Learning and Instruction 42 (2016) 123–140

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Teachers' explicit expectations and implicit prejudiced attitudes to educational achievement: Relations with student achievement and the ethnic achievement gap

E.R. Peterson [a, *], C. Rubie-Davies [b], D. Osborne [a], C. Sibley [a]

[a] School of Psychology, Faculty of Science, The University of Auckland, Private Bag 92019, Auckland, 1142, New Zealand
[b] School of Learning, Development and Professional Practice, Faculty of Education, The University of Auckland, Private Bag 92601, Auckland, 1150, New Zealand

ARTICLE INFO

Article history:
Received 6 November 2014
Received in revised form
12 December 2015
Accepted 18 January 2016
Available online 9 February 2016

Keywords:
Teacher expectations
Ethnic achievement gap
Implicit prejudiced attitudes
Explicit expectations
Achievement

ABSTRACT

In this paper, we consider whether teachers' explicit and implicit prejudiced attitudes underlie the ethnic achievement gap. To date, most research on teacher expectation effects has relied on explicit expectation measures that are prone to social desirability biases. In contrast, we examine the effects of teachers' (a) explicit ethnicity-based expectations for academic achievement and (b) implicit prejudiced attitudes about academic achievement on students' actual academic success over time. A total of 38 teachers completed both a traditional teacher expectation measure and a modified Implicit Association Task designed to assess ethnic stereotypes associated with academic achievement and failure. A multi-level analytic framework showed that students in classrooms of teachers with high expectations performed better in reading at the end of the year and that these effects were found across all ethnic groups. In contrast, whereas students' mathematics achievement scores were largely unrelated to teachers' explicit expectations, teachers' implicit prejudiced attitudes predicted student performance. Specifically, students benefited most academically when their teachers' implicit biases favored the ethnic group to which the student belonged. Findings are discussed in relation to differences in the salience of teachers' expectations and implicit prejudiced attitude in the classroom, and the ethnic achievement gap.

1. Introduction

1.1. Ethnic achievement gap

In many countries, academic achievement levels of ethnic minority groups are lower than those of the ethnic majority (Fryer & Levitt, 2004; Glock & Karbach, 2015; Harker, 2006; Jencks & Phillips, 2011; Sammons, 1995). Not surprisingly, this performance gap is of great concern to researchers, educators, and policy makers (e.g., Bishop, Berryman, Cavanagh, & Teddy, 2009; Delpit, 1995; Strand, 2014; Weinstein, Gregory, & Strambler, 2004), leading some to posit that socioeconomic status (SES) is (partly) responsible for this ethnic achievement gap (e.g., Strand, 2014). This is because SES affects the material resources available to support a child in their education, and is related to numerous health and developmental risks factors. SES also indirectly affects academic achievement through things like parental beliefs and expectations (Strand, 2014).

However, a recent longitudinal study ($N = 14{,}500$) in the United Kingdom found that, while SES differences (and a range of other contextual factors such as parental education) could partially or fully account for some achievement gaps between majority and minority groups, it could not account for all such differences (Strand, 2013). Strand suggested that the remaining unexplained ethnicity-based differences in academic achievement may be due to teachers' academic expectations which potentially bias their judgments of student achievement.

1.2. What are teachers' expectations and what difference do they make?

Teacher achievement expectations (commonly referred to as teachers' expectations) are defined as beliefs teachers hold about their students' academic capabilities and subsequent levels of

* Corresponding author.
E-mail address: e.peterson@auckland.ac.nz (E.R. Peterson).

http://dx.doi.org/10.1016/j.learninstruc.2016.01.010
0959-4752/

achievement. They are thought to be largely influenced by factors such as students' prior achievement, but also by ethnicity, SES, gender, and student diagnostic labeling. Other factors such as student behavior, the child's name, personality and friendliness, and the child's older siblings' success have been found to have lesser effects on teacher expectations (Rubie-Davies, 2015; Strand, 2013).

Differences in teacher expectations are important not only because they can influence teachers' subjective judgments of their students' academic abilities and grades, but also because when teachers hold different expectations for particular groups of students, they may engage, support, and teach their students differently. Indeed, Brophy and Good (1970) identified 17 ways that teachers responded differently to students for whom they had low (compared to high) expectations, including providing low expectation students with less time to respond to questions, less eye contact, and reduced warmth and friendliness during interpersonal interactions. Moreover, teachers' expectations can affect the types of learning tasks they set. Specifically, teachers set high-level learning tasks when they have high expectations, but low-level tasks when they have low expectations (Rubie-Davies, 2015). Because the amount that students learn depends on the learning opportunities provided by their teacher, differential expectations potentially exacerbate pre-existing achievement gaps (Rubie-Davies, 2015).

1.3. Student sensitivity to teacher expectations

Research shows that, from a young age, students are able to identify teachers who have high and low expectations from observing their teacher's behavior. Babad and Taylor (1992) conducted a study where students (Grades 4–10) were shown very brief samples (10-s audio and video clips in a language they did not understand) of teachers talking to a student. Despite the short length of the clip, students were able to detect whether the teacher was talking to a high or a low expectation student. These differences were argued to be picked up from the teachers' nonverbal behaviors because they were detected in the absence of language and verbal cues.

1.4. Teacher expectations and academic achievement

Although few question the existence of differential teacher expectations, there is debate about the size of teacher expectation effects on students' achievement. A recent meta-analysis of 674 experimental and naturalistic studies found that the average expectation effect size across all students was a modest $d = .43$ (Hattie, 2009). However, based on 11 naturalistic studies, Jussim, Robustelli, and Cain (2009) argue that teacher expectations predicted student achievement *because* their expectations were accurate. Nevertheless, even critics of teacher expectation research concede that, although the effects of teacher expectations on student achievement may be small overall, they can still be powerful in certain circumstances (Jussim & Harber, 2005; Van den Bergh, Denessen, Hornstra, Voeten, & Holland, 2010). In particular, critics note that students from low socioeconomic groups and students from ethnic minorities may be particularly susceptible to teacher expectation effects through processes like stereotype threat (Jussim, Eccles, & Madon, 1996).

1.5. Source of teachers' expectations: stereotypes and prejudiced attitudes

One reason why some teachers form different expectations for different ethnic groups may be because of stereotypes and implicit prejudiced attitudes they hold about students from particular ethnic groups (Jussim et al., 1996). A *stereotype* is defined as a belief that members of a particular group (e.g., men, women, minorities, the poor, etc.) have certain attributes or traits (Greenwald & Babaji, 1995; Wilson, Lindsey, & Schooler, 2000). Stereotypes, which are cognitive in nature, are usually associated with an affective component: prejudice. *Prejudice* is defined as the (often negative) feelings and attitudes one holds towards a particular group (Singh, 2015; Stangor & O'Brien, 2010). In the current study, we refer to these as prejudiced attitudes. A third related term is *discrimination*. This is the behavioral component of an intergroup attitude and refers to the differential treatment of people based on the group to which they belong (Vescio & Bloodhart, 2010).

Stereotypes and prejudiced attitudes can be positive or negative and can exist on an explicit and implicit level. Theoretically, explicit and implicit stereotypes and prejudiced attitudes are distinct constructs (Gawronski, Strack, & Bodenhausen, 2009). Whereas, a person is consciously aware of and has control over explicit stereotypes and prejudiced attitudes (Gawronski & Bodenausen, 2006), implicit stereotypes and prejudiced attitudes emerge via automatic processing and are typically unconscious.

1.5.1. Development and activation of stereotypes and prejudiced attitudes

Both explicit and implicit stereotypes and prejudiced attitudes develop from repeated exposure to pairings of a social group or object with a particular characteristic. One such pairing found in industrial countries is the ethnic achievement gap between majority and minority students. Most teachers are aware of this association, as attempts to address the achievement gap lie at the heart of many educational policies such as the 'No Child Left Behind' (2002) policy in the United States which targets poor and minority students or the Ka Hikitia: Accelerating Success 2013–2017 policy in New Zealand which targets Indigenous Māori under-achievement (Ministry of Education, 2012). Teachers may also be aware of the ethnic achievement gap from their own teaching experience by either directly observing it among their own students, or indirectly by talking to others.

Importantly, knowledge of a particular stereotype does not necessarily mean endorsement of it. Specifically, Devine (1989) showed that people who were high and low on an explicit measure of racial bias were equally knowledgeable about the *content* of racial stereotypes. What differentiated these two groups, however, was whether or not they suppressed the automatic activation of these beliefs in a subsequent task. Whereas those low on bias controlled the use of previously activated stereotypes, those high on bias responded to the ambiguous behavior of a target in a stereotypical manner. Critically, this means that factors that interrupt teachers' ability to suppress automatically activated stereotypes (e.g., a busy classroom) could unconsciously allow these biases to leak out affecting a teachers' behavior.

This finding highlights a key difference between explicit and implicit stereotypes and prejudiced attitudes in that explicit stereotypes and attitudes (compared to implicit ones) are easier to control. According to Chaiken and Trope's (1999) dual process MODE model, the execution of control over a prejudiced attitude depends on the *m*otivation and *o*pportunity to control or *de*termine the attitude-to-behavior process. For example, when a teacher fills in a teacher expectation questionnaire stating his or her explicit expectations for each student's achievement, there is time to reflect and provide a socially desirable response. Although the teacher expectation task does not make direct reference to any particular stereotypes, it is possible that some teachers will be explicitly aware of stereotypes surrounding the achievement of some groups in their class. Given that the teacher expectation task is not timed, teachers also have plenty of opportunity to reflect and respond in a

socially desirable way. As such, the teacher expectation task can be viewed not only as a judgment of students' future achievement, but also as a way of tapping into whether teachers hold explicit prejudiced attitudes about the educational achievement of some student groups over others.

Conversely, when a stereotype is activated and time is pressured (and/or the motivation to utilize the cognitive effort needed to provide a socially desirable response is low such as when doing an implicit attitude test or teaching in a busy classroom), a less socially desirable and more automatic response may emerge. Hence, when presented with a timed implicit association task that measures teachers' implicit prejudiced attitudes about the academic achievement of different ethnic groups, they are likely to respond automatically (i.e., without the ability to provide a socially desirable response). A description of the implicit association test used in the current study is described in more detail below.

Chaiken and Trope (1999) MODE model also suggested that implicit and uncontrollable responses could influence a person's behavior even when an explicit and more socially desirable attitude was retrieved from memory. This is particularly true of nonverbal behavior, which is more likely to operate outside of a person's conscious awareness (Asendenorf, Banse, Mücke, 2002). Indeed, implicit attitudes—which can be activated simply by being in the mere presence of a person of a particular ethnicity (Fazio, 2001) can dominate a situation (Gloch & Karbach, 2015). Therefore, although a teacher might *explicitly* reject the stereotype that an ethnic achievement gap exists in their class (as indicated on a teacher expectancy rating scale), they may still *implicitly* show discriminatory behaviors by encouraging some children to take home school work and library books, but not others (Huss-Keeler, 1997). Likewise, teachers might unconsciously refer more ethnic majority kids to gifted programs and more ethnic minority students to special education programs (Tenenbaum & Ruck, 2007).

Consistent with this thesis, Tenenbaum and Ruck's (2007) meta-analysis found that American teachers tended to have higher expectations for European American students than they did for Latino/a and African American children ($d = .46$ and $d = .25$, respectively). Moreover, teachers tended to show more positive interactions (e.g., asking questions and providing encouragement) towards European American students than they did towards Latino/a and African American students ($d = .21$). In other words, implicit stereotypes may affect teachers' behaviors without conscious awareness, particularly when the ability or motivation to control their behavior is compromised by things like competing cognitive demands or other factors that reduce working memory (De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009; Wilson et al., 2000).

1.6. Measurement of explicit and implicit stereotypes and prejudiced attitudes

As noted above, explicit measures of stereotypes and prejudiced attitudes involve deliberate conscious processing and are typically measured using self-report procedures. Dovidio and Fazio (1992) argued that these methods were generally seen as satisfactory for assessing people's explicit attitudes, with the caveat that responses given on explicit measures may reflect a person's perception of a socially desirable response (as opposed to true opinion). In contrast to explicit measures, implicit assessment procedures assess spontaneous processes that are better at capturing attitudes that may be prone to social desirable responding (i.e., attitudes towards different ethnic groups; Glock & Karbach, 2015).

