Teacher Expectation Stability for Student Achievement in Reading and Mathematics with Gender as a Moderator

Nadia Zandberg

S5327725

Department of Behavioral And Social Sciences, The University of Groningen

PAMAOW05: Master's Thesis

Supervisor: Dr. Anneke C. Timmermans

May 31, 2024

Abstract

Teacher expectations correlate with student academic achievement and are often biased due to societal stereotypes, contributing to gender disparities in student academic achievement. This study had two primary goals. The first was determining whether teacher expectations were stable for students' reading and mathematics achievement over one academic year. The second was to determine whether there was a significant difference in teacher expectation stability depending on teacher and student gender. Participants in this study were 83 teachers from 12 schools in New Zealand from varying socioeconomic regions. The student participants of this study were primary students between ages 8 and 10 and intermediate school students between 11 and 12 years old ($N = 2536$). Correlations coefficients of students' test scores and teacher expectation residuals were calculated to examine the teacher expectation stability for reading and mathematics achievement and compared to determine any gender differences. The results showed that teacher expectations were stable across one academic year and increased in stability towards the end of the year. Even though male and female teacher expectation stability differed at the beginning of the academic year for reading and mathematics, these differences faded out towards the end of the year. It was concluded that teacher expectations were stable over the year and that teacher and student gender did not impact teacher expectation stability.

Keywords: teacher expectation stability, student academic achievement, gender

Teacher Expectation Stability for Student Achievement in Reading and Mathematics with Gender as a Moderator

Teacher expectations play a pivotal role in shaping students' educational experiences, having the potential to influence student performance directly (Watson et al., 2019). Teacher expectations are the "inferences that teachers make about the future behaviour or academic achievement of their students, based on what they know about these students now" (Good & Brophy, 1997, p. 79). Teacher expectations were first documented by the pioneering study of Rosenthal and Jacobson (1968), called the "*Pygmalion in the Classroom*". The researchers concluded that teacher expectations and subsequent student achievement are correlated and that teacher expectations may operate as a self-fulfilling prophecy (Rosenthal & Jacobson, 1968). The self-fulfilling prophecy is "a false definition of the situation evoking a new behaviour which makes the originally false conception to come true" (Merton, 1948, p. 195). The Pygmalion study emphasised the importance of acknowledging the existence of teacher expectations effects and laid the foundation for further research (Brophy, 1983; Good et al., 2018; Rubie-Davies, 2010; Wang et al., 2018).

Since then, it has been confirmed that different teacher expectations may lead to varying treatment of students, leading to unequal educational opportunities (Kuklinski & Weinstein, 2000; Rubie-Davies, 2023). For example, teachers pose more complex questions and participate in high-order thinking activities with students they have high expectations for, and vice versa (Rubie-Davies, 2007). Wang et al. (2020) assert that students internalise these high and low teacher expectations and that performance differences among students correspond to different teacher expectations. Brophy and Good (1970) state that teachers must be aware that students are

exposed to teacher expectations through either verbal or nonverbal behaviours and understand the effects of their expectations on students' performance.

Although the investigation of whether teachers alter their expectations throughout an academic year has become more prominent, most studies fail to acknowledge changes in teacher expectations over extended time (Wang et al., 2020). This issue may be due to the great challenge of tracking teacher expectations for specific students for more than one academic year, as the students that teachers have may differ from year to year (Wang et al., 2020). Potential changes in teacher expectations over time are referred to as teacher expectation stability (Rubie-Davies et al., 2018). Teacher expectation stability is "whether or not teachers maintain their existing expectations in the face of contradictory information or whether they adjust their expectations as students progress through the year" (Rubie-Davies et al., 2018, p. 223). The current study investigated whether teacher expectations were stable for mathematics and reading achievement over one academic year and whether significant differences occurred between male and female teachers' expectation stability and for male and female students.

Teacher Expectation Stability and The Self-fulfilling Prophecy

The area of teacher expectations stability holds a significant research gap that has been identified by recent studies (Hao et al., 2022; Rubie-Davies et al., 2018; Timmermans et al., 2021; Wang et al., 2020). However, the few studies conducted so far were used to shed light on critical factors regarding teacher expectation stability and the self-fulfilling prophecy. For the self-fulfilling prophecy to be realised, Merton (1948) argued that teachers must over- or underestimate students when formulating their first expectations. Then, teachers must not alter their expectations and must keep them stable over time (Brophy, 1983). If teacher expectations and corresponding pedagogical behaviour remain stable, this continuum will allow the

self-fulfilling prophecy to occur, causing students to perform in line with the teacher's expectations (Timmermans et al., 2021). On the other hand, if teacher expectations become more accurate and correspond to students' actual performance, the self-fulfilling prophecy would likely not occur, and one can assume that teachers' expectations did not have significant effects on students' performances (Jussim & Harber, 2005).

There are two competing perspectives on teacher expectation stability (Timmermans et al., 2021). The first view is that teachers adjust their expectations for individual students to correlate with the students' fluctuating performance, causing some level of instability (Brophy, 1983). This approach causes expectations to vary over time, minimising the opportunity for the self-fulfilling prophecy to occur as teachers adjust their expectations as they get more familiar with students' performance over time (Timmermans et al., 2021). Wang et al. (2020) assert that when teacher expectations are being adjusted to students' performances, "the bias between teacher expectations and student achievement should become smaller" (p. 2). If teachers tend to adjust their expectations according to students performance, but students perform at a consistent level, the expectations would then remain stable due to students' stable performance (Wang et al., 2020).

The second view of teacher expectation stability is when teachers do not adjust their expectations according to students' performance regardless of student achievement (Brophy, 1983). Thus, teachers keep accurate or inaccurate expectations throughout the academic year and continue to over-and underestimate students throughout the year, optimising the potential for the self-fulfilling prophecy to occur (Brophy, 1983). This study applied the first view of teacher expectation stability, indicating that teachers adjust their expectations towards the end of the year and increase in stability as they align with students' relatively stable performance.

Mixed results regarding teacher expectation stability have been found so far. Studies regarding teacher expectation stability in primary schools found that teachers do not adjust their expectations when teacher expectations were measured over two points. For example, the study of Kuklinski and Weinstein (2000) found that teachers' reading expectations remained relatively stable and showed moderate to strong stability. The study of Martinek (1980) also indicated teacher expectation stability for physical education with correlations between .84 and .96 and that teachers did not adjust their expectations over time. However, it is crucial to note that the studies of Kuklinski and Weinstein (2000) and Martinek (1980) did not take student performance into account, as stated by Wang et al. (2020). Thus, the conclusions regarding teacher expectation stability in the aforementioned studies may have resulted from the fact that the study did not control for students' fluctuating or stable performances throughout the research. Rubie-Davies and colleagues (2018) addressed this limitation and found in their three-year study that, after considering student achievement, class-level teacher expectations remain relatively stable over time on three measurement occasions in an academic year. They also discovered that certain teachers continue to over- or underestimate student achievement for individual students.

