

Viewing students consistently: How stable are teachers' expectations?

Christine M. Rubie-Davies, Penelope W. St J. Watson, Annaline Flint, Lynda Garrett, and
Lyn McDonald

Faculty of Education and Social Work, The University of Auckland, 74 Epsom Ave,
Auckland, 1023, New Zealand

Phone: +64 9 373 7599

Corresponding author: Christine Rubie-Davies, Private Bag 92601, Faculty of Education and
Social Work, The University of Auckland, New Zealand; c.rubie@auckland.ac.nz +649 932
9274

Acknowledgement: This study was supported by the Marsden Fund Council from
Government funding administered by the Royal Society of New Zealand.

Christine Rubie-Davies is a Professor of Education, Faculty of Education and Social Work,
University of Auckland, New Zealand. Her primary research interests are teacher
expectations and beliefs that moderate expectancy effects. Christine primarily focuses on
teacher expectation effects for disadvantaged groups. She has published six books and
numerous articles and chapters in prestigious presses, has won national and international
awards for research, teaching and service, and is a Fellow of the Association for
Psychological Science.

Penelope Watson is a Lecturer at the University of Auckland's Faculty of Education and
Social Work. As well as maintaining a strong research focus on teacher expectations,

her research interests include gender, gender stereotypes and identity, gendered beliefs and expectations, and gender stereotype threat. She is a researcher on several projects exploring, for example, the influence of gendered teacher expectations of academic achievement on male and female students, and the relationship of adolescent gender self-concept with school belonging, ethnic identity, and stress and coping at school.

Annaline Flint is a lecturer and researcher in the School of Learning Development and Professional Practice at the University of Auckland's Faculty of Education and Social Work. Her current research is mainly focused on the relationships between various teacher beliefs and how these influence student achievement and link to teacher practices. Annaline also has interests in teacher expectations. Annaline is further interested in the induction and mentoring of novice teachers, and teachers' perceptions of values education. Annaline has published articles in journals such as the *British Journal of Educational Psychology* and *Professional Development in Education* and has authored/co-authored three book chapters.

Lynda Garrett is a Senior Lecturer in the School of Learning Development and Professional Practice in the Faculty of Education at the University of Auckland. Her research has focussed on the talent development process for young gifted and talented students in the verbal-linguistic domain, and the influence of teacher expectations on young gifted and talented readers. Lynda has presented at national and international conferences on gifted and talented education, and is building a publishing profile within journals such as *English Teaching: Practice and Critique*, and *The European Journal of Social and Behavioural Sciences*.

Lyn McDonald is a Senior Lecturer at the Faculty of Education and Social Work at the University of Auckland. Her primary areas of research are learning that occurs within practicum, self-regulation and the role of visiting lecturers within initial teacher education. The second body of research is in the field of raising teacher expectations, changing beliefs and enhancing student achievement. Currently Lyn is involved in two research projects –

initial teacher education partnerships between universities and schools and the beliefs of 4-6 year olds in relation to their teachers' expectations.

Funding details

This study was supported by the Marsden Fund Council from Government funding administered by the Royal Society of New Zealand.

Biographical notes

Christine Rubie-Davies is a Professor in the Faculty of Education and Social Work at the University of Auckland. Her primary research interest is in teacher expectations. Her studies focused on teacher expectations at the whole-class level have ignited debates about how teacher beliefs can moderate expectation effects on students. Her experimental study funded by a Marsden Fast Start grant showed large effects on student learning when intervention teachers learnt the practices of high expectation teachers. Christine's book *Becoming a high expectation teacher: Raising the bar* (2015) provides a replicable model for teachers and researchers to implement high expectation practices into any classroom. Christine has an extensive network of international colleagues in the US, the UK, Israel, the Netherlands, Germany and Portugal and has given many invited presentations in national and international contexts in both research and education practice settings. She is a Fellow of the Association of Psychological Studies (US), has almost 100 peer reviewed articles and chapters to her credit many in high-level outlets. Christine has three sole-authored books. She won a New Zealand national teaching excellence award in 2007 and has thrice won awards for her reviewing for *Review of Educational Research* (2013, 2015, 2017).

Penelope Watson is a Lecturer at the University of Auckland's Faculty of Education and Social Work. Her research interests are gender, gender stereotypes and identity, gendered beliefs and expectations, and gender stereotype threat. She contributes as a researcher to several projects exploring, for example, the influence of gendered teacher expectations of academic achievement on male and female students, and the relationship of adolescent gender self-concept with school belonging, ethnic identity, and stress and coping at school.

Annaline Flint is a lecturer and researcher in the School of Learning Development and Professional Practice at the University of Auckland's Faculty of Education and Social Work. Her current research is mainly focused on the relationships between various teacher beliefs

and how these influence student achievement and link to teacher practices. Annaline also has interests in teacher expectations, having been a senior researcher on an experimental Teacher Expectation Project funded by a Marsden Fast Start grant which examined how teacher expectations can be significantly raised and beliefs changed through intervention to enhance student achievement. Annaline is further interested in the induction and mentoring of novice teachers, and teachers' perceptions of values education. Annaline has published articles in journals such as the *British Journal of Educational Psychology* and *Professional Development in Education* and has authored/co-authored three book chapters.

Lynda Garrett is a senior lecturer in the School of Learning, Development and Professional Practice, within the Faculty of Education and Social Work at the University of Auckland. Her research has focused on motivation, social-emotional aspects of giftedness, and the talent development process for young gifted and talented students in the verbal-linguistic domain. As a Senior researcher and co-author within the Teacher Expectation Project research team led by Professor Christine Rubie-Davies, Lynda has contributed to collaborative research outputs considering the influence of teacher expectations on achievement outcomes for primary-school students in the New Zealand context, and undertaken specific research on the influence of teacher expectations on achievement outcomes for young gifted and talented readers. Lynda has presented at national and international conferences on gifted and talented education, and is currently building a publishing profile within journals such as *English Teaching: Practice and Critique*, and the *European Journal of Social and Behavioural Sciences*.

Lyn McDonald is a Senior Lecturer in the Faculty of Education and Social Work at the University of Auckland. She has two areas of research which sit within the discipline of educational psychology. The first area of research concerns the learning that occurs within practicum, and the role of visiting lecturers within initial teacher education. This was the

focus of Lyn's doctorate, '*Visiting lecturers promoting self-regulated learning with student teachers*' (2016). In 2018, two chapters have been published based on this research. The second body of research which illustrates Lyn's contribution to research in teacher practice is in the role as Senior Researcher in an externally funded empirical research project 2011-14 (awarded \$300,000 from the Marsden fund) entitled '*Teacher Expectations: an intervention study*'. Papers from the research project have been collaboratively written and accepted for multiple journals and presented and published at both local and overseas conferences. Lyn has also co-authored a chapter for a textbook, based on the project..

Abstract

Over the past 50 years, research has shown that teacher expectations can influence student outcomes. Many studies have examined within-year effects. There is, however, a paucity of research that has investigated the stability of teachers' expectations across a single year and there are no studies that have examined stability beyond one year. The current study examined the stability of 94 teachers' expectations within three separate years and across three years in mathematics and reading. In both subject areas, teachers' expectations remained relatively stable from beginning to end-of-year for each separate year and also across three years. Additionally, teachers who under- or overestimated their students by more than half a standard deviation, continued to do so across all periods in the study. This was despite the teachers having different students each year. The study suggests that teachers view students' capabilities similarly despite having different student cohorts.

Keywords: teacher expectations, stability, longitudinal, teacher attitudes, teacher perceptions

Viewing students consistently: How stable are teachers' expectations?