Many different implicit measures have been proposed that use different stimuli and categorization tasks (see Glock & Kovacs, 2013, for a review). These tasks typically involve the recording of response latencies or physiological responses, or participating in memory tasks (Dovidio, Kawakami, & Beach, 2008; Gawronski & Bodenausen, 2006). Critically, implicit measures of prejudice have been found to predict behavior, such as how friendly a person is towards an African American experimenter (Fazio, Jackson, Dunton, & Williams, 1995) and the size of an ethnic achievement gap in a classroom (Van den Bergh et a1., 2010). A recent meta-analysis of the Implicit Association Task found that it predicted behavior, judgments, and physiological outcomes at an average of $r = .27$ (Greenwald, Poehlman, Uhlmann, & Banaji, 2009). As such, studies now tend to focus on *when* and *how* (rather than *if*) implicit measures predict behavior (see Friese, Hofmann, & Schmitt, 2009).

The variety of implicit association tasks available is argued to be one of the reasons why the results regarding implicit attitudes towards ethnic minorities have been inconsistent (Glock & Karbach, 2015). This concern led Glock and Karbach to use the same set of stimuli in three different implicit association tasks assessing teachers' attitudes towards ethnic minorities (the Implicit Association Task, the Affect Misattribution Procedure and the Affective Priming Task). Although teachers' scores on the three tests were only weakly correlated, the authors found a consistent pattern of results indicating that teachers implicitly held more positive attitudes towards the ethnic majority (vs. minority) group members. The low correlations between measures were argued to be due to a combination of measurement error, and the structure of the underlying tests which tapped into different aspects of people's implicit attitudes. Nevertheless, Glock and Karbach's findings suggested that teachers held implicit prejudiced attitudes towards ethnic minorities, and that different implicit measures had different mechanisms for tapping into these beliefs. Findings such as these contribute to calls for the use of more implicit measures within educational research.

1.7. Explicit expectations and student achievement

As noted above, numerous studies exploring the effects of teachers' explicit expectations on actual student achievement have been conducted (e.g., Hattie, 2009; Jussim et al., 2009; Van den Bergh et al., 2010). These studies have, however, often failed to control for students' *prior* achievement. As such, it is impossible to rule out the possibility that teachers' explicit expectations for different ethnic groups are merely accurate predictions of widely observed ethnic achievement gaps. Moreover, the few studies that have controlled for prior achievement when examining ethnic differences in performance yield mixed results. For example, McKown and Weinstein (2008) showed that when teachers expected European American and Asian American students to perform better than their similarly-achieving African American and Latino/a counterparts, their teachers' expectancies accounted for between .29 and .38 standard deviations of the ethnic achievement gap at the end-of-the-year. In contrast, de Boer, Bosker, and Van der Werf (2010) found that teachers' expectations were not moderated by students' ethnicity (nor by gender, nor achievement motivation), but rather, were related to the students' socioeconomic status, intelligence, and parents' aspirations. These findings suggest the need for more research in this area.

1.8. Implicit ethnic prejudiced attitudes and student achievement

To date, only one study has examined the impact of teachers' implicit ethnic prejudiced attitudes and their effect on student achievement outcomes. Van den Bergh et al. (2010) used a traditional Implicit Association Task (IAT) to assess teachers' general implicit prejudiced attitudes to different ethnic groups, as well as the Modern Racism Scale (McConahay, 1986), to measure teachers' explicit prejudiced attitudes. In this study, Van den Bergh et al.'s IAT

required students to make judgments about students' names (which reflected three different ethnic groups) which were paired with words that had generally positive or negative connotations, but were unrelated to achievement. The task was based on the assumption that paired objects would be easier to categorize if there was an implicit association between the names and evaluative words. If people responded faster to one type pairing over another (e.g., European names with positive words), they were argued to show bias. Results indicated that teachers' *explicit* ethnic attitudes were unassociated with student achievement. In contrast, teachers' general *implicit* ethnic attitudes correlated with student achievement such that classrooms with teachers who had high levels of implicit prejudice had larger ethnic-based achievement gaps than classrooms with teachers who had low levels of implicit prejudice.

Van den Bergh et al.'s (2010) study provided the first demonstration that teachers' general ethnic-based implicit prejudiced attitudes related to student achievement. Despite its merits and novel approach of examining the ethnic achievement gap via an IAT, Van den Bergh et al.'s (2010) study had two key limitations. First, their general implicit attitude measure paired students' names with general 'good' and 'bad' words (e.g., peace, fear); it did *not* pair the names with symbols or words that related to academic achievement and failure. As such, the pairings enabled them to measure generalized ethnic prejudice, but it did *not* assess the specific association between prejudice and academic achievement. Similarly, the explicit prejudiced attitude task used was the Modern Racism Scale, which is also a *general* measure of a person's attitudes towards ethnic minorities and not specifically related to academic achievement.

Second, Van den Bergh et al.'s (2010) study was cross-sectional and did not control for students' prior achievement. As such, the teachers' expectations may be (at least partially) rooted in reality. Thus, no research to date has examined the effects of teachers' implicit prejudiced attitudes in relation to the *academic achievement of different ethnic groups* and combined that with an explicit measure of teacher expectations to assess the impact both these measures have on student achievement over time.

1.9. The current study

The current study sought to address this oversight by a) examining the effects of teachers' explicit expectation on the academic achievement of different ethnic groups and b) examining teachers' implicit prejudiced attitudes and their relationship to academic achievement and c) the combined influence of explicit expectations and implicit prejudiced attitudes on students' end-of-year achievement (controlling for prior achievement). In doing so, we examined these relationships within a sample of New Zealand teachers and their classrooms. As such, the following section provides an overview of ethnic-based differences in achievement within New Zealand schools.

1.9.1. The study context: ethnic differences in achievement in New Zealand

In New Zealand, Māori (the indigenous population) and Pasifika (those from the Pacific Islands) students tend to underachieve compared to European and East Asian[1] students, an achievement gap that emerges in elementary school and continues throughout their educational career (Hattie, 2008). This pattern of underachievement and economic disadvantage has historical roots and reflects the fact that Māori and Pasifika have repeatedly been marginalized by successive New Zealand governments. Within the educational sector, this can be most obviously traced back to 1867 when the government established Native schools. In Native schools, Māori were only taught a restrictive curriculum of basic reading, writing, agricultural skills and personal hygiene, with the implication that broader skills would not be needed for this group. Māori under achievement was first statistically identified about 100 years later in a 1960s government report. This report led to a series of educational reforms to address educational disparities with policies supportive of biculturalism and multiculturalism. Nevertheless, until very recently, there has been little real change in these inequalities over the last 50 years (Bishop et al., 2009).

1.9.2. Influences on New Zealand teachers' expectations

As noted above, underachievement and poverty typically go hand-in-hand and both factors can influence teachers' expectations. Although SES is argued to be the strongest predictor of the ethnic achievement gap internationally, it is not always easy for a teacher to know the socioeconomic status of a child's family in New Zealand. Research has shown that the way a person is categorized depends on the salience and importance of the information available (Rakić, Steefens, Mummendey, 2011). In New Zealand, it is illegal to record parental occupation as a socioeconomic indicator. And although teachers will be aware of the average SES of the community in which they work, it is not the case that every child in that school would be of similar SES. In contrast, a students' ethnicity is likely to be salient, as it is required demographic on school enrollment forms. Moreover, ethnicity is often physically salient in students' appearance. Hence, even though ethnicity and SES are frequently confounded, in New Zealand, ethnicity is arguably a more salient trigger than SES for potential prejudice among New Zealand teachers.

In summary, we argue that a major gap in the teacher expectation literature is that it over-relies on teachers' explicit expectations when examining ethnic achievement differences. Indeed, it is possible that teachers have implicit stereotypes and prejudiced attitudes about the educational achievement of particular ethnic groups in the absence of explicit biases regarding ethnic-based differences in achievement expectations. In a classroom environment where cognitive resources may be limited (i.e., teachers juggle the needs of multiple children), teachers' implicit attitudes may exhibit a greater influence than their explicit attitudes over their behaviors which, as Van den Bergh et al. (2010) suggested, may (partly) explain the ethnic achievement gap.

2. Method

2.1. Participants

Participants were 38 teachers (27 females, 11 males) from 11 schools. The majority of teachers identified as European (60%), followed by Pasifika (17%), Māori (6%), Asian1 (3%), and Other (14%). The teachers had an average of 16 years teaching experience (range 5–19) and the socioeconomic status of the schools in which they taught varied. New Zealand schools are rated nationally on socioeconomic levels from 1 to 10 with 1–3 being regarded as low socioeconomic, 4–6 considered mid and 7–10 high. Eight teachers taught in three low socioeconomic schools, 22 teachers taught in six mid-socioeconomic schools and 8 teachers taught in two high socioeconomic schools. Teachers were offered a $40 shopping voucher in return for their participation.

The teachers in the study were involved in the first year of a

[1] Note: In New Zealand the term 'Asian' is typically used to refer to people from East Asia or India, as most Asians in NZ are from East Asia (e.g., Chinese, Korean, or Taiwanese) or the Indian subcontinent. People who identify themselves as coming from other parts of Asia (e.g., Iran, Pakistan) typically select their ethnicity as 'Other' and state what this is. In this study, we use the term 'Asian' with respect to its typical interpretation within the New Zealand context.

teacher expectation project which used a randomized control trial design to teach teachers about the instructional strategies and practices of high expectation teachers (those who have high expectations for *all* their students). Teachers randomly assigned to the intervention condition participated in a series of four workshops. In the first workshop, teachers were introduced to the literature on teacher expectations and informed about the effectiveness of high expectation teachers (see Rubie-Davies, 2006, 2007, 2008; Rubie-Davies & Peterson, 2011). The next three workshops focused on three major ways in which high expectation teachers differed from low expectation teachers. Specifically, participants were told that high expectation teachers (a) use flexible grouping (rather than within-class ability grouping) to teach reading and mathematics; (b) create a warm socioemotional climate; and (c) use goal-setting with students to foster motivation, engagement, and autonomy. The workshops on these three areas involved participants learning about the practices and beliefs of high expectation teachers in the morning, whereas the afternoon was devoted to planning collaboratively how they would implement the focal practice into their classrooms.

The researchers visited schools three times throughout the year to support participants in the implementation of the new practices. Because it is compulsory in New Zealand for teachers to engage in professional development each year, participants in the control group participated in their school's regular professional development training which was often as intense as the intervention described. For more information about this intervention, see Rubie-Davies, Peterson, Sibley, and Rosenthal (2014).

All teachers who were approached volunteered to take part in the current study. Of these 38 teachers, 18 were in the intervention group and 20 were in the control group. Importantly, the explicit expectation data and beginning-year achievement data used in the current study were collected as baseline data before the intervention was introduced. In our statistical models, we also controlled for whether or not the teacher was in the intervention group.

Student-level data were collected from 1060 students (53% boys, 47% girls) who were in the previously-mentioned teachers' classes. The students' ages ranged from 6 to 13 years ($M = 9.3$; $SD = 1.5$), or Grades 3 to 7. The students circled their ethnicity from a list of the major ethnicities (European, Māori, Pasifika, Asian). There was also "Other" and a request to specify the ethnicity. In keeping with the New Zealand standard classification of ethnicity (Department of Statistics, 1993), a level 1 priority recording system was used such that when students reported a dual ethnic background, priority was assigned in the following order Māori > Pacific > Other (excluding European) > European (Cormack & Robson, 2010). In the current study, the majority of students identified themselves as European (49%), followed by Māori (19%), Pasifika (19%), Asian (11%) and Other (2%).

In the city where this study was undertaken, the population is diverse. The majority is NZ European (59%), 11% are Māori, 15% are Pasifika, and 23% are Asian (Statistics New Zealand, 2013). Nevertheless, among school-age children, the proportions of the smaller groups are larger because of a younger demographic profile within these groups (Statistics New Zealand, 2010). Most students in New Zealand attend state schools (86%), with a small proportion attending either state-integrated religious schools or private schools. All students in the current study were attending state schools.