On the other hand, Timmermans et al. (2021) found that teachers did not keep their initial expectations as is but adjusted them to correspond with students' mathematics performance throughout the year. The initial adjustments of expectations to match students' performance cause instability at first, but as teachers align their expectations with student achievement, they become more stable over time (Timmermans et al., 2021). The study by Wang et al. (2020) investigated teacher expectation stability within primary schools while controlling for Chinese, mathematics and English achievement and found that teacher expectations mostly showed low stability levels. However, the study by Wang et al. (2020) interpreted stability as continuous

over- or underestimation of student performance. Thus, the findings by Wang et al. (2020) also suggest that teachers adjusted their expectations to match student performance as they learned more about their students' abilities. These findings are not universal, as other studies have also revealed that teacher expectation stability varies between teachers and classrooms (Brophy & Good, 1974; Wang et al., 2020).

Additional studies on teacher expectation stability are essential as they may clarify contradicting findings and shed light on the long-term outcomes of teacher expectations on the self-fulfilling prophecy (Hao, 2022; Timmermans et al., 2021). According to Brophy (1983), these studies could also shed light on the precision of teacher expectations. Jussim and Harber (2005) state that these studies could also help to understand the collection of teacher expectation effects over time. Further, Timmermans et al. (2021) explain that gaining deeper insight into these topics may have practical consequences for overall education and learners' academic performance.

Biased Teacher Expectations

Teacher expectations proved to be susceptible to many teacher biases (Dusek & Joseph, 1983; Hao et al., 2022; Wang et al., 2018, 2019, 2020). Jussim (1986) states that societal stereotypes regarding gender may cause teachers to form biased teacher expectations. However, teacher expectations may often accurately estimate students' academic abilities (Jussim & Harber, 2005). Researchers have deemed teacher expectations a direct result of teacher biases that may influence teachers' judgement of students' ability to perform (Papageorge et al., 2018). Teacher expectation bias can be defined as a bias that "is the difference between the actual perception or expectation and the benchmark" (Ferguson, 2003, p. 466), in other words, the gap between performance and expectation. Examples of the biases that teacher expectations are

susceptible to include gender (Robinson-Cimpian et al., 2014; Wood et al., 2007), ethnicity (Brophy, 1983; Tenenbaum & Ruck, 2007), and socioeconomic status (Rubie-Davies et al., 2018) among others. Siems-Muntoni and Retelsdorf (2023) state that the expectancy-value theory entails that teachers' expectations and stereotypes may contribute significantly to their students' self-esteem, engagement in activities and academic performance across different subjects. Thus, it is clear that teachers must be aware of their stereotypes and biased expectations to benefit students' academic achievement, as they may have long-lasting effects. Research confirms that even minor inaccurate teacher expectations can add up, leading to unwanted student performance disparities (Jussim & Harber, 2005; Rubie-Davies et al., 2014). Therefore, Hao (2022) argues that extended research on teacher expectation bias is necessary to shed light on the impact of self-fulfilling prophecy within classrooms, as the stability of teacher expectations is influenced by teacher biases based on stereotypes formed in society.

Student Gender

A significant predictor of teacher expectations is student gender (Dusek & Joseph, 1985; Jussim et al., 1996). Student gender effects are suspected of impacting classroom practices due to the variations in academic performance between boys and girls (Robinson-Cimpian et al., 2014; Watson et al., 2019; Wood et al., 2007). For example, boys have been performing lower in reading achievement than girls with a substantial effect size ($d = 0.30$), although, for mathematics, it has not been as big ($d = 0.05$; OECD, 2020). Jussim and colleagues (1996) and Thomas (2024) confirm that these gender disparities in performance are not due to ability differences but to teachers' different expectations for boys and girls. Further, Li (1999) states that such differential teacher expectations for boys and girls cause varying treatment of learners and then evoke these gender disparities in student achievement. Differential teacher expectations for

boys and girls have also been found to pass down gender stereotypes in educational settings and cause the issues to prevail (Younger & Warrington, 2008).

Empirical studies have found that teacher expectations tend to be higher for girls' reading achievement (Hinnant et al., 2009; Li, 1999; Meissel et al., 2017; Watson et al., 2019) and higher expectations for boys in mathematics (Jaremus et al., 2020; Thomas, 2024). On the other hand, Hinnant et al. (2009) found that girls' mathematics performance was overestimated even though they expected results to indicate that boys would overestimated for math. Additionally, the recent study of Timmermans and Rubie-Davies (2023) revealed through a study in New Zealand that teacher expectations for girls' mathematics performance were higher than for boys in Grades 6 and 7. These findings may have been due to teachers' awareness of past stereotypes regarding girls' and boys' mathematics achievements, which have increased in New Zealand over recent years (Timmermans & Rubie-Davies, 2023). Wang et al. (2018) found in their review study that much research affirmed the effects of student gender on teacher expectations for reading and mathematics and stated that further research is needed as contradicting studies have also been discovered. Multiple empirical studies found no differences in teacher expectations for boys and girls and argue that gender does not influence the formulation of them (e.g. Auwarter & Aruguete, 2008; de Boer et al., 2010; Sorhagen, 2013; Watson et al., 2016). Thomas (2024) argues that whether teachers' bias regarding student gender is why gender disparities in mathematics and reading performance are found remains inconclusive due to the mixed results so far. Overall, it is clear that student gender has been impacting the way teachers treat their students, resulting in varying performances among boys and girls, and that more research is needed to clarify the mixed results discovered by researchers.

Teacher Gender

Teacher gender has been investigated as an influential factor in education (Krkovic et al., 2014). Hopf and Hatzichristou (1999) found that female and male teachers treat students differently. In addition, female teachers tend to have better interconnections with students, and give students higher grades than male teachers (Hopf & Hatzichristou, 1999). In addition, female teachers are more prominent in the educational field (Watson et al., 2019), and some argue that the lack of male teachers is to blame for boys' overall underachievement (Brown, 2012). In countries such as the U.S.A. (Ehrenberg et al., 1995), Greece (Hopf & Hatzichristou, 1999) and Germany (Neugebauer et al., 2011), students have overall higher achievement in reading in female teachers' classes than in male teachers' classes. Female teachers have also been found to have higher expectations for students in general, while male teachers tend to have lower expectations (Watson et al., 2019). More specifically, female teachers tend to have higher reading expectations for boys and girls than male teachers (Ehrenberg et al., 1995).