Abstract

Teacher expectations and their relations with student achievement have been examined within education for fifty years. However, there remain several gaps in the literature. There is a paucity of literature, for example, that has examined the stability of expectations, across a single year, and no studies that have examined stability beyond one year. The current study examined teachers' expectations for three groups of teachers: those who over- and underestimated their students, and those whose initial expectations were accurate. The study found that the between-group differences in expectations for the three teacher groups were significantly different in each of the three years of the study. Further, similar teachers maintained similar expectations throughout the study's duration, even though the cohorts of students differed each year. The importance of professional development in helping teachers to reduce bias and develop positive expectations for students is discussed.

Keywords: teacher expectations, stability, longitudinal, student achievement, teacher perceptions

Introduction

Teacher expectation research began with the seminal work of Rosenthal and Jacobson (1968) in which, in an experimental study, they showed that when teachers believed that students would do well, the students did achieve at higher levels than previously. This experiment, however, did not provide any information about the mediation of the teacher expectation effect. The work of Brophy and Good (1970) served to illustrate clearly how teachers communicated differential expectations to students and how they treated students differently when they had high or low expectations for them.

Over the intervening years since that original experiment, researchers have shown that teachers use a range of information to form their expectations. For example, student ethnicity (Tenenbaum & Ruck, 2007), social class (Westphal et al., 2016), gender (Robinson-Cimpian, Lubienski, & Ganley, 2014) and special needs labels (Moscardini, 2015) appear to influence the expectations that teachers form of their students. The largest influence on teachers' expectations is the prior achievement information about students that is passed on from teacher to teacher (Rubie-Davies, 2018); previous achievement strongly predicts how students will achieve in the ensuing year.

Researchers other than Brophy and Good (e.g., Babad, 2009) have confirmed the initial findings that teachers do interact quite differently with students for whom they have high or low expectations. These findings have been further confirmed with student explicit reports of how they know that their teacher has high or low expectations for them (e.g., Rio, 2017; Weinstein, 1993, 2002). For example, students report that teachers differentiate in the types of work that is assigned to high versus low expectation students and that teachers interact more positively with those for whom they have high versus low expectations (Babad, 1993, 1998, 2009; Weinstein, 1993; Weinstein & McKown, 1998).

Nevertheless, the vast majority of teacher expectation research has studied expectations within one year (or less). That is, researchers have investigated the relations between teacher expectations and student outcomes controlling for prior achievement and have shown that, from the beginning to the end of one academic year, teachers' expectations do predict both student academic (Glock, Krolak-Schwerdt, & Pit-ten Cate, 2015) and social-psychological outcomes (Zhu, Urhahne, & Rubie-Davies, 2018). Few studies, however, have examined the longitudinal effects of teacher expectations. Alvidrez and Weinstein (1999) obtained teachers' expectations of students aged four years and compared these with students' IQ scores. From these, they then created discrepancy scores indicating the degree to which teachers were accurate in their expectations or over- or underestimated students. The researchers then divided the students into three even groups based on the discrepancy scores. The top third became the overestimated group (high expectations), the middle third were those for whom expectations were relatively accurate, and the bottom third were those who teachers underestimated (low expectations). Alvidrez and Weinstein then conducted their analyses in relation to the over- and underestimated groups. They showed that 4-year-old students from higher socioeconomic backgrounds (SES) tended to be overestimated by their teachers whereas their counterparts were underestimated. Next, using a hierarchical multiple regression and controlling for the students' IQ, SES, and ethnicity, the researchers showed that the degree to which teachers over- or underestimated their students at age four years predicted their grade point average and Scholastic Aptitude Test (SAT) scores 14 years later.

In a more recent study, Hinnant and colleagues (Hinnant, O'Brien, & Ghazarian, 2009) regressed teachers' expectations in reading and mathematics at each of Grades 1, 3, and 5 onto student standardised scores in these subjects in order to obtain a residual score. These residual scores were used to indicate the teachers' expectations, that is, the degree to

which teachers over- or underestimated students. The residual scores were then used to predict achievement in later grades. The researchers showed that in mathematics (but not in reading), teacher expectations of student achievement in Grade 1 predicted their performance at both Grades 3 and 5. Hence, when teachers perceived students to be performing at higher levels than actual achievement suggested, students tended to achieve at higher levels in subsequent years (and the opposite). Expectations at Grades 1, 3, and 5 were correlated suggesting that once students had been over- or underestimated in Grade 1, the discrepancy between teacher expectations and student achievement continued into later years. However, there also seemed to be some carryover effects in that first grade expectations were related to fifth grade achievement and third grade expectations were also marginally associated with fifth grade achievement.

A more comprehensive view of long-term effects, in a recent paper (Rubie-Davies et al., 2014), examined within-year effects, cross-year effects of single teachers, mediated effects of single and multiple teachers, and compounded effects of multiple teachers. The participants were 110 students who were tracked from preschool to Grade 4. As has been found in the association of teacher expectations and student outcomes, accounting for preschool ability, within-year teacher expectations predicted student end-of-year achievement in reading at Kindergarten, Grade 1 and Grade 4. No direct effects were found, however, of teacher expectations at one-level predicting student achievement *across* years. In other words, the direct path from students' class teacher one year to future achievement was not shown to endure in this study. However, the effect of teachers' expectations at one year was associated with achievement in subsequent years (an indirect effect). Expectations in Kindergarten predicted achievement at the end of the year, and achievement in Kindergarten predicted achievement at both Grades 1 and 4. The effects were similar between Grades 1 and 4; controlling for prior achievement, teacher expectations predicted achievement at Grade

1 and Grade 1 achievement predicted Grade 4 achievement. Finally, the authors tested the compounding of teacher expectation effects across successive years, and for both students who were underestimated versus overestimated at four years old. The results showed that the more teachers overestimated students when they were four, the greater were their gains by fourth grade. Similarly, students made fewer gains than their peers when they were underestimated at four years of age. Hence, several years of expectation effects were additive in predicting student achievement at Grade 4, and these effects were evident over and above achievement. Further, the effects were found after controlling for initial ability, gender, and whether or not the child attended preschool.

A further view of long-term teacher expectation effects relates to the stability of expectations, a construct that has been infrequently tested in the literature. Stability relates to 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. As outlined in the research above, if teachers over- or underestimate students, this can have long-term implications for student achievement because if teachers maintain their initially erroneous expectations (they over- or underestimate students), they are likely to plan learning opportunities for students that reflect their expectations. Over one year, this could mean that students achieve more or less than they may have been capable of because of the learning opportunities provided. If successive teachers maintain the initially erroneous expectations, this could further exacerbate the teacher expectation effects. Brophy (1983), however, contended that teachers would most likely adjust their expectations in line with student achievement over the school year which could, therefore, reduce teacher expectation effects. However, other researchers have noted that for particular teachers, this may not be the case. For example, it has been suggested (Babad, 2009; Kuklinski & Weinstein, 2000) that those who are more easily biased by stereotypical information about students would be more likely

to maintain their initial expectations and may disregard disconfirming information.

Nevertheless, very few papers have tested these assumptions and, yet, because expectations are associated with achievement, it is important to investigate whether teachers do adjust their expectations over time, whether they alter them in line with new information, and whether there may be particular teachers who are more likely to maintain initial expectations than others. The purpose of the current paper was to investigate these relations. That is, this paper examined whether teachers maintained their expectations across each of three years and then whether teachers who held initially high or low expectations for all their students in one year, maintained similarly high or low expectations in subsequent years with different groups of students.