2.2. Instruments

2.2.1. Teachers' explicit expectations for academic achievement

Teachers' explicit expectations of each of their student's predicted achievement levels in both reading and mathematics were measured on a 1 (very much below average) to 7 (very much above average) Likert scale. Teachers were asked to list the students in their class and, without referring to school records, indicate the level they expected each student to reach by the end of the academic year (relative to national curriculum levels). Teachers were provided with information about the average national achievement of students at various year levels, but not individual student achievement data.

2.2.2. Teachers' implicit prejudiced attitudes for academic achievement

Teachers' implicit prejudiced attitudes for academic achievement were assessed using a modified version of the self-esteem Implicit Association Task (IAT) developed by Greenwald and Farnham (2000). We called this task the Teacher Implicit Academic Achievement Association Task (TIAAAT), which measured the relative strength of the association between (a) European, Māori, and Asian surnames and (b) images/words associated with academic success and failure (see Appendix Tables 1 and 2). Specifically, images depicting success and failure consisted of symbols teachers in New Zealand commonly use to grade assignments in local schools. For example, images showing assignments that received grades of either "10/10", "A+" and "Excellent" were used to depict success, whereas images showing assignments that received grades of either "1/10", "F-" and "Poor" were used to depict failure.

Two experts were used to identify 6 images depicting success and 6 images depicting failure. Only images that received unanimous agreement between coders were used in this study. A similar process was used to identify surnames that clearly depicted the intended ethnicity (Note, surnames rather than first names were used to avoid adding a confounding variable of gender). For example, Wong, Wairau, and Clarke were used to respectively depict Asian, Māori and European surnames common in New Zealand. All trials and blocks were counterbalanced so that each set of surnames was paired with signs of achievement and signs of failure and the ethnicity pairings (e.g., European vs. Māori, European vs. Asian, Māori vs. Asian) were also counterbalanced across participants.

The general structure of the TIAAAT was the same as other IAT computer tasks. Four blocks of attribute discrimination practice trials were given before each test block. The first required participants to sort stimuli (e.g., surnames) which were presented in the center of the screen into two categories that were presented either on the left or the right of the screen (e.g., EUROPEAN– left and ASIAN– right). Participants were told that the surnames would only be from one of the two groups, and they needed to press the 'E' on the keyboard if the surname matched the category on the left or the 'I' on the keyboard if the surname matched the category on the right. In the second attribute discrimination practice session, the left and right categories were switched. The third and fourth attribute discrimination practice sessions asked participants to sort symbols of achievement and failure into two groups (SUCCESS and FAILURE) in a similar way to how they grouped the surnames; these were also subsequently counterbalanced in the next practice trial.

In the test trials, participants were told to " react as quickly as you can and make as few mistakes as possible", and that this time the four types of stimuli the participants had practiced categorizing in the practice sessions (e.g., EUROPEAN, ASIAN and SUCCESS, FAILURE) would now be paired together. For example, in the first trial block 'SUCCESS' and 'ASIAN' would appear on the top left of the screen and 'FAILURE' and 'EUROPEAN' would appear on the top right of the screen. The participants were then presented with symbols or surnames in the center of the screen that related to one of the four categories and were asked to categorize the symbols or surnames into the appropriate category. For example, if a symbol relating to success was presented (e.g., A+), it would go in the SUCCESS/ASIAN category, not the FAILURE/EUROPEAN category.

Likewise, if a European surname was presented (e.g., Clarke), it would also go into the FAILURE/EUROPEAN category. The theory behind the task is that easier pairings of categories (measured by faster response times) are more strongly associated in memory than more difficult parings of categories (which attract a slower response). Hence, the blocks of test trials are set up so that stereotype-compatible categories are paired together in one block and stereotype-incompatible categories are compared in another block. For example, in the current task, a stereotypical response would be to link European surnames with success (compatible categories), but not Māori surnames with success (incompatible categories). Two compatible blocks of trials and two incompatible blocks of trials were given, with options to have a break (with no time limit) between each block. Each block consisted of 30 trials, with an inter-trial interval of 250 ms, and the blocks were counterbalanced across participants. Participants were informed when their response was incorrect, but no error penalties were given.

The resulting reaction-time data were analyzed following the recommendations outlined by Greenwald, Nosek, and Banaji (2003). In keeping with previous research, we calculated an IAT-D score for each participant. The IAT-D score measures the difference between the average response latencies between contrasted conditions (European vs. Māori, European vs. Asian, and Māori vs. Asian) divided by the standard deviation of response latencies across the conditions. As such, larger (positive) values represented a stronger prejudiced attitude towards the first group in each pairing, whereas negative values reflected a stronger prejudiced attitude towards the second group in each pairing (see Nosek, Bar-Anan, Sriram, & Greenwald, 2013, for technical details). Three IAT-D scores were calculated in the current study: European vs. Māori; European vs. Asian; and Māori vs. Asian.

To avoid participant fatigue on the TIAAAT, we were only able to include three ethnic groups. Adding an additional ethnic group to the three chosen would have meant we needed to include an additional three contrasting groups, which would have effectively doubled the length of the test. Therefore, in line with previous IAT studies, we chose the ethnic majority group (Europeans) in New Zealand for inclusion in our stimuli. We then selected two minority groups with contrasting levels of achievement (i.e., a typically high performing group (namely, Asian) and a typically low performing group (namely, Māori)) that were relevant to New Zealand. The Ministry of Education (2012) has actively discouraged teachers' low expectations of Māori, so the measurement of implicit bias was one way to assess the success of the Ministry initiatives.

All participants responded with more than 92% accuracy and no response times were greater than 10,000 ms. This suggested that the teachers remained focused on the task. As such, the implicit data from all participants were included in the analysis.

2.2.3. Student achievement

Student achievement in reading and mathematics was measured using the Assessment for Teaching and Learning (e-asTTle; the 'e' indicates that it is electronically created and marked), a commonly employed standardized assessment used in Grades 2–11 in New Zealand. The e-asTTle system enables teachers to construct standardized tests online by selecting the appropriate levels for the test, the length of the tests, and the specific learning areas to be tested. The computerized system then generates tests according to the guidelines requested. In this instance, all tests were generated by the second author who consulted with deputy principals at the schools involved in the project to ensure that the tests were appropriate in terms of the content to be measured.

All tests were 40 min in length. In reading, the areas tested were processes and strategies, ideas, and language features; and in mathematics, number knowledge, number sense, and algebra. All schools completed the same tests at the appropriate levels for their students. That is, tests were generated at Levels 2 to 5 in order to cater for the different grade levels involved in the project. All items in the e-asTTle system were calibrated using Item Response Theory (IRT) scoring procedures and the standard error (SE) is estimated to be 22 points. Thus, students' total scores can be compared across classes, years, and schools *irrespective* of which test they take. Scores on e-asTTle in both reading and mathematics can range from 1100 to 1900. Thus, it would be expected that students in Grade 2 would score closer to 1100, whereas those in Grade 12 would be closer to 1900. Hence, in most instances, student scores increase across time and grade, enabling student progress to be tracked. The mean learning gain across New Zealand in e-asTTle scores for Grades 2 to grade 7 students is 30 points each year. For the tests used in the current study, the scores ranged from 1171 to 1672 in reading ($M = 1395.80$, $SD = 107.57$) and from 1194 to 1754 in mathematics ($M = 1423.88$, $SD = 97.95$).

2.3. Procedure

Student achievement data using e-asTTle were collected at the beginning, middle, and end of the academic year, but only data from the beginning and end-of-year were included in this study. The tests were administered by the teachers and returned to the researchers for marking. As such, the teachers were unaware of their students' actual achievement on these measures.

Teachers' explicit academic expectations were collected at the beginning and middle-of −year, but only beginning-of-year expectations were used in the current study. In contrast, implicit prejudiced attitudes for academic achievement were collated towards the end of the academic year, when the measure became available.

Researchers administered the first teacher expectation questionnaires approximately four weeks into the academic year. This provided an ideal baseline measurement, as schools had not completed standardized testing in mathematics and reading at that time of the year and teachers were asked not to refer to school records of prior achievement. As such, teachers were likely to use subjective information to inform their judgments of their students' capabilities. The intervention study was begun after these data were collected. To assess their implicit prejudiced attitudes for academic achievement, teachers were asked to complete the online TIAAAT on their personal laptops or computers in their own time.

2.4. Data analysis

First, a series of *t*-tests were used to examine possible differences between students with missing end-of-year achievement data and students with no missing end-of-year achievement data. Although 22% of students had no achievement data at end-of-year, no differences were found in the beginning-of-year mathematics or reading achievement scores for these students ($p > .05$). Likewise, no ethnic group differences were found ($p > .05$). Teachers did however have slightly higher explicit expectations for students with both beginning and end-of-year mathematics ($t[1008] = -.216$, $p = .03$) and reading scores ($t[1008] = -2.95$, $p = .003$).

Multivariate analysis of covariance (MANCOVA) with listwise deletion for missing data was then conducted in IBM SPSS version 21. This analysis was carried out in order to examine whether ethnic and gender differences would be found in students' mathematics and reading achievement, and in teachers' mathematics and reading expectations (controlling for age). Post hoc *t*-tests with Bonferroni corrections were conducted to further explore any significant group differences.

Next, a series of multilevel models were created with *Mplus version* 7.1 to investigate the extent to which students' end-of-year

mathematics and reading achievement was explained by student level variables (i.e., Level 1) or school/teacher level variables (i.e., Level 2). All Level 1 variables were centered at the group mean and all Level 2 variables were centered at the grand mean. Missing values were estimated using Full Information Maximum Likelihood.

At Level 1, ethnicity was assessed given our focal interest in the ethnic–based achievement gap described above and possible ethnic differences in teachers' expectations. The remaining Level 1 variables were assessed as plausible alternative explanations for teachers having higher expectations of one ethnic group and lower expectations of another. For example, current achievement was controlled and we controlled for age because (as noted above) scores on the New Zealand standardized achievement test increase with age. Finally, student gender (female 1, male 0) was assessed in order to take into account any differential achievement that could be attributed to gender, such as boys being perceived as better at mathematics than girls (see Good & Findley, 1985).

With respect to the Level 2 analysis, a sample size of 30 is argued to be the smallest acceptable number for conducting multilevel modeling (Kreft et al., 1998). Mass and Hox's (2005) multilevel modeling simulation study found that, with 30 groups, the estimates of regression coefficients were unbiased. However, the non-coverage rate for the Level 2 intercept and slope variance was nearly 9% (which is greater than the nominally acceptable 5%), implying that the standard errors of the Level 2 variance components were about 15% too small (Mass & Hox, 2005). Together, these findings suggested that, although estimating fixed effects with small samples is acceptable, there can still be substantial bias in the estimation of random effects with small sample sizes.

We also needed to control for three Level 2 variables, thereby placing further demands on Level 2 of our multilevel model. The first level 2 variable we controlled for was whether the teachers in the study were in the intervention (1) or control group (0). This was done in order to control for potential differences in the teaching practices of control and intervention teachers over the course of the year. We could also examine whether being in either group moderated teacher implicit achievement biases collected towards the end-of-year. Next, we controlled for the socioeconomic status of the school which was treated as an ordinal variable in this analysis. This variable was included as previous research has suggested that socioeconomic status may be confounded with ethnicity (Magnuson & Duncan, 2006). Finally, we modeled teachers' implicit prejudiced attitudes for academic achievement at Level 2.

In order to reduce the number of Level 2 factors being modeled simultaneously and thereby lessen the modeling demands, only one TIAAAT score was modeled at a time (European vs. Māori $N = 718$, European vs. Asian $N = 635$, Māori vs. Asian $N = 309$). This resulted in three multi-level models being generated predicting end-of-year reading achievement and three models predicting end-of-year mathematics achievement. Further, in each model, the Level 1 student ethnicity variable was coded to correspond with the level two TIAAAT outcome variable. For example, for the European vs. Māori TIAAAT model, the Level 1 student ethnicity variables were coded European 1, Māori 0. Similarly, for the European vs. Asian TIAAAT analysis, the Level 1 ethnicity variable was coded European 1, Asian 0. Finally, for the Māori vs. Asian TIAAAT analysis, the Level 1 ethnicity variable was coded Māori 1, Asian 0.