The combined effects of teacher and student gender were investigated as a possible influence on teacher expectation stability in this study, as male and female teachers have shown varying interactions with opposite and same-sex students and have different teacher expectations for students (Krkovic et al., 2014; Niemi, 2010; Watson et al., 2019). These findings correspond to the gender schema theory, which posits that students connect better with teachers of the same sex (Bem, 1981). Thus, researchers speculate that the unequal teacher-gender matching ratio influences gender performance disparities over time, as female students perform better in classes with female teachers (Krkovic et al., 2014). An empirical study revealed that teacher-student gender matching levels may positively influence student achievement (Dee, 2007), whereas other studies have found adverse effects (Antecol et al., 2015). For teacher expectations, however, Watson et al. (2016) found no significant effects of teacher-student gender matching levels. This

area of research needs more exploration as mixed results continuously surface regarding the influence of teacher-student gender-matching levels on student achievement (Hwang & Fitzpatrick, 2021). Thomas (2024) confirms that more longitudinal investigations are needed to gain deeper insight into the long-term effects of teacher expectations, with both student and teacher gender as possible moderators of student achievement. The researcher of this study further identified a lack of empirical studies regarding the influence of teacher and student gender and teacher-student gender matching on teacher expectation stability.

The Current Study

This study investigated teacher expectation stability to determine whether teachers adjusted their expectations according to students' performance and whether expectations and achievement correlated more over one academic year. In addition, this study researched if teacher and student gender were moderators of teacher expectation stability for students' reading and mathematics achievement. The following research questions were posed for this study:

1. To what extent are teacher expectations stable for individual students' academic performance in mathematics and reading over one academic year?
2. How does teacher expectation stability differ among male and female teachers regarding individual students' mathematics and reading performance over one academic year?
3. How does teacher expectation stability regarding individual students' mathematics and reading performance over one academic year differ between female and male students?

4. Is there a significant difference in teacher expectation stability regarding individual students' mathematics and reading performance over one academic year between male and female teachers for male and female students?

Methods

The data used for the current study was drawn from a three-year longitudinal study by Rubie-Davies et al. (2018). For this study, only the first year of the collected data was used for the analyses.

Participants

The sampling procedures described in the original study by Rubie-Davies et al. (2018) included separating the schools in one geographical region into high, middle, and low socioeconomic levels. Schools from all socioeconomic levels were asked to join the study by random selection, ensuring a good representation of every socioeconomic level within the sample. Teachers from the schools then voluntarily participated in the study. The sample participants were 89 teachers in 12 New Zealand primary and intermediate public schools. Of the teachers, 71.9% were female, and of the students, 49% were female. Teachers participated in the study voluntarily. Student participants from the primary schools were 8-10 years old, and intermediate school students were between 11 and 12 years old ($N = 2536$). The schools that participated were from various socioeconomic areas ranging from low to high based on a ten-point scale (1 = low, and 10 = high). Student ethnicities varied, including Asian, Pacific Island, Māori, and New Zealand/European.

Instruments

Teacher expectations

Teachers had access to the information regarding students' academic achievement in the previous grade but had not tested the students themselves. Teachers were asked to use the Likert scale (Norman, 2010) to predict what they believed represented the students' abilities to perform academically by the end of the year. The scale ranged from 1 to 7, with 1 being *very much below average* and 7 being *very much above average*. To measure the students' mathematics and reading achievements on the same three occasions as when teachers had to formulate their expectations, a standardised tool known as the Electronic Assessment Tools for Teaching and Learning (e-asTTle) was used (Rubie-Davies et al., 2018). The tests had a forty-minute duration, and students had to take two tests at each measurement occasion, one for reading and one for mathematics. The participant teachers had the choice of what level of test the students should take.

Student Achievement

Different levels of reading and mathematics tests were possible to administer within one cohort of students as the curriculum used in New Zealand entails that each level requires two years to complete. For example, Rubie-Davies et al. (2018) describe that nine-year-old students would be at the start of Level 3 in New Zealand's national curriculum. Therefore, teachers could decide to administer a level 2 or 3 test, depending on the student and their judgement. All the test results from the e-asTTIe were then adjusted using the item response theory (IRT) that equalises measurements according to Embretson and Reise (2000). These procedures allowed all the participant students to achieve similar scores despite the test level they took, as the scores were comparable.

According to Rubie-Davies et al. (2018), scores had varying significance for students from different grade levels. An example given by this study includes that a score of 1460 could

mean that a Year 4 student performs at the mean level of their cohort. The same score for a Year 8 student would indicate that the student achieved considerably lower than the average of that grade level. Therefore, the standardised scores of all students were subtracted from the national average for student grades and time of the year. These procedures were done so that the scores could be comparable in a purposeful way (Rubie-Davies et al., 2018). The curriculum content tested in the reading tests was "processes and strategies, ideas, and language features" (Rubie-Davie et al., 2018, p. 228). The content tested in the mathematics tests was "number knowledge, number sense, and algebra" (Timmermans et al., 2021, p. 5).

Student and Teacher Gender

The gender variables of this study were coded for male teachers as 'Male' = 1. For female teachers, it was coded as 'Female' = 2. The variable was coded as 'Boy' = 1 for male students. For female students, the variable was coded as 'Girl' = 2.

Procedures

Ethics were considered by obtaining approval from the original study's 'institutional committee' (Rubie-Davies et al., 2018, p. 229). Teachers were surveyed three times during one year to record student expectations in mathematics and reading. The first survey was issued in February, the first month of the new school year. Another was administered in June of the first year of this study and at the end of the year in October. Simultaneously, students completed the e-asTTIe tests in February, June (only in the first year), and October. The tests were created by the second author of the original study. The tests were couriered to teachers, and after students had taken them, they were couriered back to the researchers for assessment (Rubie-Davies et al., 2018).

Data Analyses

The data was analysed by using SPSS 28.0 (Garson, 2020). The data's normality was checked to ensure the regression analyses' validity. The normality of the data was assumed due to the central limit theorem, as the number of cases in the data set exceeded the minimum number to be considered normal (Wilcox, 2009). There were missing values for the students' scores for both reading and math across the academic year. The missing values for student scores were coded, with the present values = 0 and the missing = 1 and were missing at random. After observing the loss of many cases when excluding the missing values and the significant reduction of the sample size, it was decided to continue without filtering out the missing values. The outliers present did not skew the data and were thus also not filtered out (Wilcox, 2009). The teacher sample was reduced to 83 teachers, as this study only needed the teachers who participated in phase 1, meaning in the first year of this study. The relevant variables in relation to the research questions are discussed below.