Stability of Teacher Expectations

Only two studies have specifically studied the stability of expectation effects; one measured changes over eight weeks (Martinek, 1980) and the other, over six months (Kuklinski & Weinstein, 2000). The findings related to these papers will be presented below.

Martinek (1980) surveyed six physical education teachers from different schools, each of whom taught classes at Grades Two, Four, and Six. He measured the teachers' expectations for each student in each of their three classes (a total of 179 students). Teachers rated their expectations of students in terms of their overall performance in physical education, social relations, cooperation, and cognitive reasoning. Martinek then repeated the measurements eight weeks later. The correlations between the first and second measurement, at all grade levels were very high across all teachers, particularly in terms of overall performance, social relations, and cooperation (ranging from .78 to .96 for individual teachers). The correlations between measurements for cognitive reasoning were slightly lower and ranged from .68 to .92. Martinek argued that this showed that teacher expectations were stable. However, it is interesting that the only area where the correlations were a little

lower was the only one that was not directly observable by teachers. That is, teachers could directly observe students' physical performance, interactions with others (social relations), and cooperation, whereas the area where the correlations were a little lower (cognitive reasoning) may have been more open to teacher bias. However, this study also had limitations in that the study was focused on physical education, rather than a more 'academic' area; it was conducted over a short time frame (only 8 weeks), during which teachers may have recollected their prior expectations; and the study had only a small sample of students.

Kuklinski and Weinstein (2000) addressed some of these limitations in their study. They examined the stability of teacher expectations in two studies. In the first study, 48 Grade One, Three, and Five teachers rank ordered their 464 students in reading during the fall of one academic year. They also completed a rating for each student from 'poor' to 'outstanding' in terms of the teachers' expectations for students' reading achievement. The teachers then ranked and rated their students again in the spring of the following year. In terms of the rank ordering of students, the stability of teacher expectations ranged from moderate to strong (median tau = .69). No differences were found for particular types of teachers (i.e., high and low differentiating teachers – see below) or by grade level. Similarly, the ratings (rather than the rankings) were also similar from the beginning to the end of the academic year. Only 6% of students were rated with more than a one-level change. However, in terms of the ratings (rather than the rankings), the researchers found that teacher expectations were more stable in some classrooms than in others. In classrooms where teachers provided quite different work for high and low achievers and treated these groups quite differently (high differentiating teachers), expectations were much more stable than in classes where teachers treated all students similarly (low differentiating teachers). Low differentiating teachers were much more likely to alter their expectations of students. Further, Kuklinski and Weinstein (2000) found that the ratings were more stable for students

in Grade Five than for those in Grade One. The authors reasoned that this may have been because of accumulated teacher information over time, meaning that teachers believed that prior information about students was an accurate representation of students' capabilities. This may be why other researchers (Borko & Niles, 1982) have shown that the largest effect on teachers' expectations is the prior achievement information that they receive from the previous teacher.

In the second sample of 138 students and 12 teachers, the teachers were again asked to rank-order their Grade One, Three, and Five students in fall and then again in spring and their rank-ordering was compared for the six-month period. The rank ordering showed either moderate or strong stability. The median tau was .65. As with the earlier sample, no differences were found for the rank ordering by teacher type (level of differentiation) or by grade.

Both the studies by Martinek (1980) and by Kuklinski and Weinstein (2000), suggested that teacher expectations appeared to remain comparatively stable for individual students both across an eight-week period as well as over six months (although neither study controlled for student achievement). Whether expectations remain stable for longer than six months is currently unknown. Further, in the study by Martinek, only expectations in physical education were measured and in the Kuklinski and Weinstein research, both studies related to reading. As yet, from the authors' literature searches, no studies have examined the idea of stability of teachers' expectations across more than one curriculum area. For example, it could be that expectations may remain stable in one curriculum area but not in another. In areas like mathematics, for example, where student question responses are generally either right or wrong (i.e., the assessment is fairly objective), teachers may adjust their expectations more readily than in a curriculum area like written language where teacher assessments are more subjective and, therefore, open to bias.

Further, the stability of expectations across several years and with different students have not previously been tracked over time to see whether teachers adjust their own expectations or not as cohorts of students change. This question relates to whether particular teachers may or may not maintain either high or low expectations for their students overall, across more than one year, and with different cohorts of students as well as with the same cohort. In the study by Kuklinski and Weinstein (2000), one aspect of stability that was tested was whether the expectations of some teachers were more stable than the expectations of other teachers. However, although stability was tested at the teacher level rather than the student level, the focus was not on teachers who overestimated their students on average compared with those who underestimated them. The current study was designed to address these gaps in the literature.

Moreover, the Kuklinski and Weinstein study (2000) examined rank order stability, that is, whether or not the teachers' ordering of students from most to least capable varied over time. They also measured rating stability, that is, whether ratings of individual students varied from 'poor' to 'outstanding' over time. On the other hand, Martinek used teacher ratings and examined whether ratings at the first point of data collection were correlated with ratings eight weeks later. The current study also used teacher ratings of students from 'very much below' to 'very much above average' but the data were aggregated at the teacher level in order to examine whether there were particular types of teachers whose expectations were more stable than those of other teachers. Hence, the research questions for this study were:

- 1) Across all teachers in the current study, do they maintain similar expectations within each of three separate years in mathematics and in reading?
- 2) When teachers initially over- or underestimate their students or have accurate expectations for their students in mathematics and reading, are there statistically

significant differences in the expectations of these groups within one year (at two or three time points) and across each of a further two years?

- 3) When groups of teachers over or under-estimate their students at the beginning of one year, do these groups of teachers maintain similar expectation levels (over- or underestimate, or are accurate) even when cohorts of students differ each year?

Materials and Methods

Participants

The participants in this study were 94 teachers who taught in 12 primary (Years 4-6; aged 8-10 years) and intermediate public schools (Years 7 and 8; aged 11-12 years) in low, medium, and high socioeconomic areas. Within New Zealand, schools are not allowed to ask students or parents to provide information about their family socioeconomic status. For example, they cannot ask about parent education or employment. Schools, however, are ranked by the government on a 1-10 scale, largely based on socioeconomic information for the area but the information also includes data related to the percentage of minority group families in the local area. Generally, in New Zealand all students attend their local school. Very few travel to another school. On the basis of the ranking that schools receive, with '1' being a school in a very poor area and '10' being a school in an affluent area, schools are then inversely funded; schools in poorer areas receive significantly more funding than schools in middle class areas. The first author purposively divided schools within one geographical area of a large city into schools in high, middle, and low socioeconomic areas. Schools were then randomly selected within those socioeconomic categories and invited to participate in the study, thus providing a representative sample of students from within each socioeconomic group. Teachers from those 12 schools then volunteered to be part of the study. The teachers' mean teaching experience was 10.45 years ($SD = 9.80$). As illustrated in Table 1, in New Zealand, teachers often teach two class levels in one class (e.g., Year 5/6) particularly

in primary and intermediate schools. The four major ethnic groups are New Zealand European, Māori (the indigenous group), Pacific Island (peoples originating from a Pacific Island such as Samoa, Tonga, Niue or Fiji) and Asian (from South East Asia) and Other.

Further details about teacher participants are contained in Table 1.

