

**Exploring the Impact of Student Gender and Ethnicity on the Stability of Teacher
Expectations**

Emma Aue

S4335163

Master's Thesis: PAMAOW05

Supervisor: Dr. A.C. Timmermans

Second Evaluator: Dr. M. I. Deunk

May 31, 2024

Abstract

Teacher expectation stability about student achievement was investigated for mathematics and reading during one academic year in New Zealand. Additionally, teacher expectation stability based on student characteristics, namely, gender and ethnicity, was investigated to test whether teachers' expectations are based on stereotypes about their students. Previous research has shown that students adapt to teachers' expectations about their performance. Therefore, it is essential to investigate whether teachers adjust their expectations over time if students perform differently from what was expected. It was hypothesized that teacher expectations were more stable for girls in mathematics, boys in reading, and minority groups in both domains compared to majority groups. The results of the regression and comparison of correlations show that teacher expectations remained relatively stable throughout the school year while being more stable towards the end of the year. The analysis yielded statistically insignificant differences between boys and girls in mathematics and reading. Regarding differences between minority and majority groups, some significant results were found, especially for Pacific Island minority students, but no systematic differences were found between ethnicities. Overall, it remains unclear if stability in teacher expectations, which are not accurate based on student achievement, is due to stereotypes or other influences. Further investigation is needed to understand which factors influence whether a teacher adjusts their expectations.

Keywords: Teacher Expectation Stability, Student Achievement, Gender, Ethnicity, Stereotypes.

Exploring the Impact of Student Gender and Ethnicity on the Stability of Teacher Expectations

What teachers expect of their students is influenced by specific student characteristics such as gender, ethnicity, or socioeconomic status (e.g., Timmermans & Rubie-Davies, 2022; Wang et al., 2018). Research has shown that teacher expectations, once they are formed, often remain relatively stable despite changes in student achievement (Timmermans et al., 2021; Wang et al., 2020). Teacher expectations can have an impact on multiple student outcomes, such as achievement, performance, self-perception, and self-esteem (e.g., Johnston et al., 2019; Wang et al., 2018). However, to date, very little is known about whether teachers' expectations are equally stable for various groups of students. Therefore, this study aims to investigate whether teachers' expectations about students' performance are influenced by the ethnicity and gender of the students and how they change over the school year.

Teacher Expectations

Teacher expectations are “primarily cognitive phenomena, inferential judgments that teachers make about probable future achievement and behavior based upon the student’s past record and his present achievement and behavior” (Brophy & Good, 1974, p. 129). The expectations teachers have of their students can influence various student outcomes (Johnston et al., 2019). This has been demonstrated through educational research spanning over five decades, starting with the influential “*Pygmalion in the Classroom*” study by Rosenthal and Jacobson (1968). In this study, the expectations of teachers, as influenced by the researchers in an experiment, created a self-fulfilling prophecy effect, where “teacher’s false expectations had become true” (Jussim & Harber, 2005, p. 133). However, the results of the Pygmalion study have also been criticized by researchers because of measurement problems and inadequate data analysis (Snow, 1969). Although the study itself is criticized in terms of

procedure and techniques, the existence of teacher expectation effects is not (e.g., Brothy, 1983; Good et al., 2018). The studies conducted since then collectively underscore the significant impact of teacher expectations on student achievement in education (Johnston et al., 2019). Hence, teacher expectations are pivotal in shaping students' academic experiences and outcomes (Timmermans et al., 2021; Johnston et al., 2019).

Jussim and Harber (2005) concluded after a review of studies and meta-analyses investigating whether teacher expectations have self-fulfilling prophecy effects that while the self-fulfilling effects in educational environments are genuine and can be substantial at times, they are typically small. However, other reviews, including a different variety of studies, conclude that the effects are moderately large (e.g., Wang et al., 2018). Most studies show that when teachers hold high expectations, their students rise to meet them (e.g., Wang et al., 2018; Johnston et al., 2019). Conversely, students might adjust their behavior to meet lower expectations set by teachers, particularly if they come from already marginalized backgrounds (Johnston et al., 2019).

Factors Influencing Teacher Expectations

Multiple aspects, such as students' prior performance, ethnicity, gender, and other student characteristics, can influence the expectations a teacher holds about a student (Rubie-Davies, 2006; Wang et al., 2018) and influence subsequent teaching behavior (Rubie-Davies, 2007). Johnston and colleagues (2019) found in their meta-analysis that teacher expectations can have an influence on student outcomes, yet this influence varies across different contexts and among individual students. Overall, research indicates that for students who have similar levels of accomplishment in mathematics, teachers might have different expectations based on the student's gender or ethnicity (Timmermans & Rubie-Davies, 2022).

Investigating the effect of student ethnicity on the formation of teacher expectations, most studies have shown that “[...] negative achievement stereotypes and lower teacher

expectations exist for African American and Latino students in the US, for Indigenous students in Canada, for Māori and Pacific Island students in New Zealand, and students with immigration backgrounds in Europe” (Johnston et al., 2019, p. 130). Some research has found that teachers have lower expectations for ethnic minority groups, especially for boys in reading and girls in mathematics (Wang et al., 2018). This relationship, however, seems inconsistent as some studies did find this effect while others did not (Wang et al., 2018). For example, Van den Bergh and colleagues (2010) found that teachers typically had varying expectations for students depending on the student’s ethnic background. Further, disparities in achievement between ethnic minority students and those of Dutch origin were observed in all classrooms (Van den Bergh et al., 2010). Okura (2022) found that Asian students in the United States gain advantages from teachers' racial-based expectations. This implies that varying teacher expectations contribute to educational disparities among racial groups (Okura, 2022). This can lead to positive or negative consequences for the students depending on the direction of the expectations (e.g., Meissel et al., 2017). Overall, if differences between ethnicities are found, the effect sizes vary from small to moderately large (e.g., Rubie-Davies et al., 2006).

Whether gender has an impact on the expectations of teachers of their students remains controversial (e.g., Glock & Krolak-Schwerdt, 2013; Holder & Kessels, 2017). Mixed results have been found in the past, where an effect was found or not (Wang et al., 2018) or where an effect was only found in interaction with another variable (e.g., Auwarter & Aruguete, 2008). Past studies have shown that teachers often hold higher expectations for girls in literacy subjects and boys in mathematics (e.g., Holder & Kessel, 2017; Wang et al., 2018). Contrary to this, Auwarter and Aruguete (2008) concluded that gender only had an effect on teacher expectations when socioeconomic status (SES) was taken into account. Namely, low-SES female students were rated more favorably by teachers than low-SES male

students (Auwarter & Aruguete, 2008). However, some studies suggested that there are no significant differences between gender groups (Wang et al., 2018). Overall, gender differences often seem to be relatively small (e.g., Auwarter & Aruguete, 2008; Johnston et al., 2018) but sometimes moderately large (e.g., Dickhauseer & Meyer, 2006). One possible mechanism by which differences between gender and ethnicity in teacher expectations arise could be stereotyping (Rubie-Davies et al., 2006).

Stereotypes

Sometimes, teacher expectations are based on biased beliefs about student characteristics (Dee, 2005; Glock & Krolak-Schwerdt, 2013), and the impact of those expectations can be even more harmful. Biased beliefs often rely on stereotypes, defined as “a belief about a group of individuals” (Kanahara, 2006, p. 311). These stereotypes are generalizations of group characteristics that can manifest in different ways. They can either be positive or negative and might be accurate or based on a misconception (Glock & Krolak-Schwerdt, 2013; Kanahara, 2006). For example, some people have the stereotype that boys are generally better at mathematics than girls (e.g., Nosek et al., 2009).