Although limitations to our Level 2 sample size prevented us from directly testing cross-level effects, we recoded the TIAAAT variable so that it could be modeled as a Level 1 fixed effect. This allowed us to examine whether the degree of alignment between teachers' implicit prejudiced attitudes and the students' ethnicity resulted in additional achievement gains. That is, if a teacher had an implicit bias of .5 towards Europeans vs. Māori names, then the TIAAAT score was recoded such that all European students in that teachers' class were given an TIAAAT score of +.5 and all Māori students in the class were give a score of −.5. This enabled us to examine the possibility that, for example, a European student would benefit if his or her teacher's implicit prejudiced attitude to academic achievement was pro-European.

3. Results

3.1. Ethnic achievement gap

Mean levels of student achievement at the beginning and end-of-year for each ethnic group are shown in Table 1. Analysis of Variance confirmed a main effect for ethnicity (Mathematics: $F[1532] = 14.18$, $p < .001$; Reading: $F[1527] = 26.53$, $p < .001$). Post hoc-t-tests with Bonferroni Corrections revealed the expected ethnic achievement trend such that European and Asian students typically achieved at higher levels than Māori and Pasifika students (p's $< .01$). However, when controlling for prior achievement, Analysis of Covariance (ANCOVA) found no main effect for ethnicity on students' end-of-year achievement scores. Instead, two main effects on students' end-of-year achievement were found; one for prior achievement and one for age. That is, higher performing students on the beginning-of-year assessment performed better on the end-of-year assessment (mathematics: $F[1458] = 25.438$, $p < .001$; reading: $F[1454] = 510.09$, $p < .001$). Likewise, older students achieved higher mathematics ($F[6, 458] = 2.291$, $p = .034$) and reading ($F[6454] = 1.205$, $p < .01$) scores at the end-of-year. Together, these findings provisionally suggested that any ethnic achievement gap found in student achievement was more likely to be due to prior achievement than student ethnicity.

3.2. Descriptive statistics for teachers' explicit expectations and implicit prejudiced attitudes for academic achievement

The descriptive statistics for teachers' explicit academic achievement expectations are shown in Table 1. High scores reflected high explicit teacher expectations for their students' end-of-year achievement (range 1–7). The mean score suggested that teachers' expectations for student achievement mirrored the students' performance such that higher expectations were observed for European and Asian students than for Māori or Pasifika students.

An ANCOVA (controlling for beginning-of-year achievement scores) found no main effect for student ethnicity on teachers' explicit academic expectations. However, main effects were found for prior mathematics and reading achievement such that teachers with high expectations of students had higher beginning-of-year achievement scores in mathematics ($F[1538] = 123.7$, $p < .001$) and reading ($F[1529] = 150.5$, $p < .001$). Furthermore, teachers had higher explicit reading ($F[1529] = 3.95$, $p < .001$) and mathematics ($F[1538] = 3.213$, $p = .002$) expectations for older students than for younger students.

In order to examine whether teachers in higher ability classes had higher mean level expectations for their students than teachers in lower ability classes, we calculated teachers' mean level of expectation for their class and conducted an Analysis of Variance on the range of mathematics and reading scores within the class. No main effects were found.

With respect to the implicit prejudiced attitudes measure of academic achievement, a standardized score of .00 indicated no tendency to associate educational achievement with one ethnic group over another. In contrast, a positive score indicated a prejudice in favor of the larger or ethnic majority group, whereas a negative score indicated a prejudice in favor of the minority group. The teachers' TIAAAT scores all showed a bias (European vs. Māori

Table 1
Ethnic differences in teachers' explicit expectations at beginning-of-year, and student achievement at beginning (T1) and end-of-year (T2).

	European			East Asian			Māori			Pasifika			Other			Total		
	Mean	*SD*	*n*	Mean	*SD*	*n*	Mean	*SD*	*n*	Mean	*SD*	*n*	Mean	*SD*	*n*	Mean	*SD*	*n*
Maths expectation	4.64	1.41	499	5.45	1.13	110	4.48	1.34	184	4.54	1.47	190	4.70	1.46	27	4.68	1.410	1010
Maths achievement T1	1414	87.31	427	1458	93.26	95	1370	82.52	158	1370	75.35	163	1363	96.67	19	1401	89.810	862
Maths achievement T2	1458	82	389	1461	94.49	78	1429	71	144	1428	76.58	150	1412	89.87	24	1446	82.310	785
Reading expectation	4.81	1.48	499	4.87	1.33	110	4.64	1.36	184	4.62	1.48	190	4.48	1.42	27	4.74	1.450	1010
Reading achievement T1	1385	101	419	1404	97.73	89	1329	91.4	154	1333	83.31	149	1353	95.97	21	1367	99.780	832
Reading achievement T2	1422	99.14	386	1434	92.39	73	1373	90.43	143	1374	85.39	145	1353	95.97	21	1404	97.800	770

$M = .46$, $SD = .49$ range −.58 to 1.19; European vs. Asian $M = .50$, $SD = .32$, range −.37 to 1.05, and Māori vs. Asian $M = .31$, $SD = .46$, range −1.01 to 1.02). Three follow-up one-sample *t*-tests indicated that all three were significantly different from zero ($p < .001$). In each of these cases, the prejudiced attitudes were in favor of the first group in each pairing. In other words, teachers generally made quicker categorizations when European names were paired with symbols of achievement compared to when Māori or Asian names were paired with symbols of achievement, and vice versa for symbols of failure. Somewhat surprisingly, teachers also tended to make faster categorizations when Māori names were paired with achievement compared to Asian names.

A series of partial correlations were also run to examine the relationship between teachers' implicit prejudiced attitudes for academic achievement bias and teachers' explicit expectations, controlling for students' prior reading and mathematics achievement. The correlations were run separately for each of the three combinations of ethnic groups used in the TIAAAT (i.e., European vs. Māori, European vs. Asian, and Māori vs. Asian). In all cases, no statistically significant correlations were found between teachers' implicit prejudiced attitudes for academic achievement and teachers' explicit academic achievement expectations.

We also found no differences in European versus non-European teachers' explicit expectations for mathematics or reading achievement. We did, however, identify one statistically significant difference between European and non-European teachers' implicit prejudiced attitudes for academic achievement for Europeans vs. Māori. Specifically, European teachers had a significantly stronger bias in favor of European students than did non-European teachers ($t(1050) = 5.24$, $p < .001$).

3.3. Multilevel model results: predictors of students' end-of-year achievement

To examine whether ethnic-based differences in teachers' implicit prejudiced attitudes for academic achievement and teachers' explicit expectations were related to differences in students' actual achievement levels in mathematics and reading, multilevel models were constructed (see Tables 2a–2c and 3a-3c). The correlations among the Level 1 and Level 2 variables are given in Table 4. As expected, these results demonstrated that the highest correlations were between students' mathematics and reading achievement scores and between teachers' explicit expectations for mathematics and reading.

Note we are unable to report the fit of the models as, our multilevel regression model is just-identified (i.e., our model has zero degrees of freedom). As such, fit statistics are inapplicable in this context (see Byrne, 2006; Kline, 2005). In fact, because our model is fully saturated (i.e., there are no degrees of freedom), standard indices of model fit indicate that our model provides a perfect fit to these data (i.e., CFI = 1.00). Indeed, as Byrne (2006) stated, "despite the capability of the model to yield a unique solution for all parameters, the just-identified model … has no degrees of freedom and therefore can never be rejected" (p. 31). Thus, the analyses presented in our manuscript are comparable to a set of sequential regressions with the exception that our analyses are conducted simultaneously and, thus, utilize all of the available data.

3.3.1. Model 1 results: empty model

Two separate empty unconditional multilevel models were constructed first to establish what proportion of variance in student end-of-year (a) mathematics and (b) reading achievement could be explained by teacher/school (level 2) factors (see model 1 in Tables 2a–c and 3a-c). The ICCs of these initial empty models indicated that approximately 20% of the variance in mathematics achievement and 30% of the variance in reading achievement at end-of-year was attributable to factors residing at the teacher/school level. Comparing the empty model to subsequent models allowed estimation of the value of additional predictors in the model. Models 2 and 3 of Tables 2a–c and 3a-c display the unstandardized regression coefficients, their standard errors and statistical significance, as well as the variance estimates and R^2 statistic at the student level and the teacher level.

3.3.2. Model 2: control model

Model 2 in each analysis was a control model that contained mostly the demographic factors that might influence student academic achievement outcomes. Specifically, model 2 contained the Level 1 variables of student gender, age, ethnicity, and beginning-of-year mathematics or reading achievement and the Level 2 variables of SES of the school, and whether the teachers were in the intervention or the control group.

As expected, in all these models (and subsequent models), the students' beginning-of-year mathematics or reading achievement scores were significant predictors of their end-of-year mathematics or reading scores. Not surprisingly, academically stronger students at beginning-of-year performed better at the end-of-year than their academically weaker counterparts. In addition, student gender was a statistically significant predictor of academic performance such that boys outperformed girls in end-of-year mathematics in one of the models (Māori and Asian students, Table 2c).

With respect to the Level 2 predictors, school SES was a statistically significant predictor in all but one of the models. Specifically, students in higher SES schools tended to perform better on end-of-year mathematics and reading achievement tests than students in lower SES schools. The exception to this trend was in the European vs. Asian mathematics model (Table 2c). SES also remained a significant predictor of student achievement in all but one of the subsequent models. No differences were found based on whether the teacher was in the intervention or the control group.

Overall, with respect to end-of-year reading achievement, the addition of the model 2 control predictors reduced the Level 1 prediction error by between 56% and 61% in reading and by between 6% and 9% in mathematics. At Level 2, the control variables reduced the prediction error from 13 to 23% in reading and 13–26% in mathematics.

Table 2a
Teachers' European vs Māori explicit expectations and implicit prejudiced attitudes, and associations with end-of-year mathematics achievement.

		Model 1 ($n = 718$)	Model 2 ($n = 471$)	Model 3 ($n = 471$)	Model 4 ($n = 471$)
		Intercept only	Control	Explicit expectation	Explicit expectations and implicit bias
		Coefficient (SE)	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)
Level 1	Student Gender (female 1, male 0)		−.096 (.06)	−.097 (.06)	−.096 (07)*
	Student Age		.015 (.06)	−.017(.06	−.019 (.06)
	Student Ethnicity (European 1, Māori 0)		.108 (.07)	.107 (.07)	.002 (.05)
	Mathematics Achievement Beginning of the Year		.215 (.09)*	.199 (09)*	.201 (.09)*
	Explicit Exptn in Mathematics			.020 (.04)	.014 (.04)
Interactions	Explicit Exptn in Mathematics × Student Ethnicity			.053 (.06)	−.038 (.051)
	TIAAAT European vs Māori and Student Ethnicity Match				.133 (.05)**
	Explicit Mathematics Exptn × TIAAAT European vs Māori and Student Ethnicity Match				.012 (.05)
Level 2	School Socioeconomic Status		.107 (.05)*	.107 (.05)*	.114 (.04)**
	Intervention (0) or Control (1)		.131 (.13)	.131 (.13)	.067 (.13)
	TIAAAT European vs Māori				.265 (.13)*
Interaction	Intervention or Control × Implicit Achievement Bias: European vs Māori				−.158 (.23)
	Intercept	14.46 (.07)	14.51 (.06)***	14.51 (.06)	14.39 (.07)***
	Residual variance level 1	.498 (.03)***	.425 (.04)***	.434 (.04)	.421 (.04)***
	Residual variance level 2	.158 (.05)***	.113 (.04)***	.113 (.04)	.096 (.03)*
	ICC	.243	.257	.257	.258
	R^2 at level 1		.060	.062	.069
	R^2 at level 2		.227	.277	.398
	R^2 difference level 1			.002	.007
	R^2 difference level 2			.050	.121

Note 1. Expt = expectation; TIAAAT = Teacher Implicit Attitude to Academic Achievement Task; ICC = intraclass correlation; E = European; M = Māori; *$p < .05$; **$p < .01$ ***$p < .001$.
Note 2. For modeling purposes the achievement scores were divided by 100 and unstandardized regression coefficients are reported.

Table 2b
Teachers' European vs Asian explicit expectations and implicit prejudiced attitudes, and associations with end-of-year mathematics achievement.