INSERT TABLE 1 ABOUT HERE

Materials

Teacher expectations. Expectations were measured during the first month of school each year for three successive years. In that first month, teachers had access to information about students from their previous teachers, including achievement levels, but had not yet conducted their own standardised testing. The teacher expectation survey was also administered in the middle of the first year (not in the middle of other years) and at the end of all three years when teachers predicted the levels they expected each of their students to reach by the end of the following year when they would be with a new teacher. On a 1-7 Likert scale, teachers used a one-item survey to predict the levels they believed each student in their class would have reached by the end of the year, ranging from 1 = very much below average to 7 = very much above average. All teachers used the entire 1-7 range in indicating their expectations for their individual students. For teachers later identified as high expectation teachers (see below), their mean expectation was 5.54 (*SD* 1.31) for mathematics and 5.52 (*SD* 1.32) for reading. Thus, overall, they expected their students to be achieving at moderately above average levels by the end of the year. In contrast, those later identified as low expectation teachers expected their students to reach slightly below average levels in mathematics ($M = 3.85$, $SD = 1.34$) and in reading ($M = 3.86$, $SD = 1.34$). Mid expectation teachers fell between the high and low groups in their expectations, predicting that their students would reach slightly above average levels in mathematics ($M = 4.70$, $SD = 1.36$) and in reading ($M = 4.71$, $SD = 1.46$). Mean teacher expectations on the original 1-7 Likert scale

at each time point can be found in Table 2 (Mathematics) and Table 3 (Reading). Teachers have a different cohort of students each year and, therefore, their own expectations in each successive year relate to a different group of students.

Student achievement. Student achievement was measured using e-asTTle, a standardised measure commonly used in New Zealand. The e-asTTle test enables teachers and researchers to create tests in reading and mathematics by selecting the length and level of the test and the specific areas of the curriculum they wish to be tested. For the purposes of the current study, all tests were 40 minutes long and teachers selected the most appropriate level for the various students in their class to complete. For example, average students in Year 5 (aged 9 years) would be beginning Level 3 of the New Zealand curriculum (Ministry of Education, 2007) because students take two years, on average, to complete each level. Hence, depending on student levels in a particular class, a Year 5 teacher might choose a Level 2/3 test, a Level 3 test, and a Level 3/4 test for her students. Because item response theory (IRT) scoring procedures (Embretson & Reise, 2000) were used to calibrate all items in e-asTTle, students should gain a similar score in their test, no matter which test they sit. In reading, for all tests, the curriculum strands that were tested were processes and strategies, ideas, and language features, and for mathematics these were number knowledge, number sense, and algebra. The e-asTTle system also enables researchers and teachers to create similar tests to ones that students have completed earlier. Hence, the standardised scores from tests administered at different time periods across three years could be compared.

Because of the pre-calibration of the IRT score values, student scores can be tracked over time and compared across classes, year levels, and schools. However, a score of 1460, for example, has a different meaning for a student in Year 4 compared to a student in Year 8. For a Year 4 student, this score would represent the average for their year level whereas for a Year 8 student, the same score would indicate that the student was performing at well below

the average level for their age group. Hence, all student standardised scores were subtracted from the national means for student grade and time of year (means are available for each of four quarters) in order to provide scores that could be meaningfully compared across grade levels.

Procedures

Having gained ethical approval for the study from the institutional committee of the authors, teachers completed their estimates of student achievement in February (beginning year), October and June (first year only) as outlined above.

In parallel, students completed their e-asTTle tests in February and October (and June in the first year). Tests were created by the first author and then couriered to teachers. The first sets of tests were created in consultation with the deputy principals of the schools involved in the study. Subsequent tests were created using the facility within e-asTTle, described above, that enables similar tests to be created for later administrations. Once their students had completed the tests, they were couriered back to the researchers who marked them.

Data Analytic Procedures

Data imputation. Given that this was a longitudinal study conducted over three years, there were some missing data, ranging from 16-28% for individual students. As recommended by Graham (2009), the expectation maximization (EM) method of imputing data was used to impute missing student expectation and achievement scores which were later aggregated to the teacher level as mean residuals. The Little's MCAR test resulted in $\chi^2 [577] = 575.03; p > .05$ showing that there was no discernible pattern in the missing data and that they could therefore be considered missing completely at random (MCAR). Jamshidian (2009) showed that for missingness from 10% to 30%, when the data were MCAR, multiple imputation, EM, and listwise deletion generated similar and largely unbiased estimates.

Further, Dong and Chao-Ying (2013) concluded that the standard errors using EM were closest to those in the complete dataset (when compared with other imputation methods) and that, therefore, EM was the superior method of imputation, especially when data were MCAR.

Teacher expectations. One aim of this paper was to determine whether or not the beliefs of teachers about the capability of their students moderated the stability of their expectations. For the purposes of the current study, differences between teachers who, overall, had high expectations for all their students (high expectation teachers), those whose expectations were accurate, and those whose expectations were low for all students were compared. First, teacher expectations for each individual student were regressed onto achievement providing a standardised residual for each student, indicating the degree to which the teacher over- or underestimated each student (or indicating accuracy). At this point, the residuals were then aggregated at the teacher level for each of the seven time points when expectations and achievement were measured.

The class-level residuals ranged from -1.73 to 1.80 in mathematics and -1.77 to 1.51 in reading at the first time point. Teachers whose mean expectations for their class were more than half a standard deviation above or below the mean at the beginning of the first year of the study were classified as high versus low expectation teachers respectively. Those whose expectations were less than half a standard deviation from actual student achievement were classified as those teachers whose expectations were accurate. Half a standard deviation was selected as the cut-off for grouping teachers, in line with other research that has used standardised residuals and half a standard deviation to differentiate teacher groups as high, mid, or low in terms of their expectations (Rubie-Davies & Peterson, 2016). Further, because one of the aims of this paper was to investigate if particular teacher groups were more likely than others to maintain their initial expectations (in line with the work of Kuklinski &

Weinstein, 2000), it was necessary to create groups that were large enough to enable statistical analyses on the groups. A cut-off of 1.0, for example, would have precluded group comparisons. This resulted in 13 teachers whose expectations were low on average for all students at the initial measurement, 59 whose expectations were accurate, and 22 whose expectations were high in mathematics. In reading, 18 teachers had expectations that were low on average for all students at the initial measurement, 56 had expectations that were accurate and 20 had expectations that were high. Therefore, there was some overlap such that some teachers whose expectations were high in mathematics, for example, also had high expectations of their students in reading, but there was also some variation by curriculum area.

Data Analyses

For the first research question, we asked whether teachers maintained similar expectations within each of three separate years in mathematics and in reading. In order to answer this question, repeated measures Friedman's ANOVAs and Wilcoxon tests were employed to examine within-year stability of expectations.

For the second research question, we asked whether there were statistically significant differences in teachers' expectations for those who initially over- or underestimated their students or had accurate expectations for their students in mathematics and reading, within one year (at two or three time points) and then across each of a further two years. In order to answer this research question, a repeated measures ANOVA was conducted in order to compare the expectations of the different groups of teachers (those who over- or underestimated their students and those who were accurate) over time.

For the final research question, we asked whether when groups of teachers over or under-estimated their students at the beginning of one year, these groups of teachers maintained similar expectation levels (over- or underestimate, or are accurate) even when

cohorts of students differed each year. Repeated measures ANOVAs were conducted by group to investigate whether or not the initial differences in expectations remained similar over time or not.

Results

The means and standard deviations of the standardised residuals for overestimating, accurate, and underestimating expectation teachers can be found in Table 2 (Mathematics) and Table 3 (Reading). It is only when looking at the standardised residuals for individual teacher groups that the residuals can be compared over time. Nevertheless, for the high expectation group in both mathematics and reading, the mean residuals are all above .50 at each time point. In contrast, at every time point, those for low expectation teachers are consistently below -.50 whereas those for mid expectation teachers are greater than -.50 but less than .50. Hence, the mean residuals were similar across years, but were high, mid, or low for the different teacher groups. The degree of variation between the teacher groups and across time was tested in subsequent analyses.