Persistent cultural beliefs about the academic abilities and potential of students from various ethnic backgrounds have been found to reduce teachers' expectations, consequently leading to unequal educational opportunities (Allen & Webber, 2019). This means that teachers who hold stereotypes about specific student groups relate them to the abilities of students who are part of those groups (Allen & Webber, 2019) and then hold lower expectations for those students (e.g., Glock, 2016; Holder & Kessels, 2017). Additionally, the acceptance of negative racial stereotypes by influential figures, such as teachers, can impact academic results (Allen & Webber, 2019). In the context of New Zealand schools, prevailing stereotypes about the capabilities of ethnically diverse students, particularly Māori and Pacific Island students, are predominantly negative, portraying them as less intelligent and

lacking academic engagement (Allen & Webbers, 2019). Studies have indicated that students from culturally diverse backgrounds are more susceptible to the effects of stereotypes compared to other students (Ford et al., 2008).

Gender stereotypes are structured beliefs about attributes of women and men, such as interests and competencies (Ashmore & Del Boca, 1979, as cited in Kollmayer et al., 2016). The effects of gender stereotypes in education include differential ratings on students' abilities (e.g., Holder & Kessels, 2017), differences in treatment or provision of learning opportunities (e.g., Wang et al., 2018), or suggestions made by teachers for career pursuit (Kollmayer et al., 2016). While most teachers expect girls to go in the direction of education or medicine, they expect boys to enter STEM (Kollmayer et al., 2016). These stereotyped beliefs have an impact on teacher behavior, which then impacts student outcomes, such as performance or motivation (Kollmayer et al., 2016).

Influenced by strong stereotypes that teachers hold about specific student characteristics, their expectations about those students can become self-fulfilling prophecies (Jussim & Harber, 2005). Literature suggests that teachers hold different expectations about male and female students (Li, 1999). The deep-rooted cultural views and beliefs that suggest females are less proficient in math and related subjects can limit girls' accomplishments and career choices, as they are influenced by stereotypes that mirror societal expectations, standards, and values (Andersen, 2023). A tendency seems to exist for teachers to stereotype boys as being better in mathematics than girls, which shows through the overestimation of boys' abilities in mathematics (Li, 1999). Therefore, teachers might expect females to be less proficient in math than boys. In reading achievement, boys generally perform lower than girls (Watson et al., 2019). Previous research has shown that teachers have formed the assumption that their male students are not as proficient in reading compared to their female counterparts (Watson et al., 2019). Similarly, students from ethnic minority backgrounds tend to have

lower academic performance compared to their peers from the ethnic majority (Dee, 2005). While numerous factors play a role in these disparities, the stereotyped expectations held by teachers often lead to teachers anticipating that students from ethnic minorities perform less well than their ethnic majority counterparts (Dee, 2005; Lorenz, 2021).

Teacher Expectation Stability

While teacher expectations have been extensively researched, only a few research studies have focused on the stability of those expectations (Rubie-Davies et al., 2018). Stability of expectations refers to the consistency of teachers' expectations over time, namely whether they hold onto their initial assumptions despite conflicting evidence or modify their expectations based on a student's development throughout the year (Rubie-Davies et al., 2018). Researching teacher expectation stability requires multiple measures of teacher expectations as well as student achievement over time (e.g., Timmermans et al., 2021). It is crucial to investigate the stability of those to understand the possibility of self-fulfilling prophecies (e.g., Timmermans et al., 2021) because continuous over- or underestimation could potentially increase the likelihood of self-fulfilling prophecies.

Early research on the stability of teachers' expectations indicated that formed expectations by teachers remain stable (Martinek, 1980). However, the longitudinal research was only conducted over a period of two months (Martinek, 1980), and limited background information on how the study was conducted was provided, which makes drawing broader conclusions difficult. More recent research, however, suggests that teachers' expectations remain fairly stable over the school year, and despite working with different cohorts of students, educators often perceive the potential and abilities of their students in a consistent manner (Rubie-Davies et al., 2018). Additionally, research conducted by Wang et al. (2020) showed that teachers consistently overestimated high-achieving students, and the degree of this overestimation increased throughout the school year. Conversely, students categorized as

low achieving were persistently underestimated throughout the school year. Nonetheless, recent research shows that teachers take past student performance into account and adjust their initial expectations accordingly for some students, especially at the beginning of the school year. This process of adaptation persists throughout the academic year (Timmermans et al., 2021). As expectations become more precise, the likelihood of self-fulfilling prophecies emerging and further impacting future student performance diminishes (Timmermans et al., 2021). A study by Hao and colleagues (2022) investigated teacher expectation bias stability in China. The results showed that teachers adjusted their initial expectations during the first few months of teaching the students, but the adjusted expectation bias remained stable afterward. Additionally, the study found that most teachers either over- or underestimated their students across two school years.

Considering that, to date, only a few studies have investigated the stability of teacher expectations, it is crucial that further longitudinal research is conducted. The studies that investigated stability usually used various time lags ranging from eight weeks (Martinek, 1980) over three to five months (Good et al., 1980; Rubie-Davies et al., 2018; Wang et al., 2020) to six months (Kuklinski & Weinstein, 2000) and even one year (Clifton, 1981). Additionally, the studies conducted earlier than 2018 did not make use of control variables (e.g., Good et al., 1980; Kuklinski & Weinstein, 2000). Without controlling for, e.g., student achievement, it is difficult to say whether the finding of stability is due to stability in teacher expectations or stability in student performance as they correspond to a large extent (Timmermans et al., 2021). This was also shown by Wang et al. (2020), where the correlations indicative of stability was clearly lower after controlling for student performance. Therefore, to be able to draw meaningful conclusions, more studies with similar time lags and inclusion of control variables are needed.

Stereotypes and Expectation Stability

Research has shown that the perceptions teachers have of their students may be influenced by stereotypical beliefs (Tiedemann, 2000). Reliance on stereotypical beliefs rather than objective evaluation could be indicated if expectations change for different student groups under similar circumstances or if expectations remain stable despite changes in the behavior or performance of the individual (e.g., Hao et al., 2022).

Additionally, research by Kuklinski and Weinstein (2000) posits that teachers who are more susceptible to biases from stereotypical information about students might be more inclined to stick to their initial expectations, potentially overlooking information that contradicts those beliefs. They found that in settings where teachers clearly distinguished between high and low achievers, both in tasks assigned and interactions, their expectations tended to remain steady. Conversely, teachers who approached all students with a uniform standard were more inclined to modify their expectations for students over time. Further research is necessary to understand the role of stereotypes in educational settings, especially their influence on the stability of teacher expectations.

Research Questions

Researching the stability of teacher expectations in the context of stereotyping is crucial as it is the goal to promote equity, ensuring every student has a fair chance of success. It should not be limited because of teachers' biased expectations (Watson et al., 2019). Understanding the impact of stereotypes can limit negative self-fulfilling prophecies for particular groups of students and guide the development of teacher training programs to foster a more understanding and unbiased educational environment. This leads to the following research questions: To what extent do teacher expectations of individual students' academic performance in reading and mathematics change or remain stable over the academic year (RQ 1)? To what extent is the level of teacher expectation stability dependent on students' gender (RQ 2)? To what extent is the level of teacher expectation stability

dependent on students' ethnicity (RQ 3)? To what extent is the level of teacher expectation stability dependent on students' gender and ethnicity combined (RQ 4)?

