the students' reading score fell into the proficient category, her mathematics score fell in the basic category. In addition, she was the only student included in this study who scored in the basic category on either test.

The initial hypothesis that the accuracy of teachers' judgments varies by race was not supported by this data. With the exception of Teacher B, all of the teachers gave consistent predictions and ratings for the Caucasian student and the minority student in his or her classroom in most cases. Although this finding is consistent with Brophy's conclusion that teachers who are well acquainted with their students tend to form accurate judgments, it is inconsistent with findings of other researchers who identify race as a factor that affects teachers' judgments (e.g. St. George, 1983).

#### *Limitations and Directions for Future Research*

There are several limitations to this study that preclude definitive conclusions based on the collected data. First, the sample size is too small from which to make generalizations. Time constraints and the demographics of the school district from which the participant sample was

drawn did not allow for collection of a larger sample size that could be used to quantitatively test the proposed hypotheses.

A second limitation of the study is that some of the pairs could not be matched in order to control for certain variables. Two of the four pairs of students were male-female pairs. Although the impact of this factor is not known for certain, it does not appear to have affected the teachers' judgments. Students A1 and A2 and D1 and D2 were the mixed gender pairs. The predictions and ratings made by Teachers A, D, and E, however, were consistent, for the most part, for both measures and both skill domains. The only inconsistency was Teacher A's predictions and ratings for mathematics. In each case, the discrepancy for the male, minority student was larger than for the female, Caucasian student.

A third limitation of the study is that, due to time constraints, PSSA scores from the 2005-2006 school year were used. Therefore these scores were available prior to the teachers' ratings and predictions. Teachers were not explicitly told which years' scores were being used; however, some teachers may have looked to the previous years' scores to make a prediction about this years' scores.

Due to the limitations and preliminary nature of this study, there are several directions for future research efforts. Future research should test these preliminary findings again with a larger sample size, ensuring that outside variables, such as gender, are controlled. The replication of this finding in a different setting would suggest that differential judgments, and consequently differential expectations, are not related to the achievement gap between minority and Caucasian students. Therefore, future research in this area should also address other possible factors related to this achievement gap.

In addition, the inconsistency of Teacher B's ratings and predictions raises the issue of the potential influence of years of experience on judgment accuracy. Future research in this domain should explore the relationship between teaching experience and the accuracy of teachers' judgments.

### *Conclusion*

The finding that teachers' judgments of students' academic achievement tend to agree with standardized test scores suggests that teachers are able to make accurate judgments to inform classroom instructional decisions. In addition, although the literature on teacher judgments suggests there are numerous factors that affect the accuracy of such judgments, the preliminary data reported in this study do not support the hypothesis that race is an influencing factor.

In light of the findings of the expectancy literature that indicate potentially negative outcomes resulting from low expectations, the data in this study provide some positive reassurance for teachers, parents, and administrators. While many researchers have confirmed the existence of expectancy effects and explored the potential negative outcomes of differential expectations (Alvidrez & Weinstein, 1999; Rosenthal & Jacobson, 1968; St. George, 1983), accurate judgments can help minimize such unintentional expectancy effects. Teachers who are able to make accurate judgments should be able to avoid these potential threats to their students' educational experience. Due to the limited sample size and the context in which this study was conducted, however, further research is necessary to draw firm conclusions and make generalizations that can reliably inform educational practices.

This research does, however, have implications for my own teaching practice. Brophy (1983) suggested that in-service teachers, when asked to make judgments about their own

students' academic achievement, used the best available information. This is in contrast to teachers forming judgments about students with whom they have not had the chance to interact. These judgments tend to be less accurate than judgments made for a teachers' own students. Because it is necessary for teachers to make judgments about their students in order to be effective teachers, it is essential that they use the best and most accurate information that is available. In addition, adjusting judgments and perceptions as new information becomes available is just as important. The data collected in this study supports the finding that teachers who have built relationships with their students make accurate judgments of their students' academic achievement, regardless of race. Therefore, in my own teaching, it will be essential for me to utilize the most appropriate and accurate information when making judgments. Using a variety of information (i.e. test performance, class participation, performance on assignments, formal and informal observations, teacher/student interactions, etc.) to inform my judgments will help me build the necessary relationships. From these personal relationships and the incorporation of a variety of pieces of information, accurate judgments are more likely to be formed. This, in turn, will help me plan and provide the best education for all of my students.