INSERT TABLE 2 ABOUT HERE

The first research question asked whether or not across all teachers in the current study, their expectations were stable over one year in mathematics and in reading. A series of Friedman's repeated measures ANOVAs (which compare rank differences over time) and Wilcoxon tests showed that in mathematics, across the whole sample, teacher expectations did not differ from the beginning to the end of the year for the first year, $\chi^2(2) = 0.53, p > .05$. Wilcoxon tests were used to follow up this finding. It appeared that teacher estimation of their students in mathematics did not change from the beginning of the first year to the middle of the year, $T = 2,317, r = .03$, from the beginning of the year to the end of the year, $T = 2,313, r = .03$, or from the middle of the year to the end of the year, $T = 2,232, r = -.08$. According to Cohen (1988), the first two of these are very small effect sizes whereas the

change from the middle to the end of the year, although not statistically significant, was large. Wilcoxon tests were also used to compare beginning and end of year estimations in mathematics in the second year, $T = 2,567$, $r = .12$, and in the third year of the study, $T = 2,480$, $r = .09$.

Similarly, in reading, teacher expectations for the whole sample appeared to remain stable in the first year, $\chi^2(2) = 0.53$, $p > .05$. Wilcoxon tests were again used to follow up this finding. It appeared that teacher estimation of their students in reading did not change from the beginning of the first year to the middle of the year, $T = 2,029$, $r = -.07$, from the beginning of the year to the end of the year, $T = 2,219$, $r = -.004$, or from the middle of the year to the end of the year, $T = 2,741$, $r = .18$. These are all small effect sizes. Wilcoxon tests were also used to compare beginning and end of year estimations in reading in the second year, $T = 2,150$, $r = -.03$, and in the third year of the study, $T = 2,294$, $r = .02$.

Hence, for each of the three years, in the current study, teacher expectations remained stable within each individual year and, apart from the change from the middle to end of the first year in mathematics, the effect sizes suggest little variation from year to year.

The second research question related to whether or not teachers' expectations for mathematics and reading changed significantly from the beginning to the end of three successive years for teachers who were identified as over- or underestimating, or having accurate expectations for all their students at the beginning of the study. A series of repeated measures ANOVAs in mathematics and reading for each group were performed, and as the means in Tables 2 and 3 suggest, showed that mostly there were no statistically significant differences in mathematics or in reading for any group (high, mid, or low expectations) from the beginning to the end of the three years of the study. However, there was one exception related to the high expectation group in both mathematics and reading as shown below.

For the overestimating group, their expectations were similar in mathematics for the first, $F(2,42) = .86, p > .05$ ($r = -.02$ beginning and mid-year; $r = .11$ beginning and end-of-year; $r = .16$ mid-year and end-of-year); and second; $F(1,21) = .26, p > .05, r = .05$; but not the third year; $F(1,21) = 6.21, p < .02, r = .44$). As can be seen from Table 2, the mean of the overestimating group in mathematics increased in the final year and the effect size was large. For the underestimating group, their expectations in mathematics did not vary for the first, $F(2,24) = .16, p > .05$ ($r = .01$ beginning and mid-year; $r = .10$, beginning and end-of-year; $r = .12$ mid-year and end-of-year); second, $F(1,12) = 2.0, p > .05, r = .28$; and third years, $F(1,12) = .01, p > .05, r = .003$, but although the effect sizes for the first and third years were small, for the second year, the effect size of the difference from the beginning to the end of the year was medium. For the accurate group, their expectations in mathematics were also similar across the first, $F(2,116) = .41, p > .05, (r = .03$ beginning and mid-year; $r = -.03$, beginning and end-of-year; $r = -.09$ mid-year and end-of-year); second, $F(1,58) = .07, p > .05, r = -.14$; and third years, $F(1,58) = .13, p > .05, r = .03$.

Turning to reading, for the overestimating group, their expectations were similar in reading for the first, $F(2,38) = 1.30, p > .05$ ($r = .13$ beginning and mid-year; $r = .20$, beginning and end-of-year; $r = .12$ mid-year and end-of-year); but not in the second, $F(1,19) = 4.82, p < .05, r = -.32$; but were again similar across the third year, $F(1,19) = 4.22, p > .05, r = -.27$. Although the effect size of these differences was small in the first year, it was medium in both the second and third years. As can be seen in Table 3, the mean of the overestimating group in reading increased in both the second and final year from the beginning to the end of the year. For the underestimating group, their expectations in mathematics did not vary for the first ($F(2,34) = .47, p > .05$ ($r = -.12$ beginning and mid-year; $r = -.13$, beginning and end-of-year; $r = -.02$ mid-year and end-of-year); second, $F(1,17) = .49, p > .05, r = .13$; and third years, $F(1,17) = 3.05, p > .05, r = .25$; but although

the effect size was small in the first and second years, it was medium in the final year. For the accurate group, their expectations in mathematics were also similar across the first, $F(2,110) = 1.81, p > .05$ ($r = .12$, beginning and mid-year; $r = .01$, beginning and end-of-year; $r = -.11$ mid-year and end-of-year); second, $F(1,55) = 1.73, p > .05, r = .13$; and third years, $F(1,55) = .15, p > .05, r = .14$. All effect sizes were small for this group showing that there was little variation from the beginning to the end of the year in expectations for the accurate group.

When the expectations in the first year, the second year, and the third year were compared for high, mid, and low expectation groups, however, there was a statistically significant difference in the expectations by group in mathematics for the first year, $F(2, 91) = 114.13, p < .001, partial \eta^2 = .72$; the second year, $F(2, 91) = 56.19, p < .001, \eta^2 = .55$; and the third year, $F(2, 91) = 46.37, p < .001, partial \eta^2 = .51$. All these between-group differences were very large, and differences were found between all groups, $p < .001$, for all differences. The same pattern was found in reading whereby the expectations of the different groups were significantly different from each other in the first year, $F(2, 91) = 127.26, p < .001, partial \eta^2 = .74$; the second year, $F(2, 91) = 51.54, p < .001, partial \eta^2 = .53$; and the third year, $F(2, 91) = 65.04, p < .001, partial \eta^2 = .59$. Again, the effect sizes were very large showing that there were large differences in the expectations of the different groups of teachers. The differences were between all groups at each time point, $p < .001$ for all differences. Of note, however, in both mathematics and reading, the effect sizes did decrease a little in the second and third year compared with those in the first year indicating that although the statistically significant differences remained, these differences were less extreme over successive years of the study.

The final research question asked whether or not expectations were stable overall across the entire three years, as well as for the three teacher groups, rather than just within

each year. A Friedman repeated measures ANOVA was used to examine change in teacher expectations across the entire three years in mathematics and then in reading. This was followed by investigating the group x time interaction using a repeated measures ANOVA, to determine whether the expectations of some of the over-, underestimation and accurate groups remained more similar than in other groups. Across all groups, there was no statistically significant difference in teacher expectations across the three years of the study in mathematics, $\chi^2(6) = 11.52, p > .05$; or in reading, $\chi^2(6) = 5.07, p > .05$. This meant that even though the cohorts of students varied from year to year, teacher expectations remained similar across the three years of the study. An examination of a repeated measures ANOVA group by time interaction showed that in mathematics, $F(7.80, 354.76) = 1.02, p > .05, partial \eta^2 = .02$; and in reading, $F(8.93, 354.76) = 1.23, p > .05, partial \eta^2 = .03$; the interaction was not statistically significant. This showed that also at the group level (high, mid, and low), expectations remained at similar levels over three years, even though the cohorts of students differed.