In this study, conducted in New Zealand, it is hypothesized that teachers' expectations of student's future academic performance will generally remain stable over the school year (H1) (Rubie-Davies et al., 2018) and that due to stereotyped beliefs about specific student characteristics, low and high teacher expectations remain more stable. Namely, suppose a teacher has strong stereotyped beliefs about girls generally not being capable at math. In that case, this will remain relatively stable despite having female students in math class who are very proficient (Andersen, 2023). It is hypothesized that teachers' expectations of future mathematics performance are likely to be more stable for girls than for boys (H2a), and teachers' expectations of future reading performance are likely to be more stable for boys than for girls (H2b) (Watson et al., 2019). Further, for teachers who hold the stereotype that Māori or Pacific Island students (minority) are low achieving, it is expected that teachers' expectations of future academic performance are likely to be more stable for minority students than for majority students (H3). Lastly, it is hypothesized that teacher expectations of future academic performance are likely to be more stable for minority girls than for majority girls (H4a) and minority boys than for majority boys (H4b) (Timmermans & Rubie-Davies, 2022).

Method

Context

This study will use an existing data set by Rubie-Davies (e.g., Rubie-Davies et al., 2018) collected in New Zealand. The compulsory education sector in New Zealand consists of elementary and high school levels. The elementary school attended from year one to year eight, where students are between five and twelve years old. Afterward, students attend high school from Year 9 until Year 13 (Ministry of Education, 2023). Most students attend their

local elementary school and do not travel (Ministry of Education, 2023). The New Zealand government ranks each school on a scale (1 = *school in a very disadvantaged area* to 10 = *school in a wealthy area*), which is mainly based on the socioeconomic information for that specific area. However, other information, such as the percentage of minority group families, is also included (Rubie-Davies et al., 2018). The funding is allocated inversely, which means that schools in less affluent areas are awarded substantially more funding compared to those in middle-class regions (e.g., Timmermans et al., 2021). Further, schools in New Zealand operate autonomously by a board that includes the principal, a staff member, and multiple members from the community who oversee the management of the school (Timmermans et al., 2021). For the current data set, schools were divided into high, middle, and low socioeconomic areas. Following this categorization, schools were randomly selected from these socioeconomic groups and asked to join the study. This ensured a diverse sample of students across different socioeconomic backgrounds. Subsequently, teachers from 12 selected schools agreed to participate in the research (Rubie-Davies et al., 2018).

Design

The current study is longitudinal panel research in order to be able to compare whether teacher expectations remain stable over the course of the school year (Timmermans et al., 2021; Rubie-Davies et al., 2018). The data was collected three times throughout the school year in 2014. The data regarding teacher expectations is collected through questionnaires administered to teachers, and additional achievement information is gathered from students' achievement tests.

Participants

The final dataset included 2536 students in primary schools (Year 4 – 8; aged 8 – 13). The data set includes the gender of students (51% male and 49% female) and their ethnicity from which their minority or majority status can be inferred. 49.3% of students were New

Zealand European, 17.1% Māori (Indigenous), 15.8% from Pacific Islands (Pacifica), 14.2% Asian, and 3.6% from other ethnic backgrounds.

Additionally, the data set included 89 teachers (28,6 % male; 71.4% female), their years of teaching experience (1-5 years 32.1%; 6-10 years 23.3%; 11-17 Years 18.9%; 18-25 Years 7.8%; > 25 Years 18.0%), and the years of teaching experience in New Zealand ($M = 10.5$).

Materials

Teacher Expectations

Teacher expectations regarding both mathematics and reading were measured first in the month of February, which is the start of the school year. Although they have not yet carried out their own standardized assessment of the students, teachers have received information about the students from the previous teachers, encompassing the achievement levels (Timmermans et al., 2021). The teacher expectation questionnaire was also administered in June, the middle of the school year, and October, which is at the end of the school year.

Teachers indicated on a 7-point Likert scale the level they expected each individual student to reach at the end of the school year (1 = *very much below average* to 7 = *very much above average*). This was only a single measure item that did not allow for psychometric assessment (Timmermans et al., 2021). Using a single item is, however, a common approach in teacher expectation research (Timmermans et al., 2021; Rubie-Davies et al., 2018; Kuklinski & Weinstein, 2001; Bohlmann & Weinstein, 2013).

Student Achievement

Further, contemporaneous to the measurement of teacher expectations, student scores of reading and mathematics achievement are recorded three times, namely in February, June, and October, using Electronic Assessment Tools for Teaching and Learning (e-asTTle), which

is a standardized measure often used in New Zealand (Timmermans et al., 2021). In this study, each test was designed to have a duration of 40 minutes. Teachers were responsible for choosing the test level they deemed most suitable for the diverse needs of the students in their classes, ensuring that each student received the test that best matched their individual capabilities. The test scores were standardized using national norms in order to compare the results of the three measurement points (Rubie-Davies et al., 2018; Timmermans et al., 2021). This resulted in normed scores per domain per measurement occasion. In mathematics, number knowledge, algebra, and number sense were tested (Timmermans et al., 2021), and scores ranged from 1194 to 1765. For reading, strategies, ideas, processes, and language features were tested (Rubie-Davies et al., 2018), and scores ranged from 1165 to 1780.

Procedure

Rubie-Davies obtained ethical approval from her educational institution to conduct the research (Rubie-Davies et al., 2018; Timmermans et al., 2021). Approval of the parents for the participation of their children was obtained before conducting the longitudinal study. As part of the study, what the teachers expect of each of their students and student's academic performance in mathematics and reading using the e-asTTle at three different time points throughout the school year, namely in February, June, and October, was recorded.

Rubie-Davies created the tests, which were couriered to the teachers. After the students had completed the tests, the researchers graded them (Rubie-Davies et al., 2018; Timmermans et al., 2021).

Data Analysis

Throughout the analysis, IBM SPSS Statistics (Version 29, 2023) was used, and a p-value of .05 was used as the significance threshold.

Missing Values

Prior to the data analysis, the data was investigated for missing values. The original sample size was $N = 2483$. The dataset included some participants from phase 2, which were excluded from this study, and one student was excluded because of a missing indication of the phase. For the predictor variables, namely student gender and ethnicity, there were no missing values. For teacher expectations, it was decided to exclude all students ($n = 21$) from one teacher (id = 83) because no teacher expectations were measured from that teacher. Further, there were no other missing values for teacher expectations in reading or mathematics. After excluding those cases, the dataset included $N = 2461$ cases. For the achievement tests in Mathematics, the number of missing values for time points one ($n_{\text{missing}} = 556$), two ($n_{\text{missing}} = 608$), and three ($n_{\text{missing}} = 965$) was investigated. Additionally, the number of missing values for reading achievement at time points one ($n_{\text{missing}} = 632$), two ($n_{\text{missing}} = 708$), and three ($n_{\text{missing}} = 957$) were investigated as well. Despite missing values on these variables, cases with missing values on the reading or mathematics tests were retained for the analyses. SPSS does pairwise exclusion for the variables used for the analysis; therefore, there are different sample sizes throughout the full analysis.

Analysis of Hypotheses

Before beginning the analysis, the assumptions for regression and correlation were investigated. The residuals were continuous and linearly related. Due to the large sample size, it can be assumed that the data is normally distributed (Agresti, 2018). Additionally, outliers were investigated, which ranged from $n = 1$ (math measurement occasion 3) to $n = 12$ (reading measurement occasion 3). No major differences were found between the correlations, whether the outliers were included or excluded in the dataset (see syntax). Therefore, it was decided not to remove the outliers.