Individual-Student-level Teacher Expectations

The methods applied in the study by Wang et al. (2020) were used as guidance to answer Research Question 1. First, student achievements in mathematics and reading were rescaled to make the scores comparable by standardising student achievement scores. Then, to examine whether students were over- or underestimated by teachers' expectations and expose possible expectation bias, the unstandardised scores were regressed separately for mathematics and reading. These linear regressions created residuals of teacher expectations and students' mathematics and reading achievements on the three measurement occasions. Thus, each student had three sets of residuals for both their mathematics and reading expectations. The greater the residuals differ from zero, the more inaccurate teacher expectations were. Positive residuals indicated that students were overestimated, while negative residuals indicated that

underestimation occurred regarding students' actual performances. Residuals from each regression model were stored. Every time point that teachers had to document their expectations and students had to take tests was named a ‘measurement occasion’ throughout this study. As there were three such measurement occasions throughout the year, the measurement occasions were called ‘measurement occasion 1’, ‘measurement occasion 2’, and ‘measurement occasion 3’, respectively, throughout this study.

Next, Pearson’s correlation coefficient analysis was used to determine the overall association between the teacher expectations for students’ reading and mathematics achievements over the three measurement occasions over one year. The general strength and direction of the correlations were first interpreted based on Pearson’s Correlation Coefficient scale, which indicates that correlation effects are small ($r \approx .10$), medium ($r \approx .30$), and large ($r \approx .50$) (Cohen, 1977). When the correlation coefficient is a positive number, the relationship is positive. Thus, as the independent variable increases, the dependent variable also increases, and the same counts for a negative correlation (Cohen, 1977).

In terms of teacher expectation stability, however, the interpretation of the correlations was based on the study by Wang et al. (2020). These correlation interpretations differ from the general Pearson’s correlation coefficient scale and would thus indicate the extent of teacher expectation stability between two measurement occasions, specifically. These were specified as a negative group ($r < -0.3$), low stability group ($0.3 > r \geq -0.3$), some stability group ($0.5 > r \geq 0.3$), moderate stability group ($0.7 > r \geq 0.5$), and strong stability group ($r \geq 0.70$). The closer the correlation result between two measurement occasions is to one, the more stable the reading or mathematics teacher expectations are between measurement occasions (Norman, 2010; Wang et al., 2020). Correlations were calculated for reading and mathematics achievement residuals of

measurement occasions 1 and 2, 2 and 3, and 1 and 3. Next, the Fisher's Z test was used to compare the correlations between male and female teachers and students through an online calculator and indicated whether the correlations were significantly different from each other (Lenhard & Lenhard, 2014). The significant level for the test was 0.05.

Individual Student-level Teacher Expectation Stability Across Genders

To answer question 2, the data file was split by teacher gender to determine if there was a difference in teacher expectation stability for mathematics and reading achievement between male and female teachers over the three measurement points. To answer question 3, the file was split by student gender to view whether there were differences between the teacher expectation stability of boys' and girls' reading and math achievement. To answer question 4, the data file was split by teacher and student gender simultaneously, and Pearson correlations were executed. The correlations were then compared using the Fisher's Z test to determine whether correlations were significantly different across different measurement occasions and genders.

Results

Teacher Expectation Stability for Reading and Mathematics Achievement (Research Question 1)

Regression Analyses

The simple linear regression models revealed medium to large effect sizes and that students' reading achievement was positively correlated with teacher expectations. Students' reading achievement significantly predicted the teachers' expectations at measurement occasion 1 ($\beta = 0.428$, $t(1829) = 20.25$, $p < .001$, at measurement occasion 2 ($\beta = 0.492$, $t(1785) = 23.88$, $p < .001$, and at measurement occasion 3 ($\beta = 0.525$, $t(1541) = 24.21$, $p < .001$. Further, the relationship between teacher expectations and students' reading achievement strengthened

towards the end of the year. This indicated that the expectations for reading achievement became more correlated with student reading achievement throughout the year. For mathematics, an overall positive relationship between students' mathematics achievement and teacher expectations was found with medium to large correlation coefficients. However, the strength of the positive relationship decreased between the second and third measurement occasions, and only a small effect size was seen during the third measurement occasion. This indicated that teacher expectations and the students' results became less correlated by the end of the year for mathematics. The results can be viewed in Table 1.

Table 1

Regression Coefficients for Reading and Mathematics

	Measurement Occasion	*β*	*t*	*SE*	*R*	R^2	*F*	*df1*	*df2*	*p*
Reading	1	.428	20.246	1.296	.428	.183	409.892	1	1829	<.001
	2	.492	23.878	1.305	.492	.242	570.151	1	1785	<.001
	3	.525	24.206	1.253	.525	.275	585.930	1	1541	<.001
Math	1	.459	22.519	1.214	.459	.210	507.113	1	1905	<.001
	2	.508	25.605	1.223	.508	.258	655.611	1	1884	<.001
	3	.303	12.429	1.311	.303	.092	154.491	1	1531	<.001

Note. Dependent variable = Teacher expectations; Independent variable = Students' normed scores.

Unstandardised Residuals Correlations

Table 2 presents the correlation coefficients that were calculated to determine whether the unstandardised residuals, as derived from the regression models presented in Table 1, showed stability over the academic year. The correlations of students' reading and mathematics achievement and teacher expectations were large and positively correlated, showing linear relationships. This indicated that students who were either under- or overestimated at one measurement occasion were also likely to be under- or overestimated at the following measurement occasion. In addition, all correlations were significant, indicating that teacher expectations for reading and mathematics were substantially stable over the academic year.

When interpreting the teacher expectation stability using the guidelines of Wang et al. (2020), the correlation between the reading residuals of measurement occasions 1 and 2 indicated moderate stability; $r(1434) = .667$, $p < .001$. The correlation between the residuals of measurement occasions 1 and 3 indicated strong stability; $r(1241) = .700$, $p < .001$. The correlation between the residuals of measurement occasions 2 and 3 indicated strong stability; $r(1278) = .920$, $p < .001$. Overall, all the correlations increased in stability towards the end of the academic year for reading and math achievement, and the results can be viewed in Table 2.

The Fisher's Z test revealed that the reading correlations for measurement occasions 1 and 2 were significantly different and lower than for measurement occasions 2 and 3 ($z = -20.316$, $p < .001$). For mathematics, the correlations for measurement occasions 1 and 2 were significantly different and lower than for measurement occasions 2 and 3 ($z = -13.369$, $p < .001$). The results are shown in Table 2. Scatterplot matrices were created for both reading (Figure 1) and mathematics (Figure 2) unstandardised residuals to display the positive linear relationships

between student achievement and teacher expectations and the increase in strength towards the end of the year (See Appendix 1 and 2). Overall, it can be seen here that reading residual correlations were more robust than mathematics residual correlations throughout the year. Note that all tables from here onwards contain the results of two significance tests that were executed: Pearson's correlations coefficients test and Fisher's Z test.