## References

- Alvidrez, J., & Weinstein, R. S. (1999). Teacher perceptions and later student academic achievement. *Journal of Educational Psychology, 91*, 731-746.
- Borko, H., Cone, R., Russo, N. A., & Shavelson, R. J. (1979). Teachers' Decision Making. In P. L. Peterson & H. J. Walberg (Eds.), *Research on Teaching: Concepts, Findings, and Implications* (pp. 136-160). Berkley, CA: McCutchen Publishing Corporation.
- Brophy, J. E. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology, 75*, 631-661.
- Brophy, J.E., & Good, T. L. (1970). Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. *Journal of Educational Psychology, 61*, 365-374.
- Brophy, J.E., & Good, T.L (1974). *Teacher – Student Relationships: Causes and Consequences*. New York: Holt, Rinehart, & Winston.
- Coladarci, T. (1986). Accuracy of teacher judgments of students responses to standardized test items. *Journal of Educational Psychology, 78*, 141-146.
- CTB McGraw-Hill. (2006, August). *Pennsylvania grade 3 assessment: Mathematics and reading technical report*. Monterey, CA.
- Currie, J., & Thomas, D. (1996, October). *Does Head Start help Hispanic Children?* Labor and Population Program, Working Paper Series 96-17 (Report: DRU-1528-RC).
- DiPerna, J. C., & Elliott, S. N. (2000). *Academic Competence Evaluations Scales*. San Antonio: The Psychological Corporation.

- Dougherty, J., & Conolly, M. (1985) How accurately can primary school teachers predict the scores of their pupils in standardised tests of attainment? A study of some non-cognitive factors that influence specific judgments. *Educational Studies, 11*, 41-60.
- Dusek, J. B., & Joseph, G. (1983). The bases of teacher expectancies: A meta-analysis. *Journal of Educational Psychology, 75*, 327-346.
- Hall, V. C., Howe, A., Merkel, S., & Lederman, N. (1986). Behavior, motivation, and achievement in desegregated junior high science classes. *Journal of Educational Psychology, 78*, 108-115.
- Helmke, A., & Schrader, F-W. (1987). Interactional effects of instructional quality and teacher judgment accuracy on achievement. *Teaching and Teacher Education, 3*, 91-98.
- Hoge, R. D., & Butcher, R. (1984). Analysis of teacher judgments of pupil achievement. *Journal of Educational Psychology, 76*, 771-781.
- Hoge, R. D. (1984). The definition and measurement of teacher expectations: problems and prospects. *Canadian Journal of Education, 9*, 213-228.
- Kuklinski, M. R., & Weinstein, R. S. Classroom and developmental differences in a path model of teacher expectancy effects. *Child Development, 72*, 1554-1578.
- Matthews, G. P. (1982). An example of the teacher expectation effect in mixed ability teaching. *Journal of Research in Science Teaching, 19*, 497-502.
- McCandless, B.R., Roberts, A., & Starnes, T. (1972). Teachers' marks, achievement test scores, and aptitude relations with respect to social class, race, and sex. *Journal of Educational Psychology, 63*, 153-159.