		Model 1 ($n = 635$)	Model 2 ($n = 422$)	Model 3 ($n = 413$)	Model 4 ($n = 413$)
		Intercept only	Control	Explicit expectation	Explicit expectations and implicit bias
		Coefficient (SE)	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)
Level 1	Student Gender (female 1, male 0)		−.11 (.08)	−.114 (.08)	−.124 (.08)
	Student Age		−.020 (.05)	−.022 (.05)	−.018 (.05)
	Student Ethnicity (European 1, Asian 0)		.056 (.12)	.108 (.14)	.210 (.17)
	Mathematics Achievement Beginning of the Year		.201 (.09)#	.163 (.09)#	.166 (.09)#
	Explicit Exptn in Mathematics			.033 (.05)	.037 (.05)
Interactions	Explicit Exptn in Mathematics × Student Ethnicity			−.074 (.09)	−.234 (.13)#
	TIAAAT European vs Asian and Student Ethnicity Match				−.109 (.13)
	Explicit Mathematics Exptn × TIAAAT European vs Asian and Student Ethnicity Match				.149 (.08)#
Level 2	School Socioeconomic Status		.089 (.05)*	.081 (.05)	.073 (.05)*
	Intervention (0) or Control (1)		.210 (.14)	.183 (.15)	.178 (.13)
	TIAAAT European vs Asian				.533 (.21)*
Interaction	Intervention or Control × Implicit Achievement Bias: European vs Asian				−.074 (.40)
	Intercept	14.53 (.08)***	14.56 (.07)***	14.57 (.07)	14.31 (10)***
	Residual variance level 1	.554***	.484 (.05)***	.473 (.05)	.473 (.05)***
	Residual variance level 2	.160**	.110 (.07)**	.128 (.04)	.091 (.03)**
	ICC	.214	.227	.243	.232
	R^2 at level 1		.043	.042	.047
	R^2 at level 2		.261	.194	.393
	R^2 difference level 1			−.001	.005
	R^2 difference level 2			−.067	.199

Note 1. Expt = expectation; TIAAAT = Teacher Implicit Attitude to Academic Achievement Task; ICC = intraclass correlation; E = European; A = Asian; *$p < .05$; **$p < .01$; ***$p < .001$. #$p = .07$.
Note 2. For modeling purposes the achievement scores were divided by 100 and unstandardized regression coefficients are reported.

Table 2c
Teachers' Māori vs Asian explicit expectations and implicit prejudiced attitudes, and associations with end-of-year mathematics achievement.

		Model 1 ($n = 309$)	Model 2 ($n = 204$)
		Intercept only	Control
		Coefficient (SE)	Coefficient (SE)
Level 1	Student Gender (female 1, male 0)		−.294 (.11)**
	Student Age		−.065 (.12)
	Student Ethnicity (Māori 1, Asian 0)		−.202 (.134)
	Mathematics Achievement Beginning of the Year		.221 (.09)*
	Explicit Exptn in Mathematics		
Interactions	Explicit Exptn in Mathematics × Student Ethnicity		
	TIAAAT Māori vs Asian and Student Ethnicity Match		
	Explicit Mathematics Exptn × TIAAAT Māori vs Asian and Student Ethnicity Match		
Level 2	School Socioeconomic Status		.058 (.05)
	Intervention (0) or Control (1)		.139 (.15)
	TIAAAT Māori vs Asian		
Interaction	Intervention or Control × Implicit Achievement Bias: Māori vs Asian		
	Intercept	14.41 (.08)***	14.40 (.06)***
	Residual variance level 1	.552 (.06)***	.432 (.049)***
	Residual variance level 2	.112 (.07)***	.109 (.06)#
	ICC	.180	.208
	R^2 at level 1		.094
	R^2 at level 2		.131

Note 1. Expt = expectation; TIAAAT = Teacher Implicit Attitude to Academic Achievement Task; ICC = intraclass correlation; M = Māori; A = Asian; *$p < .05$; **$p < .01$; ***$p < .001$. #$p = .09$ ##$p = .07$.
Note 2. For modeling purposes the achievement scores were divided by 100 and unstandardized regression coefficients are reported.

Table 3a
Teachers' European vs Māori explicit expectations and implicit prejudiced attitudes, and association with end-of-year reading achievement.

		Model 1 ($n = 718$)	Model 2 ($n = 481$)	Model 3 ($n = 470$)	Model 4 ($n = 470$)
		Intercept only	Control Model	Explicit expectations	Explicit expectations and implicit bias
		Coefficient (SE)	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)
Level 1	Student Gender (female 1, male 0)		.044 (.05)	.015 (.04)	.012 (.04)
	Student Age		−.040 (.05)	−.093 (.04)*	−.092 (.04)*
	Student Ethnicity (European 1, Māori 0)		.076 (.05)	.140 (.05)**	.188 (.04)***
	Reading Achievement Beginning of the Year		.750 (.04)***	.630 (.04)***	.631 (.04)***
	Explicit Exptn in Reading			.165 (.03)***	.166 (.03)***
Interactions	Explicit Exptn in Reading × Student Ethnicity			−.053 (.03)	−.063 (.04)
	TIAAAT European vs Māori and Student Ethnicity Match				.044 (.04)
	Explicit Reading Exptn × TIAAAT European vs Māori and Student Ethnicity Match				−.002 (.03)
Level 2	School Socioeconomic Status		.169 (.06)**	.164 (.06)**	.164 (.06)**
	Intervention (0) or Control (1)		.068 (.17)	.044 (.17)	.040 (.18)
	TIAAAT European vs Māori				.012 (.18)
Interaction	Intervention or Control × Implicit Achievement Bias European vs Māori				.033 (.34)
	Intercept of M2	14.04 (.10)	14.12 (09)***	14.119 (.09)***	14.11 (.09)***
	Residual variance level 1	.664***	.265 (.02)***	.216 (.02)***	.215 (.02)***
	Residual variance level 2	.320***	.245 (.05)***	.251 (.06)***	.251 (.06)***
	ICC	.332	.360	.364	.364
	R^2 at level 1		.561	.637	.639
	R^2 at level 2		.278	.261	.261
	R^2 difference level 1			.076	.002
	R^2 difference level 2			−.017	.000

Note 1. Expt = expectation; TIAAAT = Teacher Implicit Attitude to Academic Achievement Task; ICC = intraclass correlation; E = European; M = Māori; *$p < .05$; **$p < .01$ ***$p < .001$.
Note 2. For modeling purposes the achievement scores were divided by 100 and unstandardized regression coefficients are reported.

3.3.3. Model 3: addition of teachers' explicit expectations for academic achievement

Model 3 in each analysis examined the impact of adding in the Level 1 factor of teachers' explicit expectations for each student on students' end-of-year achievement. Also included in model 3 was the interaction between student ethnicity and teachers' explicit expectations. Results from these analyses indicate that teachers' explicit expectations of students had a strong relationship with students' end-of-year reading achievement ($\beta = .48–.67$), but not with their mathematics achievement. Importantly, these explicit

Table 3b
Teachers' European vs Asian explicit expectations and implicit prejudiced attitudes, and association with end-of-year reading achievement.

		Model 1 ($n = 1061$)	Model 2 ($n = 413$)	Model 3 ($n = 406$)	Model 3 ($n = 406$)
		Intercept only	Control Model	Explicit expectations	Explicit expectations and implicit bias
		Coefficient (SE)	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)
Level 1	Student Gender (female 1, male 0)		−.005 (.05)	−.011 (.04)	−.009 (.04)
	Student Age		−.016 (.05)	−.06 (.05)	−.059 (.05)
	Student Ethnicity (European 1, Asian 0)		−.096 (.10)	−.008 (.10).	−.104 (.16)
	Reading Achievement Beginning of the Year		.791 (.04)***	.661 (.05)***	.665 (.05)***
	Explicit Exptn in Reading			.174 (.03)***	.172 (.03)***
Interactions	Explicit Exptn in Reading × Student Ethnicity			−.083 (.06)	−.112 (.13)
	TIAAAT European vs Asian and Student Ethnicity Match				−.053 (.06)
	Explicit Reading Exptn × TIAAAT European vs Asian and Student Ethnicity Match				−.024 (.09)
Level 2	School Socioeconomic Status		.161 (.04)**	.143 (.05)**	.145 (.05)**
	Intervention (0) or Control (1)		.066 (.15)	.027 (.15)	.026 (.14)
	TIAAAT European vs Asian				.269 (.20)
Interaction	Intervention or Control × Implicit Achievement Bias: European vs Asian				−.887 (.39)*
	Intercept of M2	14.15 (.09)***	14.24 (.08)***	14.25 (.08)U***	14.12 (.11)***
	Residual variance level 1	.699 (.04)***	.273 (.02)***	.223 (.02)***	.222 (.02)***
	Residual variance level 2	.236 (.07)***	.176((.03)***	.191 (.04)***	.159 (.04)***
	ICC	.295	.294	.295	.293
	R^2 at level 1		.569	.643	.644
	R^2 at level 2		.327	.269	.385
	R^2 difference level 1			.074	.001
	R^2 difference level 2			−.058	.116

Note 1. Expt = expectation; TIAAAT = Teacher Implicit Attitude to Academic Achievement Task; ICC = intraclass correlation; E = European; A = Asian; *$p < .05$; **$p < .01$ ***$p < .001$.
Note 2. For modeling purposes the achievement scores were divided by 100 and unstandardized regression coefficients are reported.

Table 3c
Teachers' Māori vs. Asian explicit expectations and implicit prejudiced attitudes, and association with end-of-year reading achievement.

		Model 1 ($n = 309$)	Model 2 ($n = 182$)	Model 3 ($n = 182$)	Model 3 ($n = 182$)
		Intercept only	Control Model	Explicit expectations	Explicit expectations and implicit bias
		Coefficient (SE)	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)
Level 1	Student Gender (female 1, male 0)		.044 (.08)	.052 (.08)	.057 (.07)
	Student Age		−.077 (.07)	−.070 (.07)	−.078 (.07)
	Student Ethnicity (Māori 1, Asian 0)		−.121 (.08)	−.126 (.08)	−.126 (.10)
	Reading Achievement Beginning of the Year		.487 (.07)***	.488 (.07)***	.485 (.07)***
	Explicit Exptn in Reading			.246 (.05)***	.257 (.06)***
Interactions	Explicit Exptn in Reading × Student Ethnicity			−.054 (.07)	−.082 (.07)
	Implicit Exptn Māori vs Asian × Student Ethnicity				.003 (.08)
	Explicit Reading Exptn × Implicit Achievement Bias for MA students				.052 (.06)
Level 2	School Socioeconomic Status		.101 (.06)**	.101 (.06)	.096 (.06)
	Intervention (0) or Control (1)		.204 (.198)	.204 (.19)	.306 (.20)
	TIAAAT Māori vs Asian and Student Ethnicity Match				.025 (.20)
Interaction	Intervention or Control × TIAAAT Māori vs Asian and Student Ethnicity Match				−.246 (.40)
	Intercept	13.93***	13.97***	13.96 (.09)***	13.96***
	Residual variance level 1	.623 (.06)***	.205***	.024***	.203***
	Residual variance level 2	.274 (.08)***	.291***	.291***	.288***
	ICC	.378	.388	.389	.388
	R^2 at level 1		.611	.613	.615
	R^2 at level 2		.132	.132	.140
	R^2 difference level 1			.002	.002
	R^2 difference level 2			.000	.008

Note 1. Expt = expectation; TIAAAT = Teacher Implicit Attitude to Academic Achievement; ICC = intraclass correlation; M = Māori; A = Asian; *$p < .05$; **$p < .01$; ***p 001.
Note 2. For modeling purposes the achievement scores were divided by 100 and unstandardized regression coefficients are reported.

Table 4
Correlations among the level 1 and level 2 factors in the multilevel models.