Table 2

Unstandardised Residuals Correlations of Teacher Expectations for Students' Reading and Mathematics Achievement and *Comparisons of Correlations*

Reading			*1,2*		*1,3*	
Measurement Occasion	*r*	*n*	*z*	*p*	*z*	*p*
1, 2	.667**	1434				
1, 3	.700**	1241	-1.594	0.055		
2, 3	.920**	1278	-20.316	0.001	-18.088	0.001
Maths			*1,2*		*1,3*	
Measurement Occasion	*r*	*n*	*z*	*p*	*z*	*p*
1, 2	.560**	1597				
1, 3	.643**	1279	-3.472	0.001		
2, 3	.810**	1356	-13.369	0.001	-9.322	0.001

Note. ** $p < .001$.

Teacher Expectation Stability of Male and Female Teachers for Reading and Mathematics Achievement (Research Question 2)

From Table 3, it was seen that all correlations had positive large effect sizes, and were significant for both reading and mathematics for male and female teachers. In terms of teacher expectation stability, male teachers' reading expectations showed moderate stability for the correlation between the residuals of measurement occasions 1 and 2; $r(401) = .613$, $p < .001$ and 1 and 3; $r(410) = .633$, $p < .001$. The highest stability for male teachers was seen towards the end of the year, where the correlation between the residuals of measurement occasions 2 and 3 showed strong stability; $r(430) = .918$, $p < .001$.

The Fisher's Z test revealed that the reading achievement correlation of female teachers was significantly stronger than for males at measurement occasions 1 and 2 ($z = -1.768$, $p = .039$) and at measurement occasions 1 and 3 ($z = -2.527$, $p = .006$). For mathematics, the correlation of male teachers was significantly weaker than the correlation of female teachers at measurement occasions 1 and 2 ($z = -2.277$, $p = .011$).

Table 3

Unstandardised Residuals Correlations of Male and Female Teacher Expectations for Students' Reading and Mathematics Achievement and *Comparison of Correlations*

Reading			
	Male Teacher	Female Teacher	Fisher's Z

Measurement Occasion	*r*	*n*	*r*	*n*	*z*	*p*
1, 2	.613**	401	.674**	1033	-1.768	0.039
1, 3	.633**	410	.716**	831	-2.527	0.006
2, 3	.918**	430	.914**	848	0.419	0.338
Maths						
	Male Teacher		Female Teacher		Fisher's Z	
Measurement Occasion	*r*	*n*	*r*	*n*	*z*	*p*
1, 2	.487**	446	.578**	1151	-2.277	0.011
1, 3	.608**	396	.640**	883	-0.864	0.194
2, 3	.828**	440	.799**	916	1.477	0.07

Note. ** $p < .001$.

Teacher Expectation Stability of Boys and Girls for Reading and Mathematics Achievement (Research Question 3)

As seen in Table 4, the correlations had large and positive effect sizes, showing stability over the academic year. In terms of stability, the correlation between the residuals of the boys' reading showed moderate stability during measurement occasions 1 and 2; $r(692) = .672$, $p < .001$, and strong stability for the remainder of the correlations between the residuals. For mathematics, boys showed moderate stability for measurement occasions 1 and 2; $r(793) = .569$,

$p < .001$ and 1 and 3; $r(629) = .642$, $p < .001$. Strong stability was present at measurement occasions 2 and 3; $r(672) = .810$, $p < .001$. The Fisher's Z tests indicate no significant differences in teacher expectation stability for reading and mathematics achievement of boys and girls. Table 4 shows the results.

Table 4

Unstandardised Residuals Correlations of Teacher Expectations for Boys' and Girls' Reading and Mathematics Achievement and *Comparisons of Correlations*

Reading						
	Boys		Girls		Fisher's Z	
Measurement Occasion	*r*	*n*	*r*	*n*	*z*	*p*
1, 2	.672**	692	.661**	742	0.374	0.354
1, 3	.702**	595	.697**	646	0.172	0.432
2, 3	.921**	624	.920**	654	0.117	0.454
Maths						
	Boys		Girls		Fisher's Z	
Measurement Occasion	*r*	*n*	*r*	*n*	*z*	*p*
1, 2	.569**	793	.549**	804	0.58	0.281
1, 3	.642**	629	.645**	650	-0.091	0.464

2, 3	.810**	672	.812**	684	-0.107	0.457

Note. ** $p < .001$.

Teacher Expectation Stability of Male and Female Teachers for Boys' and Girls' Reading and Mathematics Achievement (Research Question 4)

As seen in Table 5, the correlations were all large in effect size and positive, indicating a strong linear relationship between students' mathematics and reading achievement and teacher expectations, regardless of the gender of teachers and students. Teacher expectation stability of male and female teachers for boys' and girls' reading achievement increased towards the end of the year, with strong stability present between the residuals at measurement occasions 2 and 3. The Fisher's Z test results showed no significant differences in male and female teacher expectation stability between boys and girls for reading achievement for all measurement occasions. Results can be viewed in Table 5. For mathematics, results can be viewed in Table 6.

Table 5

Unstandardised Residuals Correlations of Male and Female Teacher Expectations for Boys' and Girls' Reading Achievement and *Comparisons of Correlations*

Reading						
Male Teachers						
	Boys		Girls		Fisher's Z	
Measurement	*r*	*n*	*r*	*n*	*z*	*p*

Occasion						
1, 2	.604**	196	.618**	205	-0.222	0.412
1, 3	.628**	202	.635**	208	-0.117	0.453
2, 3	.913**	210	.922**	220	-0.586	0.279
Female Teachers						
	Boys		Girls		Fisher's Z	
Measurement Occasion	*r*	*n*	*r*	*n*	*z*	*p*
1, 2	.683**	496	.666**	537	0.499	0.309
1, 3	.721**	393	.714**	438	0.207	0.418
2, 3	.917**	414	.911**	434	-0.529	0.298

Note. ** $p < .001$.

Table 6

Unstandardised Residuals Correlations of Male and Female Teacher Expectations for Boys' and Girls' Mathematics Achievement and *Comparisons of Correlations*

Maths
Male Teachers

	Boys		Girls		Fisher's Z	
Measurement Occasion	*r*	*n*	*r*	*n*	*z*	*p*
1, 2	.478**	231	.503**	215	0.345	0.365
1, 3	.620**	199	.598**	197	0.345	0.365
2, 3	.830**	224	.827**	216	0.1	0.46
Female Teachers						
	Boys		Girls		Fisher's Z	
Measurement Occasion	*r*	*n*	*r*	*n*	*z*	*p*
1, 2	.595**	562	.560**	589	0.889	0.187
1, 3	.637**	430	.645**	453	-0.201	0.42
2, 3	.799**	448	.800**	468	-0.042	0.483

Note. ** $p < .001$.