Discussion

The first research question in the current study asked whether or not teachers maintained similar expectations within each of the three separate years of the study. Overall, the results suggested that, in line with earlier studies (Kuklinski & Weinstein, 2000; Martinek, 1980), teachers' expectations appeared to remain relatively stable over one year. The studies of Kuklinski and Weinstein (2000) were carried out over a six-month period whereas that of Martinek (1980) took place over eight weeks. The current study suggested, that even over a longer period of one year, teachers' expectations showed little relative change. We identified that over one year and for each of three years, when teachers held specific levels of expectations (high, mid or low) for their class, the rank order of the residuals stayed at similar levels each year in both mathematics and in reading. That is, it

seemed that once teachers formed high or low expectations for their class, the lack of change in the rank order of the residuals suggested that they tended to adhere to their expectations for each of three years.

The second research question explored whether any initial differences in expectations between the various groups (over or underestimation, or accurate estimations) remained within the first year of the study and then within the next two years. The study showed that there were large differences in teachers' expectations between teachers who over- or underestimated their students or who were accurate in their expectations. These differences were evident at the beginning of the study and at every other subsequent measurement. These large differences in teachers' expectations are likely to have implications for students in the ways that they are taught by the different teachers. Students whose teachers overestimated their achievement were likely to benefit from teacher practices and behaviours associated with high expectations (Rubie-Davies, 2007). On the other hand, for students who were underestimated, it was possible that they were exposed to lower-level learning opportunities, thereby reducing their chances of making substantial learning gains in one year (Rubie-Davies, 2015). Within the teacher expectation field, there is now much accumulated evidence (e.g., de Boer, Bosker, & Van der Werf, 2010; Glock et al., 2015; McKown & Weinstein, 2008; Rubie-Davies, 2018; Timmermans, Kuyper, & van der Werf, 2015) suggesting that the expectations that teachers hold for their students predict student end-of-year achievement over and above initial levels of achievement. Hence, for students with teachers whose expectations were high, they were likely to be advantaged compared with students who were in classes of low expectation teachers.

The third research question investigated whether or not the different groups of teachers in the study maintained similar levels of expectations from one year to another. Importantly, teachers' class-level expectations showed little variation year-to-year with

different cohorts of students. This meant that some teachers overestimated three separate cohorts of students whereas other teachers underestimated each of their cohorts in both reading and mathematics. This finding suggested that the teachers with high expectations maintained these each year. That is, even with different cohorts of students, their expectations remained high relative to achievement. Similarly, those who underestimated their students' achievement also appeared to do so for successive groups of students. These results add weight to a recent study (Li & Rubie-Davies, 2017) in which it was found that when teachers had high (or low) expectations for a particular class, they maintained similar expectation levels across multiple classes in the same year. Further, our findings are also in keeping with those of Kuklinski and Weinstein (2000) who suggested that some teachers were more negative in their assessments of their students, and perhaps in their interactions, whereas others were generally more positive. The researchers suggested that when students were advantaged by being with more positive teachers, their achievement gains were likely to be greater than if they were with more negative teachers. In earlier work, (Rubie-Davies, 2006; Rubie-Davies, 2007, 2008, 2010; Rubie-Davies & Peterson, 2011) Rubie-Davies was able to show that teachers who had high expectations for all their students had quite different pedagogical beliefs from those of teachers whose expectations were low for all students. High expectation teachers were far more positive in their assessments of students and their capabilities and, therefore, instituted highly effective teaching strategies that seemed to result in their students making much greater learning gains than students with low expectation teachers. Further, high expectation teachers created a much more positive class climate than that of low expectation teachers whereby their relationships with students were very warm and supportive but also where they encouraged students to support and collaborate with each other. These differences in teacher beliefs, instructional practices, and class climate appeared to lead to the large learning gains of students with high expectation teachers and also to

increased student self-belief compared with students in classes of low expectation teachers. This suggests that teachers with high expectations are also high quality teachers whose teaching practices and support of students lead to students in their classes making large learning gains and further fuelling their high expectations.

The findings from the current study, those of Rubie-Davies' earlier work, those of Kuklinski and Weinstein (2000), and those of Li and Rubie-Davies (2017) appear to provide evidence that teacher beliefs moderate teacher expectation effects (Rubie-Davies, 2015; Weinstein, 2002). Therefore, when teachers believe that all students can achieve at high levels, they institute instructional practices and supports which increase the likelihood of student success and so the expectation effects on achievement may be more positive than if teachers did not have such positive beliefs about how they could influence student achievement (Weinstein, 2002).

It is important that teachers can accurately judge student levels of achievement (or at least that they do not underestimate their students) if they are to assign appropriate tasks that are likely to positively influence student learning. There is some evidence (Hinnant et al., 2009; Meissel, Meyer, Yao, & Rubie-Davies, 2017), however, that suggests that some teachers do underestimate their students and that this can have long-term consequences for their students. In the study by Meissel and colleagues (2017), with 17,000 students, teachers were asked to make an overall judgement about student achievement levels in literacy. It was found that, on average, teachers underestimated boys, minority groups, second language students, and special needs students. These underestimations were large (ranging from 6 months for boys to two years below actual student achievement for special needs students) and were likely to result in teachers assigning inappropriate work to some students.

The stability of teachers' expectations in the current study across different teacher groups suggests that teachers likely vary in the degree to which they influence student

achievement. Students make greater gains when their teachers have high expectations for their achievement. Beliefs about the degree of influence that teachers can have on student achievement could arise from prior teacher success in increasing student achievement.

When teachers are confident that they can make a difference to student learning, there are associated benefits for students. For example, when teachers believe that every student can make large learning gains, they are likely to take ownership when students fail to grasp a concept. Such teachers are likely to support students until they grasp the concept being taught. In contrast, teachers who are not confident that they can make much difference to student achievement, may blame the student when the student has difficulty with a new concept. In this scenario, the teacher may credit the student with a lack of intelligence, for example, meaning that the teacher gives up more easily in trying to teach the student since the explanation may be that the student simply cannot learn (Woolfolk Hoy, Hoy, & Davis, 2009). Overall, the study suggested that teachers likely differed in their beliefs and pedagogical practices leading to different outcomes for their students, in turn reinforcing their beliefs about student capability – and, importantly, also leading to expectations that appeared to be stable across more than one year.

Limitations and Future Directions

This study has some limitations that constrain the generalisability of the findings. Firstly, the numbers of teachers who had either high or low class-level expectations were very small and so any conclusions that can be drawn from the study should be made cautiously. Secondly, there are always dangers of increasing the Type 1 error (a false positive) when multiple analyses are conducted on the same data. A further limitation relates to the aggregation of the individual student expectations to the teacher (class) level. The large effect sizes related to the analyses comparing teacher groups may be due to the use of aggregated data at the teacher level. Using the average teacher expectations for their class,

rather than the expectations for each individual student, would have decreased the within-teacher variability in expectations. This aggregation of data would have resulted in lower standard deviations than if the data had been based on individual student-level data, and because r uses the standard deviation in the calculations, a smaller standard deviation would result in larger effect sizes. This needs to be considered in interpreting the results because the effect sizes are not directly comparable to those found in other studies in the teacher expectation field. Similarly, the results are related to teachers' average classroom expectations. This result, however, is not directly generalisable to the expectations for individual students which may show more fluctuations. Future research could examine at the student level the degree to which particular teachers' expectations remain stable over time.

