THE UNIVERSITY OF MELBOURNE

Melbourne Graduate School of Education
Assessment Research Centre

Achieving Mathematics Growth for High Capacity Students

As part of the Realising the Potential of Australia's High Capacity Students Linkage Project

20th June 2018

This research is supported under the Australian Research Council's *Linkage Projects* funding scheme (LP140100236, *Realising the Potential of Australia's High Capacity Students*), in partnership with the State of Victoria as represented by the Department of Education and Training.

Table of Contents

Figures

Tables

Project Team

Chief Investigators	**Emeritus Professor Patrick Griffin (UoM)** **Associate Professor Esther Care – 2015 (UoM)** **Professor Lorraine Graham 2016 – 2018 (UoM)**
Partner Investigator	Dr Elvira Vacirca (DET) – 2016 Tamara Lindorff (DET) – 2017–2018
Lead Researcher	Dr Susan-Marie Harding (UoM)
Doctoral Researchers	Nives Nibali (UoM) Narelle English (UoM) Jolanta Szymakowski (UoM
Staff	BM Monjurul Alom (UoM) Jessica Arnup (UoM) Narelle English (UoM) Dr Zhonghua Zhang (UoM) Scott Ware (DET) Deborah Cordingley (DET) Jason Muir (DET) Stephanie Moorhouse (DET) Kenneth Cole (DET)

This document was prepared by Dr S. Harding, Assessment Research Centre, University of Melbourne.

For questions email s.harding@unimelb.edu.au.

Cite this report as:

Harding., S., Griffin., P., Graham., L., Arnup., J., Szymakowski, J., English., N., Nibali., N., Alom, BM., and Zhang., Z. (2018). *Achieving mathematics growth for high capacity students: Realising the potential for Australia's high capacity students*. Assessment Research Centre, Melbourne Graduate School of Education.

Introduction

Project Background

A previous Assessment Research Centre study “Assessment and Learning Partnerships” showed that despite overall average gains in student reading and mathematics, almost all the gains were achieved among the bottom quartile group of students and at low levels of proficiency in both reading and mathematics. This phenomenon was referred to by Professor Patrick Griffin as “flatlining” (Topsfield, 2013). The lack of growth among high capacity students (top 25% of a class in a particular subject) in reading and mathematics was notable. Analysis of value-added achievement at the teacher level showed that there were variable impacts between teachers at every level of proficiency, and within teachers, there was variable impact across levels of proficiency.

The Assessment Research Centre, in partnership with the Department of Education and Training, conducted research into “Realising the Potential of Australia’s High Capacity Students” (REAP). The project involved two data collection years, 2016 and 2017, where 58 schools asked teachers of Years 5 to 8 to assess students in March/April (T1) and then in September/October (T2) to determine their growth in mathematics. Concurrently, teachers read and completed eight professional development (PD) modules focusing on Identifying High Capacity Students, Zone of Proximal Development (ZPD), Rubrics, Assessment for Growth, Students’ Self-Regulated Learning (SRL), Teaching SRL in the classroom, and Targeted Teaching and Monitoring Progress. Teachers participated in the project to try to ameliorate the flatline through capitalising on their students’ ability to regulate their own learning and targeted teaching at the student ZPD.

While teachers completed PD modules, they also completed accompanying questionnaires asking them to reflect on and report the teaching practices they used to target their teaching for high capacity students. This allowed researchers to collate responses and create composite variables based on the methods teachers were using for high capacity students in reading comprehension. The composite teaching variables (teacher covariates) were modelled against students’ T2 reading comprehension achievement, while controlling for students’ T1 reading comprehension achievement, to determine which teacher variables were successful in achieving growth for students. This analysis required a multi-level modelling approach to control for T1 differences across teacher classes. Findings are reported for successful teacher practices for both whole class and high capacity students’ growth, as it was the intention of the REAP project to find teaching practices that allow equitable progress for high capacity students, but not at the expense of the low achieving students.