Discussion

Main Findings

This study investigated whether teacher expectation stability occurred over one academic year and if teacher and student gender acted as a moderator of the stability. It was discovered that

all correlations between the residuals of student achievement and teacher expectations showed positive and significant relationships for both reading and mathematics achievement throughout the academic year. Thus, there was overall teacher expectation stability throughout the academic year, although the levels of stability differed in strength on different measurement occasions. The teacher expectation stability mainly varied throughout the year from moderate to strong stability, answering the first research question. These findings indicated minimal variation between students' performance and teacher expectations. Teachers' accuracy with their expectations also revealed that there was minimal bias present, and if any, biased teacher expectations decreased towards the end of the year. The abovementioned findings are consistent with prior teacher expectation stability research from Kuklinski and Weinstein (2000) and Martinek (1980), who found that teachers' expectations were stable. Additionally, as this study controlled for student achievement, it was also consistent with the findings of Timmermans et al. (2021) and Wang et al. (2020), as teacher expectations were adjusted over time to align with student performance, and became more stable towards the end of the year due to students' consistent performance. A final observation was that the unstandardised residual correlations were stronger for reading achievement than mathematics. This may indicate that teachers were more accurate in their reading expectations and that more biased teacher expectations may be present in mathematics.

After investigating whether there were teacher gender differences in teacher expectation stability, it was concluded that teacher expectation stability increased throughout the academic year for male and female teachers. Thus, both male and female teachers adjusted their expectations throughout the year to match students' performance regardless of teacher gender. Interestingly, some significant differences were discovered between male and female reading residual correlations. However, only one significant difference was found for mathematics

between male and female teachers' expectation stability at the beginning of the year. These findings indicated more prominent differences in male and female teacher expectation stability in reading achievement than in mathematics. These findings were consistent with research stating that female teachers have higher expectations for students' reading achievement than male teachers have (Ehrenberg et al., 1995).

Further, female teachers had the highest correlations at all three of the significant differences, which leads to the conclusion that female teachers were significantly more accurate in their expectations at the beginning of the year than male teachers were. These findings relate to prior research findings stating that female teachers have overall higher expectations for students regardless of the subject, whereas male teachers tend to have lower (Watson et al., 2019). Overall, teacher expectation stability differed significantly for male and female teachers during the first half of the academic year for both reading and mathematics, but not the last part, answering the second research question. Thus, it was concluded that teacher gender may affect the expectations teachers have for their students and may still contribute to gender disparities among students. However, as the teacher gender differences found were not constant throughout the year, it indicated that teacher gender effects on teacher expectation stability may fade over time.

Regarding the third research question, no significant differences were found in teacher expectation stability for boys' and girls' reading and mathematics achievement throughout the year. Therefore, it was concluded that student gender did not moderate teacher expectation stability. Based on this conclusion, it was then assumed that student gender bias did not impact teacher expectations and was consistent with prior research (e.g. Auwarter & Aruguete, 2008; de Boer et al., 2010; Sorhagen, 2013; Watson et al., 2016). The fact that it seemed like student

gender did not cause any differences in teacher expectations is, however, contradicting some studies stating that there are significant differences between teacher expectations for boys' and girls' academic performance (e.g. Hinnant et al., 2009; Jaremus et al., 2020; Li, 1999; Meissel et al., 2017; Thomas, 2024; Watson et al., 2019) and that student gender is a predictor of teacher expectations (Jussim et al., 1996). Teacher expectation stability also increased towards the end of the year for both boys and girls and thus indicated that the more teachers became familiar with boys' and girls' true academic abilities, the more their expectations corresponded with students' achievement.

The fourth and final research question was answered, as there were no significant differences between male and female teachers' expectation stability for boys and girls. Thus, it can be said that teacher-student gender-matching levels did not impact expectation stability for reading or mathematics achievement over one academic year. These findings were consistent with those of Watson et al. (2016).

Strengths and Limitations

A strength of this study is that the sample data accurately represents the general population of New Zealand, as it includes many ethnicities, genders and schools from different socioeconomic areas. Another strength is that this study used three measurement occasions throughout the academic year and controlled for student achievement to analyse the teacher expectation stability, whereas more dated research used only two time points and did not control for student achievement (e.g. Kuklinkski & Weisntein, 2000; Martinek, 1980). Thus, more accurate results and conclusions may have been drawn from this study and contribute to teacher expectation stability research.

This study's main limitation is that many missing values were detected, although the cases with missing values were not filtered out to prevent data loss. More accurate results and conclusions would have been possible with a more complete data set. Another limitation would be that the findings might not be generalisable in other parts of the world. The data was collected in primary education and may not be generalisable in different educational sectors such as secondary or tertiary institutions. Another limitation to the generalisability of the results is that the teachers in New Zealand might be more aware of the possible biases they may have due to this long-term study conducted there, as stated by Timmermans and Rubie-Davies (2023). The testing effects may have also impacted the results and caused teacher participants to be more aware of their expectations in general. Thus, teachers in other countries where fewer empirical studies have been conducted regarding this research domain may be far less aware of their biased teacher expectations regarding gender. The unawareness of teachers may prolong the contribution of biased teacher expectations to the gender disparities in student achievement and society. A final limitation is that during the analysis procedures, only quantitative methods were used, and thus, there may have been valuable qualitative data that could have strengthened the findings.

A recommendation for future research is to conduct longitudinal studies investigating the influences of student and teacher gender on teacher expectation stability and controlling for student achievement in all subjects, as this study only controlled for two subjects. Additional longitudinal studies regarding teacher expectation stability are recommended to be done in countries other than New Zealand so that different populations of teachers and students can be involved and results can be generalised. Further, data collection methods should also be improved in future studies to ensure missing data is minimised by teachers and students. Finally,

mixed methods should be used to combine quantitative data with qualitative data to get a deeper insight into the true biases of teachers that may influence their expectations and the stability thereof. For example, observations could be used to determine why teachers have certain expectations for students. It could also be helpful to gather teachers' thoughts on whether they have gender biases, even though the statistical analyses suggested that they do not.