A further limitation of the current study is that although it seemed that expectations were maintained, even for different cohorts of students, no qualitative data were collected to enable more in-depth exploration of why teachers held particularly high or low expectations or how they believed these were enacted in their classrooms. Further, because of the small numbers of teachers in each expectation group, no controls were used in the data analyses. It is possible that those teachers whose expectations were high, low or accurate for their students had particular characteristics. For example, more high expectation teachers could have been in high socioeconomic schools, or had greater teaching experience, or belonged mostly to a particular ethnicity. These are factors that could be investigated in future studies as potentially being related to teachers' class-level expectations.

This is the first study, to the authors' knowledge, that has examined the stability of teachers' expectations not just across a complete academic year but also across multiple years with different cohorts of students. However, the small numbers of teachers in each expectation group mean that it will be important in future studies, that much larger numbers of teachers are included. This will enable researchers to investigate if there are particular

characteristics of teachers or schools that are associated with teachers having high, low, or accurate expectations for all students. Further, the inclusion of a qualitative component will enable a deep exploration of why some teachers consistently have high expectations for their students (as well as the other possibilities) across multiple years. In addition, the inclusion of larger numbers of teachers would enable more robust multilevel analyses to be conducted.

Conclusions

High expectations have long been associated with increased student achievement (Brophy, 1982; Brophy & Good, 1986; de Boer et al., 2010; Dusek & O'Connell, 1973) but to date, studies have not often examined the stability of teachers' expectations. The current study suggests that if expectations remain relatively stable and high at the class-level, then students who are fortunate to be with high expectation teachers are likely to gain in both achievement and self-belief (Rubie-Davies, 2006; Rubie-Davies, 2007). It is important, therefore, that teacher education programmes incorporate the benefits of high expectations for student learning into their teacher education programmes. A recent book (Good & Lavigne, 2018) showed that the inclusion of teacher expectations as a topic was now rare within teacher education and yet the current study and many others show that students can do much better than they may have previously when their teacher has high expectations for them. Both practising and pre-service teachers are not often confronted with topics such as teacher bias (inaccurate low expectations for some groups of students), stereotyping, or the detrimental effects of low expectations, and yet evidence suggests that teachers can increase their expectations if they are made aware of the benefits for students and if they are reminded of quality teaching practices and emotional care that are associated with increased achievement (Rubie-Davies, 2015). Targeted professional support may help underestimating teachers to become more positive in their expectations and to adjust their expectations to enable students to reach their potential. If teachers do not understand how their expectations

can limit or increase student progress, they are not likely to question their current practice.

Teachers play a key role in influencing student achievement and long-term outcomes. It is important that they maximise the capabilities of every student in their care.

In line with this, teachers whose expectations are high (that is, their expectations are well above student achievement) provide evidence (Rubie-Davies, 2007) that student progress could be much greater than it is currently were other teachers to structure their classrooms similarly to those of high expectation teachers. Overestimating students is likely to be far more beneficial for student progress than underestimation or even accurate estimation (Timmermans et al., 2015) and these teacher assessments have long-term consequences for student outcomes and ultimately their life chances (Alvidrez & Weinstein, 1999). Hence, policy makers, school managers, and teachers need to develop an understanding that prior achievement information is not necessarily an indication of students' futures. Students' achievement trajectories can be interrupted such that their learning gains are accelerated and students are enabled to reach the highest possible levels.

References

- Alvidrez, J., & Weinstein, R. S. (1999). Early teacher perceptions and later student academic achievement. *Journal of Educational Psychology, 91*, 731-746.
- Babad, E. (1993). Teachers' differential behavior. *Educational Psychology Review, 5*, 347-376.
- Babad, E. (1998). Preferential affect: The crux of the teacher expectancy issue. In J. Brophy (Ed.), *Advances in research on teaching: Expectations in the classroom* (Vol. 7, pp. 183-214). Greenwich, CT: JAI Press.
- Babad, E. (2009). *The social psychology of the classroom*. New York: Routledge.
- Borko, H., & Niles, J. A. (1982). Factors contributing to teachers' judgments about students and decisions about grouping students for reading instruction. *Journal of Reading Behaviour, 14*(2), 127-140.
- Brophy, J. E. (1982). How teachers influence what is taught and learned in classrooms. *The Elementary School Journal, 83*, 1-13.
- Brophy, J. E. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology, 75*, 631-661.
- Brophy, J. E., & Good, T. L. (1970). Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. *Journal of Educational Psychology, 61*, 365-374.
- Brophy, J. E., & Good, T. L. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Handbook of Research on Teaching* (Third ed., pp. 328-375). New York: MacMillan Publishing Company.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J.: Lawrence Erlbaum.
- de Boer, H., Bosker, R. J., & Van der Werf, M. (2010). Sustainability of teacher expectation bias effects on long-term student performance. *Journal of Educational Psychology, 102*, 168-179. 10.1037/a0017289
- Dong, Y., & Chao-Ying, J. P. (2013). Principled data methods for researchers. *Springerplus, 2*(Online) 10.1186/ 2193-1801-2-222
- Dusek, J. B., & O'Connell, E. J. (1973). Teacher expectancy effects on the achievement test performance of elementary school children. *Journal of Educational Psychology, 65*, 371-377.
- Embretson, S. E., & Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Glock, S., Krolak-Schwerdt, S., & Pit-ten Cate, I. M. (2015). Are school placement recommendations accurate? The effect of students' ethnicity on teachers' judgments and recognition memory. *European Journal of Psychology in Education, 30*, 169-188. 10.1007/s10212-014-0237-2
- Good, T., & Lavigne, A. (2018). *Looking in classrooms* (11th ed.). New York, NY: Routledge.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*, 549-576. 10.1146/annurev.psych.58.110405.085530
- Hinnant, J. B., O'Brien, M. O., & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology, 101*, 662-670. 10.1037/a0014306
- Jamshidian, M. (2009). Strategies for analysis of incomplete data. In M. A. Hardy & A. Bryman (Eds.), *Handbook of data analysis* (pp. 113-130). Thousand Oaks, CA: Sage.

- 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.
- Li, Z., & Rubie-Davies, C. M. (2017). Teachers matter: Expectancy effects in college foreign language classrooms. *Studies in Higher Education*, 47, 2042-2060. 10.1080/03075079.2015.1130692
- Martinek, T. J. (1980). Stability of teachers' expectations for elementary school aged children. *Perceptual and Motor Skills*, 51(3, Pt 2), 1269-1270.
- McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context and the achievement gap. *Journal of School Psychology*, 46, 235-261. 10.1016/j.jsp.2007.05.001
- 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. org/10.1016/j.tate.2017.02.021
- Ministry of Education. (2007). *The New Zealand curriculum*. Wellington: Learning Media
- Moscardini, L. (2015). Primary special school teachers' knowledge and beliefs about supporting learning in numeracy. *Journal of Research in Special Educational Needs*, 15, 37-47. 10.1111/1471-3802.12042
- Rio, N. T. (2017). *Teacher expectations and ethnicity: Student and teacher perspectives*. (Doctor of Philosophy), The University of Auckland, Unpublished doctoral thesis, Auckland.
- Robinson-Cimpian, J. P., Lubienski, S. T., & Ganley, C. M. (2014). Teachers' perceptions of students' mathematics proficiency may exacerbate early gender gaps in achievement. *Developmental Psychology*, 50, 1262-1281. 10.1037/a0035073
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*. New York: Holt, Rinehart & Winston.
- Rubie-Davies, C. M. (2006). Teacher expectations and student self-perceptions: Exploring relationships. *Psychology in the Schools*, 43, 537-552. 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. 10.1348/000709906X101601
- Rubie-Davies, C. M. (2008). Teacher beliefs and expectations: Relationships with student learning. In C. M. Rubie-Davies & C. Rawlinson (Eds.), *Challenging thinking about teaching and learning* (pp. 25-39). Haupauge, NY: Nova.
- Rubie-Davies, C. M. (2010). Teacher expectations and perceptions of student attributes: Is there a relationship? *British Journal of Educational Psychology*, 80, 121-135. 10.1348/000709909X466334
- Rubie-Davies, C. M. (2015). *Becoming a high expectation teacher: Raising the bar*. London: Routledge.
- Rubie-Davies, C. M. (2018). *Teacher expectations in education*. New York, NY: Routledge.
- Rubie-Davies, C. M., & Peterson, E. R. (2011). Teacher expectations and beliefs: Influences on the socioemotional environment of the classroom. In C. M. Rubie-Davies (Ed.), *Educational psychology: Concepts, research and challenges* (pp. 134-149). London: Routledge.
- Rubie-Davies, C. M., & Peterson, E. R. (2016). Relations between teachers' achievement over- and underestimation, and students' beliefs for Maori and Pakeha students. *Contemporary Educational Psychology*, 47, 72-83. 10.1016/j.cedpsych.2016.01.001