	Level 1 predictors	1	2	3	4	5	6	7	8	9	10
1	Student Gender (Female 1, Male 0)										
2	Student Age	−.001									
3	Mathematics Expectation	−.007	.137**								
4	Reading Expectation	−.007	.134**	.792**							
5	Mathematics Score Beginning of the Year	.011	.537**	.027	.415**						
6	Reading Score Beginning of the Year	−.002	.501**	.000	.382**	.812**					
7	Ethnicity European (1) Māori (0)	−.021	−.022	−.005	−.086**	.013	.050				
8	Ethnicity European (1) Asian (0)	−.025	.038	.041	.092**	.286**	.273**	.379**			
9	Ethnicity Māori (1) Asian (0)	.020	−.048	−.001	.074*	.012	−.064	−.058	−.304**		
Outcome Variables											
10	Mathematics Score End of the Year	−.018	.052	.068*	.071*	.670**	.012	.009	−.010	−.032	
11	Reading Score End of the Year	−.019	.050	.099**	.083**	.617**	−.052	.037	−.004	−.039	.787**

	Level 2 Predictors	1	2	3	4	5	6	Mean (*SD*)
1	Implicit Prejudice Attitudes European vs Māori							.381 (.556)
2	Implicit Prejudice Attitudes European vs Asian	.379**						.497 (.324)
3	Implicit Prejudice Attitudes Māori vs Asian	−.058	−.304**					.305 (.457)
4	School Socioeconomic Status	.155**	.387**	−.002				
5	Intervention (1) or Control Group (0)	.014	.082**	−.009	.011			
Outcome Variables								
6	Mathematics Score End of the Year	.009	−.01	−.032	.051	.041		
7	Reading Score End of the Year	.037	−.004	−.039	.06	.028	.787**	

*$p < .05$; **$p < .01$ ***$p < .001$.

expectations in reading did not interact with student ethnicity. This suggests that students of different ethnic groups were similarly affected by their teachers' high explicit reading expectations. That is, on average, the reading achievement of all students improved when teachers had high explicit expectations for their class. This model reduced the prediction error by between 2% and 7%.

Note that model 3 for Māori vs. Asian in mathematics (Table 2c) was unable to be computed due to a non-positive definite matrix, and so the results are not reported. This may reflect the smaller total number of students in the model, as these analyses compare the achievement of Māori and Asian students (i.e., two ethnic minority groups within New Zealand).

3.3.4. *Model 4: final model with explicit expectations and implicit prejudiced attitudes for academic achievement*

In the final model, we added the teachers' implicit prejudiced attitudes score and the interaction between the implicit prejudiced attitude scores and whether the teacher was in the intervention or the control group to the Level 2 predictors. At Level 1, we examined whether an alignment between teachers' implicit prejudiced attitudes and the students' own ethnicity was associated with greater academic achievement,[2] and whether teachers' explicit expectations interacted with their implicit prejudiced attitudes (matched against the student) to affect end-of-year achievement. Consistent with our findings for model 3, the Māori vs. Asian mathematics model (Table 2c) was unable to be computed due to a non-positive definite matrix, and so the results are not reported.

For simplicity, we describe below the Level 2 findings (teacher and school level factors) first. The Level 2 findings showed that the addition of teachers' implicit prejudiced attitudes had no direct effect on students' reading achievement. Implicit prejudiced attitudes were, however, associated with students' mathematics achievement (see Tables 2a and 2b). On first inspection, Tables 2a and 2b seem to indicate that teachers with an implicit prejudiced attitude in favor of Europeans (vs. Māori and Asians) had higher performing students in mathematics at end-of-year ($\beta = .26$ and $\beta = .53$, respectively) (Tables 2a and 2b). However, examination of the interactions modeled at Level 1 suggested that teachers' implicit prejudiced attitudes also interacted with the students' ethnicity to affect achievement in the European vs Māori model (see below). That is, when teachers' implicit prejudiced attitudes favored the student's ethnicity, students performed better in mathematics at end-of-year.

At Level 2, whether or not the teacher was in the intervention or the control group did not interact with the teachers' implicit prejudiced attitudes, with one exception (see Table 3b). In reading, when teachers had a high implicit prejudiced attitude in favor of Europeans vs. Asians, the students in the intervention teachers' classes (as compared to the control class) on average performed better at end-of-year in reading. This finding suggested that, in reading, the intervention may have helped buffer minority students from teachers with high implicit bias towards the ethnic majority (see Table 3b).

Overall, the addition of the Level 2 implicit prejudiced attitudes variables into model 4 reduced the prediction error (R^2 difference) in Level 2 end-of-year mathematics achievement by between 12 and 20% in mathematics, and between 0 and 12% for reading achievement.

At Level 1, the final models showed that the implicit prejudiced attitudes' interactions were unassociated with changes in reading achievement. However, several effects were found for mathematics scores. As noted above, when teachers' implicit prejudiced attitudes for European versus Māori favored the student's ethnicity, students performed better in mathematics at end-of-year. In keeping with van den Bergh's (2010) study, we calculated the size of this effect by multiplying the regression coefficient of TIAAAT European vs Māori and student ethnicity match ($\beta = .133$) by the possible range of TIAAAT European vs Māori scores (i.e., 2.10). The difference in achievement for a student whose ethnicity matched the teachers' implicit prejudiced attitudes versus one that mismatched could make a 28-point difference in students' end-of-year mathematics achievement (note that students are expected to gain an average of 30 points across a year on this standardized test).

In the European vs. Asian model (Table 2b, Fig. 1), when teachers' explicit expectations were low for all students in their class, Asian and European students performed at similar levels in mathematics at end-of-year. However, when teachers' explicit expectations for their class were high on average, the Asian students had relatively higher levels of achievement than their European counterparts (see Fig. 1). We calculated the size of this explicit

[2] See Data Analysis section as to how these data were converted to be modeled at Level 1.

expectation and ethnicity effect (β = −.234) for a teacher who stated their students' end-of-year achievement was expected to be average (4) versus moderately above average (6). This small difference in expectation was associated with a 47 point difference in the European versus Asian ethnic achievement gap, comparable to a one and a half years gain in achievement. However, this finding also needs to be understood in the context of the other significant interaction in the model between teachers' explicit expectations and their implicit prejudiced attitudes, which we discuss next (see Table 2b and Fig. 2).

Similar to our finding shown in Fig. 1, when teachers' explicit expectations for European and Asian students were low, whether or not the teachers' implicit prejudiced attitudes favored the student's ethnicity made relatively little difference to students' end-of-year mathematics achievement. However, if teachers had high explicit expectations for their class and if the teachers' implicit bias was in favor of the students' ethnicity, these students had higher mathematics scores. That is, high general explicit expectations for the class, coupled with implicit prejudiced attitudes in a student's favor, seemed to give an additional achievement boost in mathematics (see Fig. 2) (β = .15).

Together, the addition of the implicit prejudiced attitudes (and related interactions) in our final models led to a reduction in Level 1 prediction error (R^2 difference) of between .5 and .7% in mathematics and between .1 and .2% in reading. At Level 2, these variables contributed to a reduction in the prediction error of between 12 and 19% in mathematics, and between 0 and 11% for reading. Though these reductions may appear relatively small, the practical significance of these effects are substantial (i.e., in the case of mathematics achievement for Asian and European students, a small shift in teachers' explicit expectations from average to moderately above average was associated with a 47-point difference in European and Asian student achievement).

Overall, the final mathematics models explained 5–7% of Level 1 variance and 39% of Level 2 variance in mathematics achievement; in reading our models explained 62–64% of Level 1 variance and 14–39% of Level 2 variance. The difference in the amounts of variance explained at each level reflects the dominance of different level 1 and level 2 variables in the models. That is, teachers' explicit expectations measured at level 1 have a substantial impact on students' reading achievement, but not their mathematics achievement. Whereas implicit prejudiced attitudes (measured at Level 2) have almost no influence on students reading achievement, but do influence mathematics achievement.

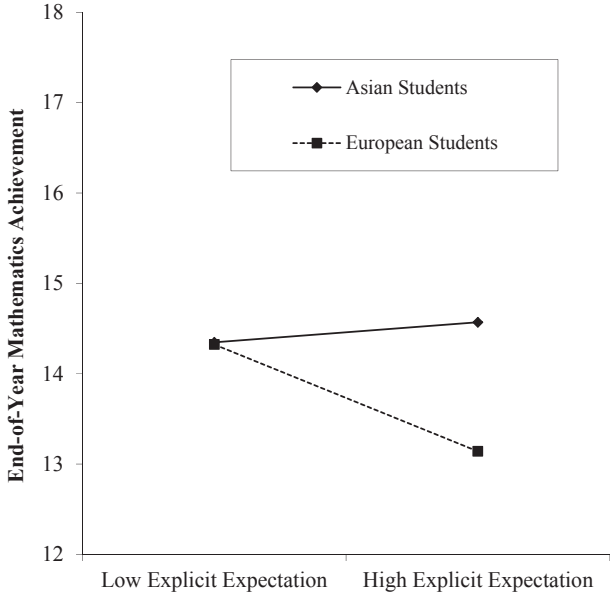


Fig. 1. Interaction between teachers' explicit expectation and student ethnicity, and their effect on European and Asian students' end-of-year mathematics achievement.

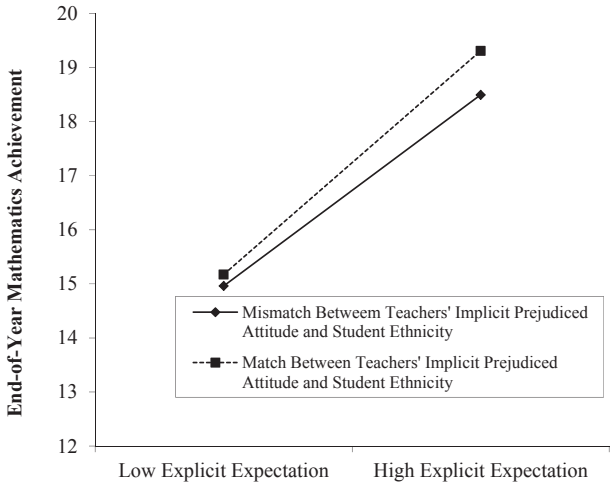


Fig. 2. Interaction between explicit expectations and teachers' implicit prejudice attitudes matched or mismatched with student ethnicity and their combined effects on European and Asian students' mathematics achievement.

4. Discussion

In this study, we examined whether teachers' explicit expectations and teachers' implicit prejudiced attitudes for the educational achievement of different ethnic groups related to students' end-of-year achievement and the ethnic achievement gap. In particular, we were interested in explaining the gap between European and Māori (and Asian and Māori) students, as this was where (both historically and in the present study) the biggest ethnic achievement gaps in New Zealand exist.

4.1. Ethnic differences in achievement outcomes and teachers' explicit expectations

Inspection of the mean achievement scores for each ethnic group, teachers' explicit expectations for each ethnic group, and teachers' implicit prejudiced attitudes for academic achievement by-in-large reflected the well-known trend whereby European and Asian students typically outperform Māori and Pasifika students. However, our results suggested that, when prior achievement was controlled for, these ethnic differences in students' end-of-year achievement and ethnic differences in teachers' explicit expectations disappeared, suggesting that teachers' explicit expectations may not have been contributing to the ethnic achievement gap. In order to unpack this finding further and to examine the possibility that teachers' implicit prejudiced attitudes might be affecting student achievement, we examined the effect of teachers' explicit expectations for mathematics and reading achievement, as well as their implicit prejudiced attitudes for general academic achievement, on student achievement in a series of ethnic-specific multilevel models.

Our multilevel models indicated that approximately 20% of the variance in mathematics achievement and 30% of the variance in reading achievement at end-of-year were attributable to factors residing at the teacher/school level (e.g., teachers' implicit prejudiced attitudes, school socioeconomic status, and the teacher group – intervention or control). These ICCs are comparable to those of other studies looking at teacher fixed effects (e.g., Ruzek, Domina, Conley, Duncan, & Karabenick, 2015; Mikami, Gregory, Pianta,

Allen, & Lun, 2011). Note, however, that researchers have also argued that multilevel modeling should be pursued regardless of how small the ICCs are, as small ICCs can still impact on significance tests when the number of individuals in a cluster is large (Heck & Thomas, 2015).