To investigate the more general research question if teachers' expectations about individual students' performance remain stable over the school year, Wang et al. (2020) and

Rubie-Davies et al. (2018) data-analytic procedures were used as a model on how to analyze the data. First, teacher expectations for each student were regressed, using a linear regression model, against their actual achievements within the same measurement occasion for mathematics and reading separately. From each of the regression models, an unstandardized residual for each student was stored, indicating to what extent the observed teacher's expectations at that measurement occasion were above or below the expected teacher expectation, given the students' achievement, thereby gauging the accuracy of these expectations. For example, the unstandardized residual in the domain mathematics for measurement occasion one will be referred to as "Math 1" and for reading "Reading 1" etc. A positive correlation indicates a relationship between two variables, high values on the one variable are associated with high values on the other variable and vice versa (Agresti, 2018). A positive correlation between the residuals of subsequent measurements indicates that teachers keep overestimating the same students as well as underestimating the same students. Therefore, high correlations are indicative of stability. Comparing the correlations between the residuals of the three different measurement time points with the Fisher Z test (Lenhard & Lenhard, 2014) indicated whether a teacher tends to overestimate or underestimate throughout the school year (high correlation) or whether the expectations change and are not stable over the school year (low correlation). Following Wang and colleagues (2020) guidelines, the interpretation of the stability correlation is as follows: strong stability ($r > 0.7$), moderate stability ($0.7 > r > 0.5$), some stability ($0.5 > r > 0.3$) and low stability ($r < 0.3$).

To test the other research questions, namely, to what extent teacher expectations are dependent on students' gender and ethnicity, the function of splitting the data set was used. The data file was split and organized by groups for both gender and ethnicity separately and thereafter together. For example, when splitting for gender, two groups are generated, namely

boys and girls. Then the analysis of correlations is conducted as before for the groups separately. Subsequently, the correlations between the residuals of the teacher expectations and the student achievement scores of each group were calculated to investigate the relationship between those two variables at different time points. Afterward, the correlations between the groups (e.g., boys and girls) were compared using the Fisher Z test (Lenhard & Lenhard, 2014) to check for significant differences.

Results

The results of the statistical analysis for the regression and the analysis of the stability of teacher expectations across the school year will be presented below.

Regression Models

The regression models between teacher expectations and students' achievement at the three time points for mathematics and reading can be found in Table 1. In all models, teacher expectations are positively related to student achievement within the same domain and measurement point.

Table 1

Regression Models for Reading and Mathematics

	B	R	Rsquared	S.E.	ΔF	df1	df2	Sig.
Math 1	.009	.526	.276	1.160	727.246	1	1903	<.001
Math 2	.011	.612	.374	1.118	1107.193	1	1852	<.001
Math 3	.005	.300	.090	1.295	147.896	1	1495	<.001
Reading 1	.008	.509	.259	1.235	638.917	1	1827	<.001
Reading 2	.010	.634	.401	1.155	1174.592	1	1752	<.001
Reading 3	.011	.672	.452	1.082	1239.597	1	1503	<.001

Note. Dependent variable: Teacher expectations; Predictors: Constant, Normed scores

Stability

The first hypothesis, namely that teacher expectations of student performance remain relatively stable over the school year and that low expectations remain more stable, was

investigated first. Table 2 shows the unstandardized residual correlations between teacher expectations and student achievement in mathematics and reading.

Table 2

Unstandardized Residual Correlations for Mathematics and Reading

	<i>n</i>	<i>SD</i>	1		2	
			<i>r</i>	<i>n</i>	<i>r</i>	<i>n</i>
1 Math 1	1905	1.159	–	–	–	–
2 Math 2	1853	1.115	.446*	1596	–	–
3 Math 3	1496	1.295	.542*	1278	.709*	1335
1 Reading 1	1829	1.234	–	–	–	–
2 Reading 2	1753	1.153	.544*	1433	–	–
3 Reading 1	1504	1.081	.568*	1239	.801*	1225

* $p < .001$.

Overall, it can be said that all correlations for mathematics and reading between the three time points were positive and significant but varied in strength. In mathematics, the correlation between timepoints one and two indicated some stability ($r(1594) = .446, p < .001$), between the first and third timepoint moderate stability ($r(1278) = .542, p < .001$), and strong stability ($r(1335) = .709, p < .001$) between the second and third measurements. In reading the correlations between the first and the second ($r(1433) = .544, p < .001$) and the first and the third ($r(1239) = .568, p < .001$) measurements both indicate moderate stability, while the correlation between the second and third measurements indicated strong stability ($r(1225) = .801, p < .001$). Comparing the correlations in Mathematics of the first half and the second half of the school year indicated a larger stability in the second half of the school year ($z = -7.100, p < .001$). In reading, there is evidence for larger stability as well in the second half of the school year compared to the first half ($z = -12.619, p < .001$).

Gender

Table 3 shows the unstandardized residual correlations between teacher expectations and student achievement in mathematics and reading split by gender, namely boys and girls. To test H2, namely, whether teacher expectations are more stable for girls in mathematics (a) and for boys in reading (b), differences in correlations between boys and girls were tested for significance. No significant differences were found for any of the correlations between girls and boys for either reading or mathematics.

Table 3

Unstandardized Residual Correlations for Mathematics and Reading and Comparison of Correlations Between Boys vs. Girls

	Measurement Occasions	Boys		Girls		<i>z</i>	<i>p</i>
		<i>r</i>	<i>n</i>	<i>r</i>	<i>n</i>		
Mathematics	1, 2	.439*	793	.454*	803	-.374	.354
	1, 3	.523*	628	.561*	650	-.96	.169
	2, 3	.696*	662	.724*	673	-1.035	.152
Reading	1, 2	.535*	692	.553*	741	-.455	.324
	1, 3	.547*	594	.587*	645	-1.035	.150
	2, 3	.792*	615	.812*	640	-.991	.161

Note. 'Measurement Occasions' = Indication for which measurement occasions correlations were assessed.

* $p < .001$.

Ethnicity

Tables 4 and 5 show the unstandardized residual correlations between teacher expectations and student achievement split for ethnicities. To test H3, namely whether teacher expectations are more stable for minority students than for majority students, differences in correlations between ethnicities were tested for significance.

In mathematics, the following correlations were significantly different from one another. For measurement occasions one and two, there was a significant difference between

New Zealand Europeans and Māori students, teacher expectations were significantly more stable for New Zealand European students than for Māori ($z = 1.797, p = .036$). Between New Zealand Europeans and Pacifica students, teacher expectations were also significantly more stable for New Zealand Europeans ($z = 3.456, p < .001$). Additionally, teacher expectations were significantly more stable for Asian students than for Pacifica students ($z = -2.796, p = .003$). For measurement occasions one and three, teacher expectations were significantly more stable for New Zealand European students than for Māori ($z = 1.797, p = .041$). Additionally, teacher expectations were significantly less stable for Māori students than for Asian students ($z = -1.776, p = .038$). Comparing the correlations between the second and third measurement occasions shows that teacher expectations were significantly more stable for Pacifica students than for New Zealand European ($z = -2.258, p = .012$), Māori ($z = -2.109, p = .017$), and Asian ($z = 2.583, p = .005$).

In reading, the following correlations had significant differences. For measurement occasions one and two, teacher expectations were significantly more stable for New Zealand European students than for Pacifica ($z = 2.408, p = .008$) and Asian students ($z = 2.166, p = .015$). For measurement occasions one and three, no significant differences were found between ethnicities. Additionally, for measurement occasions two and three, teacher expectations were significantly more stable for Pacifica students than for New Zealand European ($z = -2.064, p = .019$) and Asian ($z = 3.096, p = .001$). Further, Teacher expectations were significantly less stable for Asian students than for Māori students ($z = 1.922, p = .027$) and New Zealand European students ($z = 1.794, p = .036$). Some comparisons between ethnicities with the category “Other” are significant. However, it is not possible to draw meaningful conclusions from these results, as the category “Other” exists of a large variety of backgrounds combining groups for whom there are positive stereotypes and

groups for whom negative stereotypes apply. Therefore, the results are only mentioned in the Tables and not in text.