- Rubie-Davies, C. M., Weinstein, R. S., Huang, F. L., Gregory, A., Cowan, P., & Cowan, C. (2014). Successive teacher expectancy effects across the early school years. *Journal of Applied Developmental Psychology, 35*, 181-191. 10.1016/j.appdev.2014.03.006
- Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology, 99*, 253-273.
- Timmermans, A. C., Kuyper, H., & van der Werf, M. P. C. (2015). Accurate, inaccurate or biased teacher expectations. Do Dutch teachers differ in their expectations at the end of primary education? . *British Journal of Educational Psychology, 85*, 459-478. 10.1111/bjep.12087
- Weinstein, R. S. (1993). Children's knowledge of differential treatment in school: Implications for motivation. In T. M. Tomlinson (Ed.), *Motivating students to learn: Overcoming barriers to high achievement* (pp. 197-224). Berkeley, CA: McCutchan.
- Weinstein, R. S. (2002). *Reaching higher: The power of expectations in schooling*. Cambridge, MA: Harvard University Press.
- Weinstein, R. S., & McKown, C. (1998). Expectancy effects in "context": Listening to the voices of students and teachers. In J. Brophy (Ed.), *Advances in Research on Teaching. Expectations in the Classroom* (Vol. 7, pp. 215-242). Greenwich, Connecticut: JAI Press.
- Westphal, A., Becker, M., Vock, M., Maaz, K., Neumann, M., & McElvany, N. (2016). The link between teacher-assigned grades and classroom socioeconomic composition: The role of classroom behaviour, motivation, and teacher characteristics. *Contemporary Educational Psychology, 46*, 218-227. 10.1016/j.cedpsych.2016.06.004
- Woolfolk Hoy, A., Hoy, W. K., & Davis, H. A. (2009). Teachers' self-efficacy beliefs. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of Motivation in School* (pp. 627-653). New York: Routledge.
- Zhu, M., Urhahne, D., & Rubie-Davies, C. M. (2018). The long-term effect of teacher judgement and differential teacher treatment on students' academic outcomes. *Educational Psychology: An International Journal of Experimental Educational Psychology, 38*, 648-688. 10.1080/01443410.2017.1412399

Table 1. *Demographics for Teacher Participants (N = 94)*

Teacher Demographics (Percentage)							
Gender	Male				Female		
	27				73		
Ethnicity	European		Māori	Pacific Island		Asian and Other	
	62		10	10		18	
Teaching Experience	1-5 years		6-10 years		11-17 years		18+ years
	33		23		18		26
School SES	Low		Medium		High		
	19		62		19		
Class Level Taught	Y4	Y5	Y5/6	Y6	Y7	Y7/8	Y8
	31	4	26	4	10	15	11

Table 2. Mean Standardised Residuals, Teacher Expectations (original 1-7 Likert scale) and Standard Deviations of Teacher Expectations Overall and by High, Mid, and Low Expectation Groups over Three Years in Mathematics

		Mathematics						
		Year One			Year Two		Year Three	
		Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phase 7
High Expectation <i>n</i> = 22	Mean	.69	.71	.60	.76	.81	.66	.93
	<i>SD</i>	.52	.35	.31	.44	.47	.45	.75
1-7 Likert scale	Mean	5.54	5.18	5.29	5.53	5.14	5.56	5.57
	<i>SD</i>	1.31	1.23	1.28	1.31	1.38	1.32	1.42
Mid Expectation <i>n</i> = 59	Mean	-.003	-.03	.03	.02	.17	-.002	-.03
	<i>SD</i>	.53	.32	.30	.51	.55	.39	.56
1-7 Likert scale	Mean	4.70	4.57	4.62	4.73	4.60	4.59	4.64
	<i>SD</i>	1.36	1.38	1.41	1.37	1.48	1.47	1.60
Low Expectation <i>n</i> = 13	Mean	-.84	-.85	-.92	-.56	-.72	-.71	-.71
	<i>SD</i>	.49	.33	.32	.26	.31	.42	.90
1-7 Likert scale	Mean	3.85	3.34	3.68	4.24	4.13	4.14	4.27
	<i>SD</i>	1.34	1.47	1.42	1.43	1.71	1.45	1.61
Total Residual <i>n</i> = 94	Mean	.00	.00	.00	.00	.00	.00	.00
	<i>SD</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Residual range	Minimum	-1.73	-1.40	-1.43	-1.30	-1.13	-1.71	-1.78
	Maximum	1.80	1.09	1.02	1.51	2.56	2.08	2.44

Running Head: STABILITY OF TEACHER EXPECTATIONS

Table 3. Mean Standardised Residuals, Teacher Expectations (original 1-7 Likert scale) and Standard Deviations of Teacher Expectations Overall and by High, Mid, and Low Expectation Groups over Three Years in Reading

		Reading						
		Year One			Year Two		Year Three	
		Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Phase 7
High Expectation <i>n</i> = 20	Mean	.82	.76	.70	.62	.81	.54	.76
	<i>SD</i>	.31	.20	.28	.20	.47	.47	.30
1-7 Likert scale	Mean	5.52	5.39	5.43	5.63	5.48	5.53	5.62
	<i>SD</i>	1.46	1.43	1.44	1.44	1.50	1.48	1.52
Mid Expectation <i>n</i> = 56	Mean	.06	-.04	.05	.07	.17	-.02	.01
	<i>SD</i>	.40	.45	.40	.59	.55	.47	.51
1-7 Likert scale	Mean	4.71	4.67	4.68	4.62	4.72	4.75	4.69
	<i>SD</i>	1.46	1.43	1.46	1.44	1.50	1.48	1.52
Low Expectation <i>n</i> = 18	Mean	-.80	-.71	-.70	-.54	-.72	-.62	-.90
	<i>SD</i>	.36	.33	.36	.69	.31	.43	.63
1-7 Likert scale	Mean	3.86	3.36	3.78	4.29	4.21	4.14	4.18
	<i>SD</i>	1.34	1.52	1.40	1.49	1.80	1.74	1.81
Total Residual <i>n</i> = 94	Mean	.00	.00	.00	.00	.00	.00	.00
	<i>SD</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Residual range	Minimum	-1.77	-1.78	-1.85	-1.27	-1.98	-2.74	-2.01
	Maximum	1.51	1.06	1.41	1.90	1.81	1.24	1.88