- Phillips, M., Crouse, J., & Ralph, J. (1998). Does the black-white test score gap widen after children enter school? In C. Jencks & M. Phillips (Eds.), *The black-white test score gap* (pp. 229-272). Washington, DC: Brookings.
- Prawat, R. S., & Jarvis, R. (1980). Gender difference as a factor in teachers' perceptions of students. *Journal of Educational Psychology, 72*, 743-749.
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the Classroom: Teacher Expectations and Pupils' Intellectual Development*. New York: Irvington.
- Smith, F. J., & Luginbuhl, J. E. R. (1976). Inspecting expectancy: Some laboratory results of relevance for teacher training. *Journal of Educational Psychology, 68*, 265-272.
- St. George, A. (1983). Teacher expectations and perceptions of Polynesian and Paheka pupils and the relationship to classroom behaviour and school achievement. *British Journal of Educational Psychology, 53*, 48-59.
- U.S. Department of Education. (2006). *Digest of Educational Statistics: 2005*. (NCES Publications No. 2006030). Washington, D.C.: Snyder, T. D., Tan, A. G., Hoffman, C. M.
- West, C. K., & Anderson, T.H. (1976). The question of preponderant causation in teacher expectancy research. *Review of Educational Research, 46*, 613-630.
- Yando, R., Seitz, V., & Zigler, E. (1979). *Intellectual and personality characteristics of children: Social-class and ethnic-group differences*. Hillsdale, NJ: Erlbaum.

Appendix

Academic Skills Rating Scale

Student Information

Name \_\_\_\_\_ Date \_\_\_\_\_  
 Birthdate \_\_\_\_\_ Grade \_\_\_\_\_ Sex  Male  Female

Teacher Information

Name \_\_\_\_\_ School \_\_\_\_\_

Directions

This portion of the Academic Competence Evaluation Scales (DiPerna & Elliott, 2000) is designed to assess a student's academic skills. For each item, **two ratings** are required. The first rating describes the proficiency of a skill. The second rating describes the **importance** of the skill in **your** classroom.

Please circle **two ratings** for each item. Please do not skip any items. If you have not had the opportunity to observe the skill, make a √ in the N/O column. Below is an example.

Reading/Language Arts	Far Below	Below	Grade Level	Above	Far Above	Not Important	Important	Critical	N/O
1 Reading comprehension	1	2	3	4	5	1	2	3	

For the proficiency rating scale, the teacher circled a 2, which indicates that the student's skill is judged to be below grade level expectations. For the importance rating, the teacher circled a 3, indicating this skill is critical to academic success in his or her classroom.

Following the rating scale, there is a place for you to predict the PSSA scores of this student for the reading and writing section of the PSSA test. Please fill in this section with a prediction of one of the following ratings: below basic, basic, proficient, advanced.

Academic Skills

Reading/Language Arts	Far Below	Below	Grade Level	Above	Far Above	Not Important	Important	Critical	N/O
1 Reading comprehension	1	2	3	4	5	1	2	3	
2 Word-math	1	2	3	4	5	1	2	3	
3 Vocabulary	1	2	3	4	5	1	2	3	
4 Identifying main idea	1	2	3	4	5	1	2	3	
5 Reading fluency	1	2	3	4	5	1	2	3	
6 Spelling	1	2	3	4	5	1	2	3	
7 Punctuation	1	2	3	4	5	1	2	3	
8 Grammar	1	2	3	4	5	1	2	3	
9 Writing conventions	1	2	3	4	5	1	2	3	
10 Oral conventions	1	2	3	4	5	1	2	3	
11 Drawing conclusions from written material	1	2	3	4	5	1	2	3	

Mathematics	Fw Below	Below	Grade Level	Above	Fw Above	Met Important	Met Important	Critical	NO
12 Computation	1	2	3	4	5	1	2	3	
13 Power analysis	1	2	3	4	5	1	2	3	
14 Measurement	1	2	3	4	5	1	2	3	
15 Understanding of spatial relationships	1	2	3	4	5	1	2	3	
16 Mental math	1	2	3	4	5	1	2	3	
17 Using numbers to solve daily problems	1	2	3	4	5	1	2	3	
18 Breaking down a complex problem	1	2	3	4	5	1	2	3	
19 Problem-solving	1	2	3	4	5	1	2	3	

**PSSA Predictions**

Please make a  $\checkmark$  in the appropriate column.

	Below Basic	Basic	Proficient	Advanced
Reading				
Mathematics				