Table 4

Unstandardized Residual Correlations for Mathematics and Comparison of Correlations Between Ethnicities

Measurement Occasions	Ethnicity	<i>r</i>	n	Ethnicity	<i>z</i>	<i>p</i>	
1,2	NZ/European	.510*	804	Māori	1.797	.036	
				Pacifica	3.456	< .001	
				Asian	-0.072	.514	
				Other	0.727	.234	
	Māori	.411*	276	Pacifica	1.451	.073	
				Asian	-1.457	.073	
				Other	-0.074	.471	
	Pacifica	.299*	243	Asian	-2.796	.003	
				Other	-0.848	.198	
	Asian	.514*	227	Other	0.712	.237	
	1,3	NZ/European	.592*	612	Māori	1.735	.041
					Pacifica	1.340	.090
Asian					-0.511	.305	
Other					-1.946	.026	
Māori		.495*	216	Pacifica	-0.353	.362	
				Asian	-1.776	.038	
				Other	-2.668	.004	
Pacifica		.520*	229	Asian	-1.467	.071	
				Other	-2.477	.007	
Asian		.620*	174	Other	-1.535	.062	
2,3		NZ/European	.710*	673	Māori	0.352	.362
					Pacifica	-2.258	.012
	Asian				1.023	.153	
	Other				1.525	.064	
	Māori	.696*	216	Pacifica	-2.109	.017	
				Asian	0.582	.280	
				Other	1.287	.099	
	Pacifica	.786*	229	Asian	2.583	.005	
				Other	2.441	.007	
	Asian	.664*	176	Other	0.933	.175	

Note. Bold = Significant *p*-value at .05.

* *p* < .001.

Table 5

Unstandardized Residual Correlations for Reading and Comparison of Correlations Between Ethnicities

Measurement Occasions	Ethnicity	<i>r</i>	<i>n</i>	Ethnicity	<i>z</i>	<i>p</i>
1,2	NZ/European	.596*	724	Māori	1.583	.057
				Pacifica	2.408	.008
				Asian	2.166	.015
				Other	-0.566	.286
	Māori	.516*	254	Pacifica	0.744	.228
				Asian	0.642	.261
				Other	-1.200	.115
	Pacifica	.464*	226	Asian	-0.064	.474
				Other	-1.577	.057
	Asian	.469*	189	Other	-1.519	.064
	1,3	NZ/European	.596*	631	Māori	0.732
Pacifica					0.887	.188
Asian					1.300	.097
Other					-0.277	.391
Māori		.556*	199	Pacifica	0.128	.449
				Asian	0.480	.316
				Other	-0.609	.271
Pacifica		.547*	197	Asian	0.353	.362
				Other	-0.682	.247
Asian		.521*	186	Other	-0.890	.187
2,3		NZ/European	.791*	611	Māori	-0.521
	Pacifica				-2.064	.019
	Asian				1.794	.036
	Other				-2.029	.021
	Māori	.806*	217	Pacifica	-1.252	.105
				Asian	1.922	.027
				Other	-1.712	.043
	Pacifica	.844*	229	Asian	3.096	.001
				Other	-1.069	.142
	Asian	.723*	161	Other	-2.739	.003

Note. Bold = Significant *p*-value at .05.

**p* < .001.

Ethnicity and Gender

The following four tables show the correlations between the unstandardized residuals split for gender and ethnicity. To test H4, namely whether teacher expectations are more

stable for minority boys than majority boys (a) and for minority girls than for majority girls (b), unstandardized residual correlations were compared using the Fisher Z test. Boys and girls were compared within the same ethnicity in mathematics (Table 6) and in reading (Table 7). In mathematics, teacher expectations were more stable for New Zealand European girls than boys ($z = -1.652, p = .049$) and Asian boys than for girls ($z = 2.1655, p = .015$) for measurement occasions one and two. For measurement occasions one and three, teacher expectations were more stable for New Zealand European girls than for boys ($z = -1.655, p = .049$). In reading for measurement occasions one and two, teacher expectations were more stable for Māori girls than for boys ($z = -2.262, p = .012$), just as at time points one and three ($z = -2.229, p = .013$). At measurement occasions two and three, teacher expectations were more stable for Asian boys than for girls ($z = 1.703, p = .044$).

Table 6

Unstandardized Residual Correlations for Mathematics and Comparison of Correlations Between Boys and Girls

Measurement occasions	Ethnicity	Boys		Girls		z	p
		r	n	r	n		
1,2	NZ/European	.470*	400	.556*	404	-1.652	.049
	Māori	.373*	132	.440*	144	-0.659	.255
	Pacifica	.359*	128	.234*	115	1.055	.146
	Asian	.626*	105	.416*	122	2.164	.015
1,3	NZ/European	.549*	301	.636*	311	-1.655	.049
	Māori	.453*	102	.526*	114	-0.695	.243
	Pacifica	.524*	122	.508*	107	0.162	.435
	Asian	.639*	75	.609*	99	0.314	.337
2,3	NZ/European	.692*	342	.731*	331	-1.021	.154
	Māori	.652*	100	.734*	116	-1.146	.126
	Pacifica	.778*	121	.796*	108	-0.353	.362
	Asian	.699*	76	.642*	100	0.670	.252

Note. Bold = Significant p-value at .05.

* $p < .001$.

Table 7

Unstandardized Residual Correlations for Reading and Comparison of Correlations Between Boys and Girls

Measurement occasions	Ethnicity	Boys		Girls		z	p
		r	n	r	n		
1,2	NZ/European	.615*	358	.577*	366	0.790	.215
	Māori	.395*	118	.608*	136	-2.262	.012
	Pacifica	.460*	113	.467*	113	-0.066	.474
	Asian	.493*	81	.454*	108	0.336	.368
1,3	NZ/European	.654*	308	.631*	323	0.490	.312
	Māori	.432*	93	.655*	106	-2.229	.013
	Pacifica	.594*	101	.501*	96	0.920	.179
	Asian	.546*	68	.505*	100	0.354	.362
2,3	NZ/European	.777*	311	.809*	300	-1.062	.144
	Māori	.770*	97	.842*	120	-1.499	.067
	Pacifica	.841*	120	.849*	109	-0.209	.417
	Asian	.801*	67	.677*	94	1.703	.044

Note. Bold = Significant p-value at .05.

* $p < .001$.

Additionally, to further investigate H4, ethnicities were compared for boys and girls separately in mathematics (Table 8) and in reading (Table 9). For boys in mathematics at measurement occasions one and two, teacher expectations were significantly more stable for Asian than for New Zealand European ($z = -2.025, p = .021$), Māori ($z = -2.588, p = .005$), and Pacifica students ($z = -2.691, p = .004$). For measurement occasions one and three, there was significantly lower stability in expectations for Māori than for Asian students ($z = -1.730, p = .042$). For measurement occasions two and three, teacher expectations were significantly more stable for Pacifica students than for New Zealand Europeans ($z = -1.764, p = .039$) and Māori ($z = -1.908, p = .028$).

For girls in mathematics, at measurement occasions one and two, teacher expectations were significantly more stable for Māori than for Pacifica students ($z = 1.847, p = .032$).

Between measurement occasions one and three, there was higher stability for New Zealand European than for Pacifica students ($z = 1.688, p = .046$).