4.2. Ethnic achievement gap between European and Māori students

In our multi-level models, when prior ability was controlled for, European and Māori students' did not appear to be differentially sensitive to their teachers' explicit expectations for their achievement. That is, we found no interaction between student ethnicity (European and Māori) and teachers' explicit expectations and student achievement in reading or mathematics at end-of-year. European and Māori students did, however, seem to be differentially sensitive to teachers' implicit prejudiced attitudes, but only in mathematics. That is, if the direction of the teachers' implicit prejudiced attitude (e.g., a pro-European bias) favored the student's ethnicity (e.g., a European student), then the teacher's prejudiced attitude had a stronger effect on the student's mathematics achievement. Moreover, the degree to which a teacher's implicit prejudiced attitude was complementary with a student's ethnicity resulted in up to a 28-point difference in students' end-of-year mathematics achievement. This gain equated to almost a full year's academic advantage in mathematics. These findings highlight the importance of examining implicit prejudiced attitudes that teachers hold towards European and Māori students, as they may well be contributing to the widely-observed ethnic achievement gap in some subject areas. Note, however, the impact of the addition of these implicit effects is relatively small overall, only explaining an additional 12% of the Level 2 variance.

4.3. Ethnic achievement gap between Asian and Māori students

In keeping with the findings regarding European and Māori students' achievement, we found that Asian and Māori students were not differentially affected by their teachers' explicit expectations for their reading achievement. That is, we found no interaction between student ethnicity (Asian and Māori) and teachers' explicit expectations on student achievement in reading. Also in keeping with our findings for European and Māori students, we found no effect of teachers' implicit prejudiced attitudes on students' reading achievement. Unfortunately, saddle-point estimation problems prevented us from examining if these findings would be the same for mathematics.

4.4. Ethnic achievement gap between European and Asian students

In New Zealand, the expected ethnic achievement gap between European and Asian students is less clear, as both groups tend to perform equally well across most academic subjects (although some research has suggested that Asians slightly outperform Europeans in mathematics; see Harker, 2006). Similar to our findings above, in reading, no interaction was found between teachers' explicit expectations and whether the student was European or Asian. This suggests that European and Asian students are not differentially sensitive to teachers' explicit reading expectations. We did find however, that European and Asian students performed similarly in classes with low explicit expectation teachers. But when teachers had on average high expectations for the whole class, Asian students' had higher mathematics scores than their European peers (see Fig. 1).

Our results also identified some effects at the implicit level. Specifically, we found that when teachers' explicit expectations were on average high *and* the teachers' implicit prejudiced was in favor of the students ethnicity, these students had on average a 15-point advantage at end-of-year, approximately half a year's gain. Together, these findings suggest that even traditionally high achieving groups of students, who are potentially exposed to fewer achievement stereotypes, are sensitive to the effects of having high expectation teachers whose implicit prejudiced attitude favors their ethnicity. Note, however, that overall the impact of the addition of these implicit effects is relatively small, only explaining an additional 20% of the Level 2 variance but still making a sizeable practical difference to students' learning outcomes.

4.5. Why do teachers' explicit expectations affect reading achievement and implicit prejudiced attitudes affect mathematics achievement?

As noted above, we found no interactions between teachers' explicit expectations and students' ethnicity. We did, however, show that teachers with generally high expectations had higher achieving students in reading, but not in mathematics. There are several possible explanations for this finding. First, elementary school teachers may be worse at predicting students' achievement in mathematics compared to reading. In New Zealand, there are more established procedures for assessing reading than mathematics and development in reading is more explicitly understood (see Wilkinson & Townsend, 2000). In addition, New Zealand's recent national numeracy project has shifted elementary teachers' focus from teaching skills and knowledge to teaching mathematics strategies (see Ministry of Education, 2015). This pedagogical shift may further account for difficulties in predicting student achievement.

Another possibility is that the lack of association between teachers' explicit mathematics expectations and their students' achievement occurs because the teachers' expectations in mathematics are largely accurate, and the inclusion of prior ability in all the models may have removed any additive effect. This would be in keeping with Jussim and Harber's (2005) claim that teacher expectations predict achievement because they are accurate reflections of students' ability.

A third possibility of why significant effects were found for teachers' explicit expectations in reading, but not in mathematics, is because elementary school teachers' assessments of students' reading abilities tend to be communicated more explicitly to students than their mathematics abilities. For example, New Zealand elementary school students are frequently given reading books to take home to read to their parents, as homework, and these books clearly indicate the reading level of the child on the cover. In fact, one of the authors of this paper noted in her own five-year-old's class, a wall chart clearly displaying the reading levels of students relative to other students in the class; no such chart was visible for performance in mathematics.

Further, New Zealand has the highest within-class ability grouping rate of all OECD countries (Wagemaker, 1993). Moreover, students' reading achievement is often the basis on which teachers form these ability groups, with classroom seating plans and activities often undertaken within these groups (Rubie-Davies, 2015). Interestingly, research among New Zealand elementary teachers has found that low expectation teachers kept their students in ability groups for learning activities, whereas the high expectation teachers did not (Rubie-Davies, 2006, 2007). High expectation teachers also gave students choices of learning activities and allowed them to work with a range of peers (Rubie-Davies, 2015).

Taken together, these findings suggested that teachers' explicit expectations may have a stronger effect on achievement in reading than on mathematics because teachers' reading expectations are simply more salient to students. As such, teacher expectations about reading may be more likely to be internalized, thereby

influencing students' own sense of self-efficacy and ultimately their achievement (Hallam, Ireson, & Davies, 2004). Although we do not know the teaching practices of the current sample of teachers, it seems that teachers with high and low explicit expectations for students' reading achievement may have different pedagogical practices across these two subject areas (e.g., reading ladders with students' names clearly displayed, but no mathematics ladder) which make students more sensitive to teachers' explicit reading expectations. In contrast, we found that students were more likely to be sensitive to teachers' implicit prejudiced attitudes in mathematics. This opposite effect probably reflects the fact that teachers' mathematics expectations are less salient to New Zealand elementary school students than teachers' reading expectations (Rubie-Davies, 2015), and, hence, the students may be picking up on teachers' more subtle implicit prejudiced attitudes.

These findings also highlight the importance of teachers expressing high explicit expectations for *all* students in *all* subjects, not just in reading. Adopting such an inclusive strategy may help desensitize students to teachers' implicit prejudiced attitudes that could leak out and send negative messages about students' capabilities (Babad, 2009). To be clear, we are not advocating that teachers put up wall charts of students' rankings in all subject areas to make teachers' explicit expectations more salient. Indeed, this approach will only serve to highlight differences in students' expected achievement. Rather, our findings suggest that teachers express high expectations for *all* students and adopt the practices often found among high expectation teachers such as flexible groupings as opposed to ability groups, which serve to highlight differential expectations. The employment of flexible grouping enables all students to work collaboratively with their peers, ensures that all students are challenged, and facilitates opportunities for high-level peer modeling (Rubie-Davies, 2015).

4.6. Other multilevel predictors of student mathematics and reading achievement

In four out of the five final multilevel models, students in higher socioeconomic schools tended to earn higher end-of-year achievement scores than students in lower socioeconomic schools. This finding is consistent with national trends showing that high socioeconomic schools tend to have higher achieving students than low socioeconomic schools (Hattie, 2008). There was also some evidence of girls performing worse than boys at end-of-year in mathematics in the European vs. Māori model. This finding is in keeping with gender stereotypes which posit that girls perform less well in mathematics than boys (Good & Findley, 1985). It is not clear why this gender effect was not found in the other mathematics models.

4.7. Limitations

This study has two main limitations. First, teachers in this study were taking part in a project designed to teach them about the behaviors of high expectation teachers. As such, half of the teachers in the larger study were randomly allocated to an intervention group and the other half to a control group. Teachers' initial explicit expectation data were collected at the beginning of the year (i.e., before the intervention began) and, hence, it is unlikely that teachers' participation in the study could have influenced their responding. Similarly, the Time 1 achievement data are unlikely to be related to the intervention given that they were collected early in the year. However, the implicit bias data were collected after the intervention began. Nevertheless, in four out of the five multilevel models we ran, no interaction was found between the group teachers were in and their implicit achievement expectations on student performance. The absence of this interaction suggested that the intervention did not affect the expression of teachers' implicit prejudiced attitudes.

The one interaction that was found was in reading. Specifically, we found that teachers who had a higher implicit bias in favor of Europeans vs. Asians had students with higher reading scores compared those in the control group. As noted above, this finding suggested that, in reading, the intervention may have helped buffer minority students from teachers with high implicit bias towards the ethnic majority.

The second limitation was that we were unable to explore whether students' individual SES was associated with their teachers' expectations. Schools are not legally allowed to collect individual socioeconomic data about students. For this reason, we used the schools' nationally determined socioeconomic ranking. This ranking system is commonly used within New Zealand as a broad proxy for the socioeconomic status of the students within the school. Future research needs to explore whether the ethnic differences found in teachers implicit and explicit achievement expectations relate to the students' socioeconomic status.

4.8. Significance for theory, policy, and practice

We have demonstrated, for the first time, the impact of teachers' implicit prejudiced attitudes for academic achievement and teachers' explicit expectations for academic achievement on students' reading and mathematics achievement. Importantly, we have also controlled for prior achievement. Our findings suggested that future teacher expectation research would benefit from the inclusion of both explicit and implicit measures, and that including both may lead to different outcomes and a greater understanding of the ethnic achievement gap.

We also found that teachers with higher explicit expectations for students in reading had, on average, higher achieving students in reading. This highlights the importance of teachers having high explicit expectations for all students. In addition, we found that teachers' explicit expectations did not vary with students' ethnicity. This finding is important because the New Zealand Ministry of Education actively discourages teachers from setting low expectations for Māori, as evident in its recent Māori education strategy (Ka Hikitia) which has as one of its five key principles the setting of high expectations for Māori. Our findings suggested that, at the explicit expectations level, this strategy seemed to be working. Indeed, teachers' explicit expectations were not contributing to the European-Māori ethnic achievement gap in either reading or mathematics.

However, some teachers' implicit prejudiced attitudes did seem to have an ethnic bias, and these biases may be more likely to play a role when teachers' explicit expectations are less obvious (i.e., in mathematics). In our study, teachers' implicit prejudiced attitudes to educational achievement related to students' end-of-year mathematics achievement, but not to their reading achievement. This subject difference may be a reflection of the relative prominence of teachers' explicit expectations for students in reading and the relatively low prominence of teachers' mathematics expectations within New Zealand schools. Hence, it may be that, in mathematics where there is an absence of any obvious explicit expectations, students perceive their teachers' nonverbal cues about what is expected of them. In turn, these beliefs may be internalized and influence students' mathematics achievement.

These findings suggested that teachers need to be explicit about having high expectations for achievement in *all* subject areas and for *all* students. Taking such an approach may help buffer students from any nonverbal cues that may suggest the teachers do not have high achievement expectations. Indeed, the fact that the Teacher Expectation Project intervention buffered the effect of teachers' implicit biases on reading achievement for European vs. Asian

students suggested that the intervention may be a vehicle through which teachers' implicit biases can be controlled or masked. More research, however, is needed to confirm this finding.

Training programs that help teachers identify and reject discriminatory thoughts may also be useful in seeking to reduce their potential negative impact on some ethnic groups (Glock & Kovacs, 2013). Research has shown that, although implicit attitudes are automatic, they can be controlled if people are willing and able to exert cognitive control (Blair, 2002). Overall, more research is needed to understand the role of implicit and prejudiced attitudes within educational contexts. We therefore join Glock and Kovacs (2013) in their call for more educational research using implicit measures with a view that doing so may help reduce the potential impact of attitude driven biases on the future educational outcomes, careers and lives of our children.

Acknowledgment

We gratefully acknowledge the support of the teachers, students and the schools that took part in the Teacher Expectation Project (TEP) which was in part funded by a Marsden Fast Start Grant which is a New Zealand government grant administered by the Royal Society of New Zealand and a Cognition Trust Grant. We thank the teachers, students and for welcoming us into their schools and classroom and working with us. We also thank those on the TEP team who also assisted data collection and data entry which was no small task given the size of the project.

Appendix

Table 1
Symbols associated with success and failure used in the teachers' implicit attitude to academic achievement task.

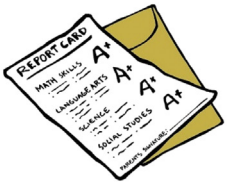
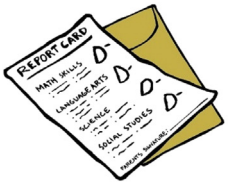
Success	Failure
	
	
	
	
	

Table 2
Surnames used in the teachers' implicit attitude to academic achievement task.