Practical Implications

Practical implications for education would be that school leadership teams should ensure that high teacher expectations are present in schools. Workshops can help teachers learn how to maintain high expectations throughout the year, as research shows that inaccurate expectations can accumulate over time and affect student outcomes. Even though such initiatives may positively affect student performance, Timmermans et al. (2021) state that it is not yet clear how big these effects would be. Further, school leaders must be aware that teacher gender may have long-term effects on student achievement. As research proved that female teachers tend to have higher expectations for boys and girls in reading and mathematics, students of male teachers may suffer lower teacher expectations and may be affected by this phenomenon. Male teachers need to be made aware of their own possible biases and that they tend to underestimate students' performances. Further, due to the high number of female teachers in the education industry, schools must pay attention to retaining male teachers to increase the balance in the educational field. With enhanced teacher education regarding biased expectations and decreasing the male and female teacher imbalance, gender disparities among students may be lessened and positively impact equality in education and society.

Conclusions

Overall, no significant differences were found between the stability of teacher expectations of male and female teachers for boys and girls when all correlations were compared simultaneously. However, due to significantly lower correlations found for male teachers than female teachers at the beginning of the year for reading and mathematics, it was concluded that some teacher gender influences may still impact teacher expectation stability. Therefore, this study recommended that future research investigate teacher gender's influence on teacher expectations and the stability of student achievement across subjects. Additional research on teacher expectation stability and student and teacher gender as possible moderators of teacher expectation stability is needed as empirical studies in this domain remain scarce.

References

Antecol, H., Eren, O., & Ozbeklik, S. (2015). The Effect of Teacher Gender on Student Achievement in Primary School. *Journal of Labor Economics*, *33*(1), 63–89. https://doi.org/10.1086/677391

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. https://doi.org/10.3200/JOER.101.4.243-246

Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review*, *88*(4), 354–364. https://doi.org/10.1037/0033-295X.88.4.354

Brophy, J. E., & Good, T. L. (1970). Teachers' communication of differential expectations for children's classroom performance: Sombehaviouralal data. *Journal of Educational Psychology*, *61*(5), 365–374. https://doi.org/10.1037/h0029908

Brophy, J. E. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology*, *75*(5), 631–661. https://doi.org/10.1037/0022-0663.75.5.631

Brown, A. L. (2012). On Human Kinds and Role Models: A Critical Discussion about the African American Male Teacher. *Educational Studies*, *48*(3), 296–315. https://doi.org/10.1080/00131946.2012.660666

Cohen, J. (1977). Statistical power analysis for the behavioral sciences (Rev. ed). Academic Press.

Dee, T. S. (2007). Teachers and the Gender Gaps in Student Achievement. *Journal of Human Resources*, *XLII*(3), 528–554. https://doi.org/10.3368/jhr.XLII.3.528

De Boer, H., Bosker, R. J., & Van Der Werf, M. P. C. (2010). Sustainability of teacher expectation bias effects on long-term student performance. *Journal of Educational Psychology*, *102*(1), 168–179. https://doi.org/10.1037/a0017289

Dusek, J. B., & Joseph, G. (1983). The bases of teacher expectancies: A meta-analysis. *Journal of Educational Psychology*, *75*(3), 327–346. https://doi.org/10.1037/0022-0663.75.3.327

Dusek, J. B., & Joseph, B. (1985). The basis of teacher expectancies. In J. B. Dusek (Ed.), *Teacher expectancies* (229–250). Lawrence Erlbaum Associates.

Ehrenberg, R., Goldhaber, D., & Brewer, D. (1994). *Do Teachers' Race, Gender, and Ethnicity Matter?: Evidence from NELS88* (w4669; p. w4669). National Bureau of Economic Research. https://doi.org/10.3386/w4669

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. L. Erlbaum Associates.

Ferguson, R. F. (2003). Teachers' Perceptions and Expectations and the Black-White Test Score Gap. *Urban Education*, *38*(4), 460–507. https://doi.org/10.1177/0042085903038004006

Garson, G. D. (2020). *Multilevel modeling: Applications in STATA, IBM SPSS, SAS, R, & HLM*. SAGE Publications, Inc.

Good, T. L. (1987). Two Decades of Research on Teacher Expectations: Findings and Future Directions. *Journal of Teacher Education*, *38*(4), 32–47. https://doi.org/10.1177/002248718703800406

Good, T. L. & Brophy, J. E. (1997). *Looking in classrooms* (7th ed.). Longman.

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*, *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. https://doi.org/10.1037/a0014306

Hopf, D., & Hatzichristou, C. (1999). Teacher gender‐related influences in Greek schools. *British Journal of Educational Psychology*, *69*(1), 1–18. https://doi.org/10.1348/000709999157527

Hwang, N., & Fitzpatrick, B. (2021). Student–Teacher Gender Matching and Academic Achievement. *AERA Open*, *7*, 233285842110400. https://doi.org/10.1177/23328584211040058

Jaremus, F., Gore, J., Prieto-Rodriguez, E., & Fray, L. (2020). Girls are still being 'counted out': Teacher expectations of high-level mathematics students. *Educational Studies in Mathematics*, *105*(2), 219–236. https://doi.org/10.1007/s10649-020-09986-9

Jussim, L. (1986). Self-fulfilling prophecies: A theoretical and integrative review. *Psychological Review*, *93*(4), 429–445. https://doi.org/10.1037/0033-295X.93.4.429

Jussim, L., Eccles, J., & Madon, S. (1996). Social Perception, Social Stereotypes, and Teacher Expectations: Accuracy and the Quest for the Powerful Self-Fulfilling Prophecy. In *Advances in Experimental Social Psychology* (Vol. 28, pp. 281–388). Elsevier. https://doi.org/10.1016/S0065-2601(08)60240-3

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*(2), 131–155. https://doi.org/10.1207/s15327957pspr0902_3

Krkovic, K., Greiff, S., Kupiainen, S., Vainikainen, M.-P., & Hautamäki, J. (2014). Teacher evaluation of student ability: What roles do teacher gender, student gender, and their interaction play? *Educational Research*, *56*(2), 244–257.

https://doi.org/10.1080/00131881.2014.898909

Kuklinski, M., & Weinstein, R. (2000). Classroom and Grade Level Differences in the Stability of Teacher Expectations and Perceived Differential Teacher Treatment. *Learning Environments Research*, *3*(1), 1–34. https://doi.org/10.1023/A:1009904718353

Lenhard, W., & Lenhard, A. (2014). Testing the Significance of Correlations. Unpublished. https://doi.org/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

Martinek, T. J. (1980). Stability of Teachers' Expectations for Elementary School Aged Children. *Perceptual and Motor Skills*, *51*(3_suppl2), 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

Merton, R. K. (1948). The Self-Fulfilling Prophecy. *The Antioch Review*, *8*(2), 193. https://doi.org/10.2307/4609267

Neugebauer, M., Helbig, M., & Landmann, A. (2011). Unmasking the Myth of the Same-Sex Teacher Advantage. *European Sociological Review*, *27*(5), 669–689. https://doi.org/10.1093/esr/jcq038

Niemi, N. S. (2010). Still failing at fairness: How gender bias cheats girls and boys in school and what we can do about it, by David Sadker, Myra Sadker and Karen Zittleman. *Gender and Education*, *22*(1), 142–143. https://doi.org/10.1080/09540250903464773

Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education*, *15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

Organisation for Economic Co-operation and Development (OECD). (2020). PISA 2018 results (Volume II): Where all students can succeed. OECD Publishing.