Table 8

Comparison of Unstandardized Residual Correlations Between Ethnicities for Boys and Girls in Mathematics

Measurement occasions	Ethnicity	Boys		Girls				
		z	p	z	p			
1,2	NZ/European	Māori	1.166	.373	Māori	1.581	.057	
		Pacifica	1.310	.095	Pacifica	3.636	<.001	
		Asian	-2.025	.021	Asian	1.764	.039	
		Other	1.188	.117	Other	-1.059	.145	
	Māori	Pacifica	0.129	.449	Pacifica	1.847	.032	
		Asian	-2.588	.005	Asian	0.236	.407	
		Other	0.581	.281	Other	-1.596	.055	
	Pacifica	Asian	-2.691	.004	Asian	-1.553	.060	
		Other	0.505	.307	Other	-2.426	.008	
	Asian	Other	2.105	.018	Other	-1.689	.046	
	1,3	NZ/European	Māori	1.108	.134	Māori	1.507	.066
			Pacifica	0.324	.373	Pacifica	1.688	.046
Asian			-1.063	.144	Asian	0.377	.353	
Other			-1.386	.083	Other	-1.300	.097	
Māori		Pacifica	-0.686	.246	Pacifica	0.180	.429	
		Asian	-1.730	.042	Asian	-0.881	.189	
		Other	-1.863	.031	Other	-1.871	.031	
Pacifica		Asian	-1.170	.121	Asian	-1.041	.149	
		Other	-1.471	.071	Other	-1.954	.025	
Asian		Other	-0.642	.260	Other	-1.398	.081	
2,3		NZ/European	Māori	0.634	.263	Māori	-0.059	.476
			Pacifica	-1.764	.039	Pacifica	-1.398	.081
	Asian		-0.105	.458	Asian	1.465	.071	
	Other		-1.904	.028	Other	0.233	.408	
	Māori	Pacifica	-1.908	.028	Pacifica	-1.108	.134	
		Asian	-0.559	.288	Asian	1.270	.102	
		Other	1.253	.105	Other	0.248	.402	
	Pacifica	Asian	1.175	.120	Asian	2.315	.010	
		Other	2.052	.020	Other	0.791	.214	
	Asian	Other	1.273	.101	Other	-0.388	.349	

Note. Bold = Significant p -value at .05.

Table 9

Comparison of Unstandardized Residual Correlations Between Ethnicities for Boys and Girls in Reading

Measurement occasions	Ethnicity	Boys		Girls				
		<i>z</i>	<i>p</i>	<i>z</i>	<i>p</i>			
1,2	NZ/European	Māori	2.789	.003	Māori	-0.471	.319	
		Pacifica	2.013	.022	Pacifica	1.394	.082	
		Asian	1.415	.079	Asian	1.518	.064	
		Other	0.382	.351	Other	-1.329	.092	
	Māori	Pacifica	-0.597	.275	Pacifica	1.548	.061	
		Asian	-0.834	.202	Asian	1.655	.049	
		Other	-0.845	.199	Other	-1.110	.133	
	Pacifica	Asian	-0.288	.386	Asian	0.121	.452	
		Other	-0.522	.301	Other	-1.824	.034	
	Asian	Other	-0.340	.367	Other	-1.878	.030	
	1,3	NZ/European	Māori	2.667	.004	Māori	-0.361	.359
			Pacifica	0.848	.198	Pacifica	1.634	.051
Asian			1.241	.107	Asian	1.614	.053	
Other			-1.212	.113	Other	1.023	.153	
Māori		Pacifica	-1.517	.065	Pacifica	1.631	.051	
		Asian	-0.923	.178	Asian	1.612	.054	
		Other	-2.448	.007	Other	1.129	.129	
Pacifica		Asian	0.445	.328	Asian	-0.037	.485	
		Other	-1.842	.061	Other	0.236	.407	
Asian		Other	-1.765	.039	Other	0.257	.399	
2,3		NZ/European	Māori	0.148	.441	Māori	-0.952	.171
			Pacifica	-1.720	.043	Pacifica	-1.135	.128
	Asian		-0.463	.322	Asian	2.509	.006	
	Other		-1.153	.125	Other	-1.676	.047	
	Māori	Pacifica	-1.475	.070	Pacifica	-0.183	.427	
		Asian	-0.500	.308	Asian	2.894	.002	
		Other	-1.156	.124	Other	-1.254	.105	
	Pacifica	Asian	0.792	.214	Asian	3.002	.001	
		Other	-0.387	.350	Other	-1.160	.123	
	Asian	Other	-0.819	.206	Other	-1.644	.004	

Note. Bold = Significant *p*-value at .05.

Discussion

The current study aimed to investigate teacher expectation stability about students' academic performance in mathematics and reading in New Zealand over one academic year.

Teacher expectations and student achievement were recorded at the beginning, middle, and end of the school year.

Hypotheses and Theoretical Implications

The results of this study provide evidence for the first hypothesis, namely that teacher expectations of student performance remain relatively stable over the course of a school year. This aligns with the findings of previous research (e.g., Kuklinski & Weinstein, 2000; Timmermans et al., 2021; Wang et al., 2020). As in the results by Wang and colleagues (2020) and Timmermans and colleagues (2021), the results of the current study indicate that teacher expectations were increasingly stable throughout the school year. The higher the stability of teacher expectation, the more likely it is to have a self-fulfilling prophecy effect for the student because continuous over- or underestimation means persistent inaccuracies that might influence student achievement over time (Timmermans et al., 2021). If expectations are adjusted, self-fulfilling prophecy effects are less likely, and there is a smaller effect of expectations in general (Timmermans et al., 2021) because the inaccuracy does not persist for long, and therefore, student achievement might not yet change.

No evidence was found for the second hypothesis. Namely, teacher expectations are more stable for girls in math (H2a) and for boys in reading (H2b). This is contrary to what was expected. However, while many studies found differences in expectations between girls and boys (e.g., Wang et al., 2018), some studies also did not find gender differences (Holder & Kessels, 2017). If differences were found, they were typically small (e.g., Johnston et al., 2018). Recently, there has been an increased awareness of social stereotypes regarding girls in STEM in New Zealand. Potentially, this made teachers more aware and led them to more careful consideration of girls' mathematic abilities (Timmermans et al., 2021).

Regarding the third hypothesis, namely that for teachers who hold the stereotype of Māori or Pacific Island (minority) students being low achieving, it is expected that teachers'

expectations of future academic performance are likely to be more stable for minority students than for the majority students (H3) it can be said that the results provide no clear support. While there are no systematic significant differences between majority or minority student ethnicities in the first half of the school year, there seem to be some significant differences between majority students and Pacific Island students as well as Asian students. Teacher expectations were more stable in the second half for Pacific Island students compared to all other ethnicities in both reading and mathematics. Previous research has found some differences in teacher expectations between minority and majority ethnicities (e.g., Johnston et al., 2018). However, clear support for this was not found in the current study. Despite not finding systematic differences between majority and minority groups, some statistically significant differences can be remarked on. Teacher expectations were relatively stable for New Zealand European students during the first half of the school year in both mathematics and reading. In mathematics, teacher expectations were more stable than for Māori and Pacific Island students; in reading, they were more stable for Māori than Pacific Island and Asian students. This is contrary to what was hypothesized, as the expectations for the majority group show more stability than for minority groups during the first half year. Past research has found that there were no significant differences between ethnicities in New Zealand in regard to teacher expectations and student achievement (Peterson et al., 2016). Due to this, it might be that there are no apparent differences between ethnicities regarding stability either. Future research could investigate whether the ethnicity of the teacher has an effect on teacher expectations, as this was not included in the current study. Teachers from the majority background might be more likely to stereotype minority students and favor their same-ethnicity students (e.g., Schuchart et al., 2021). In turn, teachers who are from a minority background as well and might experience discrimination or stereotyping themselves might be less likely to stereotype themselves because they are more

aware of the existence and potential inaccuracy. Overall, a match in ethnicity between teacher and student seems to be beneficial for student outcomes (e.g., Glock & Schuchard, 2020).