Ethnic group	Surnames
European	Johnson, Jones, Morris, Wills, Smith, Clarke
Māori	Tawhara, Wairau, Wikaira, Tautari, Ngatai
Asian	Wong, Leung, Nguyen, Li, Chan, Hunag

References

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: the case of shy behavior. *Journal of Personality and Social Psychology, 83*, 380–393. http://dx.doi.org/10.1037//0022-3514.83.2.380.

Babad, E. (2009). Teaching and nonverbal behavior in the classroom. In *International handbook of research on teachers and teaching* (pp. 817–827). US: Springer. http://dx.doi.org/10.1007/978-0-387-73317-3_52.

Babad, E., & Taylor, P. J. (1992). Transparency of teacher expectancies across language, cultural boundaries. *The Journal of Educational Research, 86*, 120–125.

Bishop, R., Berryman, M., Cavanagh, T., & Teddy, L. (2009). Te Kotahitanga: addressing educational disparities facing Māori students in New Zealand. *Teaching and Teacher Education, 25*, 734–742. http://dx.doi.org/10.1016/j.tate.2009.01.009.

Blair, I. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*, 242–261. http://dx.doi.org/10.1207/S15327957PSPR0603_8.

de Boer, H., Bosker, R. J., & Van der Werf, M. (2010). Sustainability of teacher expectation bias effects on long-term student performance. *Journal of Educational Psychology, 102*, 168–179. http://dx.doi.org/10.1037/a0017289.

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*, 365–274.

Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming* (2nd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. New York: Guilford Press.

Cormack, D., & Robson, C. (2010). *Classification and output of multiple ethnicities: Issues for monitoring Māori health*. Wellington: Te Rōpū Rangahau Hauora a Eru Pōmare.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: a normative analysis and review. *Psychological Bulletin, 135*, 347–368. http://dx.doi.org/10.1037/a0014211.

Delpit, L. (1995). *Other people's children: Cultural conflict in the classroom*. New York: The New Press.

Department of Statistics. (1993). *New Zealand standard classification of ethnicity, 1993*. Wellington: Department of Statistics.

Devine, P. G. (1989). Stereotypes and prejudice: their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5–18. http://dx.doi.org/10.1037/0022-3514.56.1.5.

Dovidio, J. F., & Fazio, R. H. (1992). New technologies for the direct and indirect assessment of attitudes. In J. M. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys* (pp. 204–237). New York, NY: Russel Sage Foundation.

Dovidio, J. F., Kawakami, K., & Beach, K. R. (2008). Implicit and explicit attitudes: examination of the relationship between measures of intergroup bias. In *Blackwell handbook of social psychology: Intergroup processes* (pp. 175–197). Malden, MA: Blackwell. http://dx.doi.org/10.1111/b.9781405106542.2002.00011.x.

Fazio, R. H. (2001). On the automatic activation of associated evaluations: an overview. *Cognition and Emotion, 15*, 115–141. http://dx.doi.org/10.1080/026999300 4200024.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline? *Journal of Personality and Social Psychology, 69*, 1013–1027.

Friese, M., Hofmann, W., & Schmitt, M. (2009). When and why do implicit measures predict behavior? Empirical evidence for the moderating role of opportunity, motivation, and process reliance. *European Review of Social Psychology, 19*, 285–338.

Fryer, R. G., Jr., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics, 86*(2), 447–464. http://dx.doi.org/10.1162/003465304323031049.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*(5), 692–731. http://dx.doi.org/10.1037/0033-2909.132.5.692.

Gawronski, B., Strack, F., & Bodenhausen, G. V. (2009). 4. Attitudes and cognitive consistency: the role of associative and propositional processes. In *Attitudes: Insights from the new implicit measures* (pp. 85–118).

Glock, S., & Karbach, J. (2015). Preservice teachers' implicit attitudes toward racial minority students: evidence from three implicit measures. *Studies in Educational Evaluation, 45*, 55–61. http://dx.doi.org/10.1016/j.stueduc.2015.03.006.

Glock, S., & Kovacs, C. (2013). Educational psychology: using insights from implicit attitude measures. *Educational Psychology Review, 25*, 503–522. http://dx.doi.org/10.1007/s10648-013-9241-3.

Good, T. L., & Findley, M. J. (1985). Sex role expectations and achievement. In J. B. Dusek (Ed.), *Teacher expectancies* (pp. 271–302). Hillsdale, NJ: Lawrence Erlbaum.

Greenwald, A. G., & Babaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem and stereotypes. *Psychological Review, 102*, 4–27.

Greenwald, A. G., & Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79*, 1022–1038. http://dx.doi.org/10.1037/0022-3514.79.6.1022.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197–216. http://dx.doi.org/10.1037/0022-3514.85.2.197.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*, 17–41. http://dx.doi.org/10.1037/a0015575.

Hallam, S., Ireson, J., & Davies, J. (2004). Primary pupils' experiences of different types of grouping in school. *British Educational Research Journal, 30*, 515–533.

Harker, R. (2006). *Ethnicity and school achievement in New Zealand: Some data to supplement the Biddulph et al. (2003) best Evidence Synthesis*. Ministry Of Education Report. Retrieved from http://www.victoria.ac.nz/education/pdf/whakapiki/ethnicity_school_achievement_nz_harker_2006.pdf.

Hattie, J. (2008). Narrow the gap, fix the tail, or close the curves: the power of words. In C. M. Rubie-Davies, & C. Rawlinson (Eds.), *Challenging thinking about teaching and learning* (pp. 19–23). New York, NY: Nova.

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analysis relating to achievement*. London: Routledge.

Heck, R. H., & Thomas, S. L. (2015). *An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus*. Routledge.

Huss-Keeler, R. L. (1997). Teacher perception of ethnic and linguistic minority parental involvement and its relationships to children's language and literacy learning: a case study. *Teaching and Teacher Education, 13*, 171–182.

Jencks, C., & Phillips, M. (Eds.). (2011). *The black-white test score gap*. Washington, D.C.: Brookings Institution Press.

Jussim, L., Eccles, J., & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: accuracy and the quest for the powerful self-fulfilling prophecy. *Advances in Experimental Social Psychology, 28*, 281–388.

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review, 9*, 131–155. http://dx.doi.org/10.1207/s15327957pspr0902_3.

Jussim, L., Robustelli, S. L., & Cain, T. R. (2009). Teacher expectations and self-fulfilling prophecies. In K. R. Wentzel, & A. Wigfield (Eds.), *Handbook of motivation in school*. New York: Routledge.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, New York: The Guilford Press.

Kreft, I. G., Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modelling*. London: Sage.

Magnuson, K. A., & Duncan, G. J. (2006). The role of family socioeconomic resources in the black-white test score gap among young children. *Developmental Review, 2*, 365–399. http://dx.doi.org/10.1016/j.dr.2006.06.004.

Mass, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*, 86–92. http://dx.doi.org/10.1027/1614-1881.1.3.86.

McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio, & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91–125). San Diego, CA: Academic Press.

McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context and the achievement gap. *Journal of School Psychology, 46*, 235–261. http://dx.doi.org/10.1016/j.jsp.2007.05.

Mikami, A. Y., Gregory, A., Allen, H. J. P., Pianta, R. C., & Lun, J. (2011). Effects of a teacher professional development intervention on peer relationships in secondary classrooms. *School Psychology Review, 40*, 367–385.

Ministry of Education. (2012). *Ka Hikitia – Accelerating success: 2013-2017*. Wellington, New Zealand: Minstry of Education.

Ministry of Education. (2015). *The numeracy development projects & number framework*. Available online http://nzmaths.co.nz/numeracy-development-projects-number-framework.

No Child Left Behind Act of 2002 Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

Nosek, B. A., Bar-Anan, Y., Sriram, N., & Greenwald, A. G. (2013). *Understanding and using the brief Implicit Association Test: I. Recommended Scoring Procedures*. Available at: http://dx.doi.org/10.2139/ssrn.2196002.

Rakić, T., Steffens, M. C., & Mummendey, A. (2011). Blinded by the accent! The minor role of looks in ethnic categorization. *Journal of Personality and Social Psychology, 100*, 16–29. http://dx.doi.org/10.1037/a0021522.

Rubie-Davies, C. M. (2006). Teacher expectations and student self-perceptions: exploring relationships. *Psychology in the Schools, 43*, 537–552.

Rubie-Davies, C. M. (2007). Classroom interactions: exploring the practices of high- and low-expectation teachers. *British Journal of Educational Psychology, 77*, 289–306. http://dx.doi.org/10.1348/000709906X101601.

Rubie-Davies, C. M. (2008). Teacher beliefs and expectations: relationships with student learning. In C. M. Rubie-Davies, & C. Rawlinson (Eds.), *Challenging thinking about teaching and learning*. Haupaugge, NY: Nova.

Rubie-Davies, C. M. (2015). *Becoming a high expectation teacher: Raising the bar*. London, UK: Routledge.

Rubie-Davies, C. M., & Peterson, E. R. (2011). Teacher expectations and beliefs: influences on the socioemotional environment of the classroom. In C. M. Rubie-Davies (Ed.), *Educational psychology: Concepts, research and challenges*. London: Routledge.

Rubie-Davies, C. M., Peterson, E. R., Sibley, C. G., & Rosenthal, R. (2014). A teacher expectation intervention: modelling the practices of high expectation teachers. *Contemporary Educational Psychology, 40*, 72–78. http://dx.doi.org/10.1016/j.cedpsych.2014.03.003.

Ruzek, E. A., Domina, T., Conley, A. M., Duncan, G. J., & Karabenick, S. A. (2015). Using value-added models to measure teacher effects on students' motivation and achievement. *The Journal of Early Adolescence, 35*, 852–882. http://dx.doi.org/10.1177/0272431614525260.

Sammons, P. (1995). Gender, ethnic and socio-economic differences in attainment and progress: a longitudinal analysis of student achievement over 9 years. *British Educational Research Journal, 21*, 465–485.

Singh, A. K. (2015). *Social psychology*. New Delhi, India: PHI Learning.

Stangor, C., & O'Brien, J. (2010). Stereotyping. In J. Levine, & M. Hogg (Eds.), *Encyclopedia of group processes and intergroup relations* (pp. 857–862). Thousand Oaks, CA: SAGE. http://dx.doi.org/10.4135/9781412972017.n265.

Statistics New Zealand. (2013a). *Demographic trends 2013*. Retrieved from http://www.stats.govt.nz/browse_for_stats/population/estimates_and_projections/demographic-trends-2010/chapter1.aspx.

Statistics New Zealand (2013b). Retrieved from http://www.stats.govt.nz/Census/2013-census.aspx.

Strand, S. (2013). *Insights: What accounts for ethnic achievement gap in secondary schools in England issue 4*. Autumn: Retrieved from http://www.education.ox.ac.uk/wordpress/wp-content/uploads/2013/12/Insights4_Ethnic-Achievement_Strand.pdf.

Strand, S. (2014). Ethnicity, gender, social class and achievement gaps at age 16: intersectionality and getting it for the white working class. *Research Papers in Education, 29*, 131–171. http://dx.doi.org/10.1080/02671522.2013.767370.

Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology, 99*, 253–273. http://dx.doi.org/10.1037/0022-0663.99.2.253.

Van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M. J., & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers: relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal, 47*, 497–527. http://dx.doi.org/10.3102/0002831209353594.

Vescio, T., & Bloodhart, B. (2010). Discrimination. In J. Levine, & M. Hogg (Eds.), *Encyclopedia of group processes and intergroup relations* (pp. 212–218). Thousand Oaks, CA: SAGE. http://dx.doi.org/10.4135/9781412972017.n65.

Wagemaker, H. (Ed.). (1993). *Achievement in reading literacy: New Zealand's performance in a national and international context*. Wellington, New Zealand: Ministry of Education.

Weinstein, R. S., Gregory, A., & Strambler, M. J. (2004). Intractable self-fulfilling prophecies fifty years after Brown v. Board of Education. *American Psychologist, 59*, 511–520.

Wilkinson, I. G., & Townsend, M. A. R. (2000). From Rata to Rimu: grouping for instruction in best practice New Zealand classrooms. *The Reading Teacher, 53*, 460–471.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*, 101–126. http://dx.doi.org/10.1037//0033-295X.107.1.101.