Papageorge, N., Gershenson, S., & Kang, K. M. (2018). *Teacher Expectations Matter* (w25255; p. w25255). National Bureau of Economic Research. https://doi.org/10.3386/w25255

Robinson-Cimpian, J. P., Lubienski, S. T., Ganley, C. M., & Copur-Gencturk, Y. (2014). Teachers’ perceptions of students’ mathematics proficiency may exacerbate early gender gaps in achievement. *Developmental Psychology*, *50*(4), 1262–1281. https://doi.org/10.1037/a0035073

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, *3*(1), 16–20. https://doi.org/10.1007/BF02322211

Rubie‐Davies, C. M. (2007). Classroom interactions: Exploring the practices of high‐ and low‐expectation teachers. *British Journal of Educational Psychology*, *77*(2), 289–306. https://doi.org/10.1348/000709906X101601

Rubie-Davies, C. M. (2010). Teacher expectations and perceptions of student attributes: Is there a relationship? *British Journal of Educational Psychology*, *80*(1), 121–135. https://doi.org/10.1348/000709909X466334

Rubie-Davies, C. M. (2023). Teacher Expectations. In P. W. S. J. Watson, C. M. Rubie-Davies, & B. Ertl, *The Routledge International Handbook of Gender Beliefs, Stereotype Threat, and Teacher Expectations* (1st ed., pp. 179–190). Routledge. https://doi.org/10.4324/9781003275763-19

Rubie-Davies, C. M., Weinstein, R. S., Huang, F. L., Gregory, A., Cowan, P. A., & Cowan, C. P.

(2014). Successive teacher expectation effects across the early school years. *Journal of Applied Developmental Psychology*, *35*(3), 181–191. https://doi.org/10.1016/j.appdev.2014.03.006

Rubie-Davies, C. M., Watson, P. W. S. J., 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

Siems-Muntoni, F., & Retelsdorf, J. (2023). Teachers' Gender-Based Expectations and Stereotypes for Student Reading. In P. W. S. J. Watson, C. M. Rubie-Davies, & B. Ertl, The Routledge International Handbook of Gender Beliefs, Stereotype Threat, and Teacher Expectations (1st ed., pp. 191–203). Routledge. https://doi.org/10.4324/9781003275763-20

Sorhagen, N. S. (2013). Early teacher expectations disproportionately affect poor children's high school performance. *Journal of Educational Psychology*, *105*(2), 465–477. https://doi.org/10.1037/a0031754

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

Thomas, H. (2024). *Sex Differences in Reading and Math Test Scores of Children: A Heterodoxical Model*. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-41272-1

Timmermans, A. C., & Rubie-Davies, C. M. (2023). 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*, *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*, 101483. https://doi.org/10.1016/j.learninstruc.2021.101483

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. (2019). Instructional practices and classroom interactions of high and low expectation teachers in China. *Social Psychology of Education*, *22*(4), 841–866. https://doi.org/10.1007/s11218-019-09507-4

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*, 101819. https://doi.org/10.1016/j.lindif.2019.101819

Watson, P. W., C. M. Rubie-Davies, K. Meissel, E. R. Peterson, A. Flint, L. Garrett, and L. McDonald. (2016). Gendered Teacher Expectations of Mathematics Achievement in New Zealand: Contributing to a Kink at the Base of the STEM Pipeline? *International Journal of Gender, Science and Technology*, *8*(1), 83–102.

Watson, P. W. St. 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), 1000–1019. https://doi.org/10.1080/09540253.2017.1410108

Wilcox, R. R. (2009). Basic statistics: Understanding conventional methods and modern insights. Oxford University Press.

Wood, D., Kaplan, R., & McLoyd, V. C. (2007). Gender Differences in the Educational Expectations of Urban, Low-Income African American Youth: The Role of Parents and the School. *Journal of Youth and Adolescence*, *36*(4), 417–427. https://doi.org/10.1007/s10964-007-9186-2

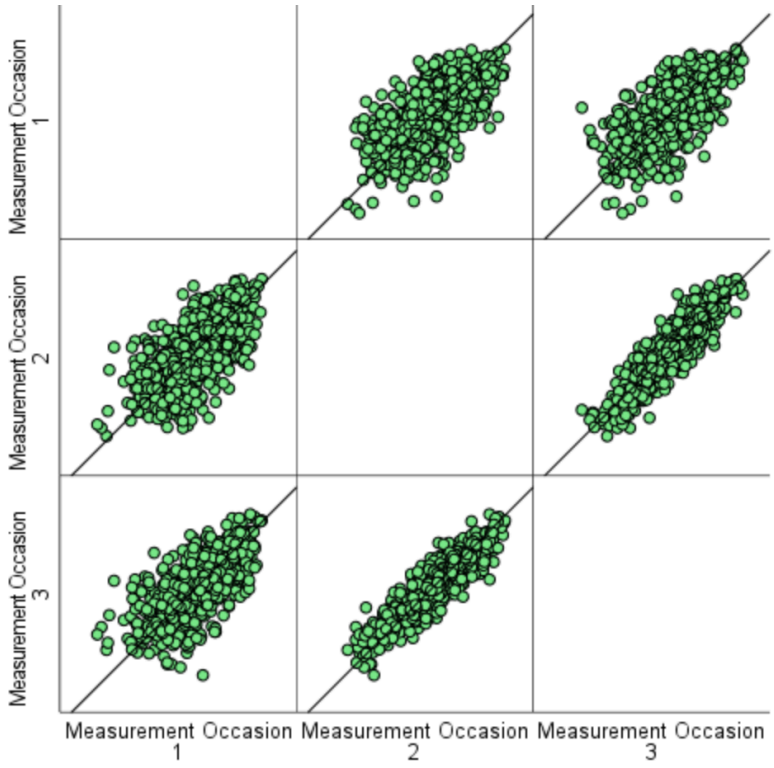
Younger, M., & Warrington, M. (2008). The gender agenda in primary teacher education in England: Fifteen lost years? *Journal of Education Policy*, *23*(4), 429–445. https://doi.org/10.1080/02680930802054396

Appendices

Appendix 1

Figure 1

Scatterplot Matrix of Unstandardised Residuals for Reading



Appendix 2

Figure 2

Scatterplot Matrix of Unstandardised Residuals for Mathematics