Therefore, future studies should take into account both students' and teachers' ethnicity when investigating teacher expectation stability.

Regarding the fourth hypothesis, which is that teacher expectations of future academic performance are likely to be more stable for minority girls than for majority girls (H4a) and minority boys than for majority boys (H4b), the results suggest no clear support. If significant differences were found, teacher expectations were more stable for Asian boys than girls, and for Māori and New Zealand European students, they were more stable for girls than for boys. No significant differences were found for Pacific Island students. Therefore, no conclusions about systematic differences can be drawn based on the results of this study. Past research has found that minority boys are especially vulnerable to low expectations about their reading achievement (Hinnant et al., 2009). Further research is needed to explore potential differences in teacher expectation stability for a combination of student characteristics.

Strengths and Limitations

Considering the few studies on teacher expectation stability (e.g., Martinek, 1980; Wang et al., 2020), this longitudinal study is a necessary addition to understanding this complex concept. The study made use of a large and somewhat representative sample as schools from high, middle, and low SES areas were included. Whether the results are generalizable outside of New Zealand is unknown. However, since the main finding, namely that teacher expectations remained relatively stable throughout the school year, has also been found in different studies conducted in, for example, China (Wang et al., 2020), it might be the case. Additionally, this study is one of the only studies investigating student characteristics as an explanation for stability in teacher expectations (Wang et al., 2020).

Another strength is that this study controlled for student achievement, which has only been done in some studies about teacher expectation stability (Timmermans et al., 2021).

Additionally, some limitations have to be remarked on. In this study, only the domains of mathematics and reading were investigated, where the correct answers are usually relatively straightforward (Timmermans et al., 2021). Therefore, it is still unclear which effect teacher expectation stability may have in domains where teachers' interpretations are more subjective (Timmermans et al., 2021). Potentially, teachers may have less difficulty adapting their expectations if they have access to “clear and objective student achievement information” (Timmermans et al., 2021, p. 10). Future research should also focus on different domains to provide a more in-depth picture, even in subjects where the evaluation of student performance may be more subjective. Another limitation of this study may be the missing data. While the data set included a large sample, there were also many students for whom only some measurements were taken. For example, only student achievement at the first two time points but not the last. It was decided to include those cases still and use pairwise exclusion during the statistical analysis. This, however, could have an impact on the results as it is unknown whether the teacher's expectations would have been stable or unstable for those students. Additionally, future research could not only investigate teacher expectation stability but also include whether expectations were higher or lower than student achievement to understand the impact of stability. Further, the study only used quantitative data to understand the stability of teacher expectations. To explore more in-depth how these expectations influence behavior, qualitative data could give additional insights (e.g., Rubie-Davies et al., 2018). Future research should adopt a qualitative or mixed-methods approach as it could provide more insight into the mechanisms behind teacher expectations. For example, by asking the teachers to specify why they have certain expectations.

Practical Implications

Overall, research shows that teachers should maintain high expectations for all students. This is even more beneficial than having accurate expectations, as often students rise to meet those expectations (e.g., Timmermans et al., 2021; Wang et al., 2018). Additionally, teachers should be aware of possible biases they might have towards specific groups of students (e.g., Clifton, 1981). Teacher training programs and workshops for practicing teachers should include awareness of teacher expectations, their stability, and effects (Timmermans et al., 2021) to avoid low expectations, which lead to inequality in education if those are based on, for example, ethnicity or gender of students (e.g., Clifton, 1981). It is crucial that teachers do not share too much information about their students with the teachers who are taking over a course, as they may adopt the same incorrect expectations (e.g., Borko & Niles, 1982; Rubie-Davies et al., 2018). It may be helpful, especially for students for whom a teacher had low expectations, to be able to have a fresh start and succeed academically without the teacher's limiting expectations.

Conclusion

Teacher expectations seem to remain relatively stable over time. This can be beneficial for some students but might disadvantage others, as higher teacher expectations have been associated with higher student achievement (e.g., Jussim & Harber, 2005; Timmermans et al., 2015). Teachers should maintain higher expectations and only adjust them if it benefits the student. Low teacher expectations that are not adjusted may lead to injustice and inequality in education (Clifton, 1981; Tenenbaum & Ruck, 2007).

References

- Agresti, A. (2018). *Statistical Methods for the Social Sciences, Global Edition*. Pearson.
- Allen, J. M., & Webber, M. (2019). Stereotypes of minorities and education. *The Palgrave Handbook of Ethnicity*, 1-21. https://doi.org/10.1007/978-981-13-0242-8_107-1
- Andersen, I. G. (2023). Teachers' gender bias in STEM: Results from a vignette study. *British Educational Research Journal*, 49(4), 833–851. <https://doi-org.proxy-ub.rug.nl/10.1002/berj.3870>
- Auwarter, A. E., & Aruguete, M. S. (2008). Effects of student gender and socioeconomic status on teacher perceptions. *The Journal of Educational Research*, 101(4), 242-246.
- Bohlmann, N. L., & Weinstein, R. S. (2013). Classroom context, teacher expectations, and cognitive level: Predicting children's math ability judgments. *Journal of Applied Developmental Psychology*, 34, 288–298. <https://doi.org/10.1016/j.appdev.2013.06.003>
- Borko, H., & Niles, J. (1982). Factors Contributing to Teachers' Judgments about Students and Decisions about Grouping Students for Reading Instruction. *Journal of Reading Behavior*, 14(2), 127–140. <https://doi.org/10.1080/10862968209547441>
- Brophy, J. E. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology*, 76, 236–247. <https://doi.org/10.1037/0022-0663.75.5.631>
- Brophy, J. E., & Good, T. L. (1974). Teacher-student relationships: Causes and consequences. Holt, Rinehart & Winston.
- Clifton, R. A. (1981). The Effects of Students' Ethnicity and Sex on the Expectations of Teachers. *Interchange on Educational Policy*, 12(1), 31-38. <https://doi.org/10.1007/BF01807395>

- Dee, T. S. (2005). A teacher like me: does race, ethnicity, or gender matter? *The American Economic Review*, *95*(2), 158–165. Doi: 10.1257/000282805774670446
- Dickhauser, O., & Meyer, W. J. P. S. (2006). Gender differences in young children's math ability attributions. *Psychology Science*, *48*(1), 3.
- Ford, D. Y., Grantham, T. C., & Whiting, G. W. (2008). Culturally and linguistically diverse students in gifted education: Recruitment and retention issues. *Exceptional Children*, *74*(3), 289-306. <https://doi.org/10.1177/001440290807400302>
- Glock, S. (2016). Does ethnicity matter? The impact of stereotypical expectations on in-service teachers' judgments of students. *Social Psychology of Education: An International Journal*, *19*(3), 493–509. <https://doi.org/10.1007/s11218-016-9349-7>
- Glock, S., & Krolak-Schwerdt, S. (2013). Does Nationality Matter? The Impact of Stereotypical Expectations on Student Teachers' Judgments. *Social Psychology of Education: An International Journal*, *16*(1), 111–127. <https://doi.org/10.1007/s11218-012-9197-z>
- Glock, S., & Schuchart, C. (2020). The Ethnic Match between Students and Teachers: Evidence from a Vignette Study. *Social Psychology of Education: An International Journal*, *23*(1), 27–50. <https://doi-org.proxy-ub.rug.nl/10.1007/s11218-019-09525-2>
- Good, T. L., Cooper, H. M., & Blakey, S. L. (1980). Classroom interaction as a function of teacher expectations, student sex, and time of year. *Journal of Educational Psychology*, *72*(3), 378. <https://doi.org/10.1037/0022-0663.72.3.378>
- Good, T. L., Sterzinger, N., & Lavigne, A. (2018). Expectation effects: Pygmalion and the initial 20 years of research. *Educational Research and Evaluation*, *24*(3-5), 99-123. <https://doi.org/10.1080/13803611.2018.1548817>
- Hao, L., Rubie-Davies, C. M., & Watson, P. W. S. J. (2022). Do teachers maintain their expectation bias for students? a longitudinal investigation. *Social Psychology of*

Education: An International Journal, 25(4), 719–744. <https://doi.org/10.1007/s11218-022-09714-6>

Hinnant, J. B., O'Brien, M., & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology*, 101(3), 662–670. Doi: 10.1037/a0014306

Holder, K., & Kessels, U. (2017). Gender and ethnic stereotypes in student teachers' judgments: A new look from a shifting standards perspective. *Social psychology of education*, 20, 471-490. <https://doi.org/10.1007/s11218-017-9384-z>

IBM Corp. (2023). IBM SPSS Statistics for Windows, Version 29.0 [Computer software]. IBM Corp.

Johnston, O., Wildy, H., & Shand, J. (2019). A decade of teacher expectations research 2008–2018: Historical foundations, new developments, and future pathways. *Australian Journal of Education*, 63(1), 44-73. <https://doi.org/10.1177/0004944118824420>

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, 9(2), 131–55.

Kanahara, S. (2006). A Review of the Definitions of Stereotype and a Proposal for a Progressional Model. *Individual Differences Research*, 4(5), 306–321.

Kollmayer, M., Schober, B., & Spiel, C. (2018). Gender stereotypes in education: Development, consequences, and interventions. *European Journal of Developmental Psychology*, 15(4), 361-377. <http://dx.doi.org/10.1080/17405629.2016.1193483>

- Kuklinski, M. R., & Weinstein, R. S. (2000). Classroom and grade level differences in the stability of teacher expectations and perceived differential treatment. *Learning Environments Research*, 3, 1–34. doi:10.1023/A:1009904718353
- Kuklinski, M. R., & Weinstein, R. S. (2001). Classroom and developmental differences in a path model of teacher expectancy effects. *Child Development*, 72, 1554–1578. <https://doi.org/10.1111/1467-8624.00365>
- Lenhard, W. & Lenhard, A. (2014). Hypothesis Tests for Comparing Correlations. *Psychometrica*. Retrieved March 2024, from <https://www.psychometria.de/correlation.html>. DOI: 10.13140/RG.2.1.2954.1367
- Li, Q. (1999). Teachers' beliefs and gender differences in mathematics: A review. *Educational Research*, 41(1), 63-76. <https://doi.org/10.1080/0013188990410106>
- Lorenz, G. (2021). Subtle Discrimination: Do Stereotypes among Teachers Trigger Bias in Their Expectations and Widen Ethnic Achievement Gaps? *Social Psychology of Education: An International Journal*, 24(2), 537–571. <https://doi-org.proxy-ub.rug.nl/10.1007/s11218-021-09615-0>
- Martinek, T. J. (1980). Stability of teachers' expectations for elementary school-aged children. *Perceptual and Motor Skills*, 51, 1269–1270. <https://doi.org/10.2466/pms.1980.51.3f.1269>
- Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education*, 65, 48-60. <https://doi.org/10.1016/j.tate.2017.02.021>
- Ministry of Education. (2023). Education in New Zealand. Retrieved from <https://www.education.govt.nz/our-work/our-role-and-our-people/education-in-nz/>

- Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., et al. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, *106*, 10593–10597. doi:10.1073/pnas.
- Okura, K. (2022). Stereotype Promise: Racialized Teacher Appraisals of Asian American Academic Achievement. *Sociology of Education*, *95*(4), 302–319. <https://doi-org.proxy-ub.rug.nl/10.1177/00380407221119746>
- Peterson, E. R., Rubie-Davies, C., Osborne, D., & Sibley, C. (2016). Teachers' explicit expectations and implicit prejudiced attitudes to educational achievement: Relations with student achievement and the ethnic achievement gap. *Learning and Instruction*, *42*, 123-140. <http://dx.doi.org/10.1016/j.learninstruc.2016.01.010>
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*. Holt, Rinehart & Winston.
- Rubie-Davies, C. M. (2006). Teacher expectations and student self-perceptions: Exploring relationships. *Psychology in the Schools*, *43*, 537–552. doi:10.1002/pits.20169
- Rubie-Davies, C. M. (2007). Classroom interactions: Exploring the practices of high- and low-expectation teachers. *British Journal of Educational Psychology*, *77*, 289–306. <https://doi.org/10.1348/000709906X101601>
- Rubie-Davies, C., Hattie, J., & Hamilton, R. (2006). Expecting the best for students: Teacher expectations and academic outcomes. *British Journal of Educational Psychology*, *76*(3), 429-444. <https://doi.org/10.1348/000709905X53589>
- Rubie-Davies, C. M., Watson, P. W., Flint, A., Garrett, L., & McDonald, L. (2018). Viewing students consistently: How stable are teachers' expectations? *Educational Research and Evaluation*, *24*(3–5), 221–240. <https://doi.org/10.1080/13803611.2018.1550836>

- Schuchart, C., Glock, S., & Dunkake, I. (2021). The influence of in-group and out-group favouritism on the disciplinary practice of ethnic majority and minority preservice teachers. *Social Psychology of Education: An International Journal*, 24(3), 691–715. <https://doi-org.proxy-ub.rug.nl/10.1007/s11218-021-09627-w>
- Snow, R. E. (1969). Unfinished pygmalion. Review of the book *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*, by R. Rosenthal & L. Jacobson]. *Contemporary Psychology*, 14(4), 197–199. <https://doi.org/10.1037/0010293>
- 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(2), 253–273. <https://doi.org/10.1037/0022-0663.99.2.253>
- Tiedemann, J. (2000). Gender-related Beliefs of Teachers in Elementary School Mathematics. *Educational Studies in Mathematics*, 41, 191–207. doi:10.1023/A:1003953801526.
- Timmermans, A. C., & Rubie-Davies, C. M. (2022). Gender and minority background as moderators of teacher expectation effects on self-concept, subjective task values, and academic performance. *European Journal of Psychology of Education : A Journal of Education and Development*, 38(4), 1677–1705. <https://doi.org/10.1007/s10212-022-00650-9>
- Timmermans, A. C., Rubie-Davies, C. M., & Wang, S. (2021). Adjusting expectations or maintaining first impressions? The stability of teachers' expectations of students' mathematics achievement. *Learning and Instruction*, 75. <https://doi.org/10.1016/j.learninstruc.2021.101483>

Van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal*, *47*(2), 497-527.

Wang, S., Rubie-Davies, C. M., & Meissel, K. (2018). A systematic review of the teacher expectation literature over the past 30 years. *Educational Research and Evaluation*, *24*(3-5), 124-179. <https://doi.org/10.1080/13803611.2018.1548798>

Wang, S., Rubie-Davies, C. M., & Meissel, K. (2020). The stability and trajectories of teacher expectations: Student achievement level as a moderator. *Learning and Individual Differences*, *78*. <https://doi.org/10.1016/j.lindif.2019.101819>

Watson, P. W. S. J., Rubie-Davies, C. M., Meissel, K., Peterson, E. R., Flint, A., Garrett, L., & McDonald, L. (2019). Teacher gender, and expectation of reading achievement in New Zealand elementary school students: essentially a barrier? *Gender and Education*, *31*(8). <https://doi.org/10.1080/09540253.2017.1410108>