Differential Growth

The flatline phenomenon refers to reduced growth or lack of growth for high capacity students in comparison to their class-matched peers. There is a need to acknowledge and describe the various causes, or possible causes, of a flatline in growth and to analyse and interpret data with caution based on those potential causes. Rather than continue to describe a flatline, it might be beneficial to use a term such as "differential growth", as this term encompasses and includes situations where certain sub-populations of students (such as high achievers) have reduced growth or reduced achievement rather than no growth at all. The term flatline implies that the students in the sub-population in question do not achieve anything within the scope of the testing; however, this is not always the case. Below, the various causes of differential growth are described, with reference to literature that supports arguments for each as existing within pre–post student testing data.

1. *Cognitive-/age-based differential growth*

Cognitive differential growth refers to the notion that as students reach the higher end of proficiency on a latent construct, their rate of learning slows. High order and complex skills are proposed to take longer to acquire and master than basic skills. The rate of growth at higher levels of development may then be lower than the rate of growth at lower levels, as can be demonstrated by data collected in the REAP project (Figure 1). Shanley (2016), argues that the cognitive-based differential growth pattern could result from the "sensitivity of early mathematics assessments, which test discrete whole number skills like counting, cardinality, number identification, and basic calculation, lend themselves to more impressive rates of growth, whereas the more applied nature of late elementary and middle school mathematics achievement makes it more difficult to track and measure progress or growth" (p. 8).

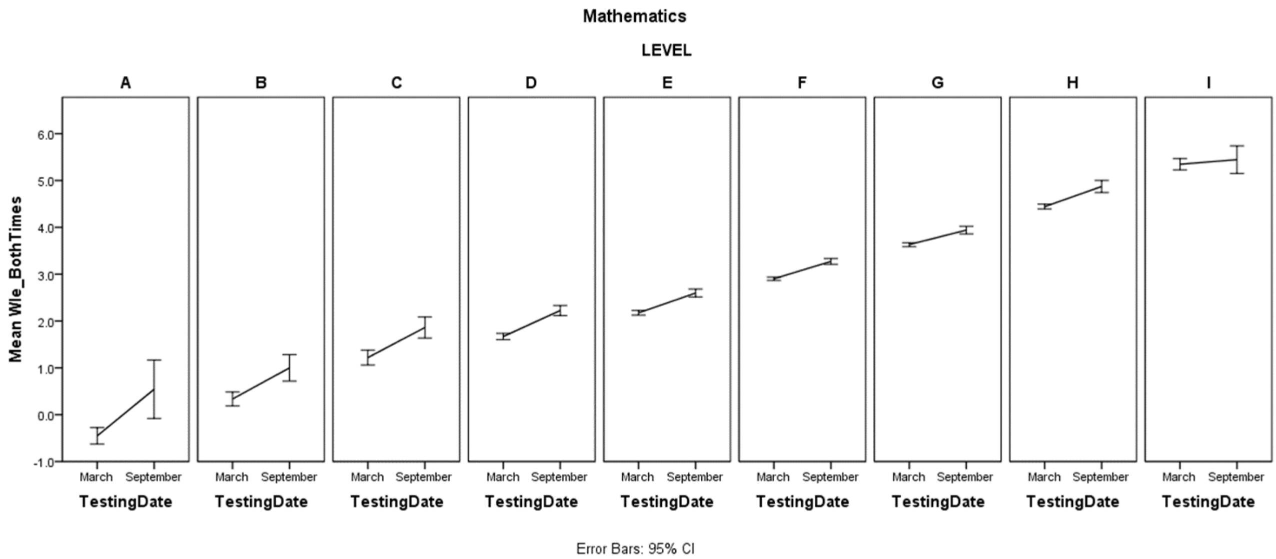

Figure 1. Cognitive-based differential growth (REAP data).

This phenomenon can manifest as *reduced average student growth as students progress through their schooling years* because students who are in older year levels are (on an aggregate basis) students whose ZPD is at the higher end of the latent continuum (Figure 2). Thus, an effect based on cognitive skill acquisition may also be described as an age- or grade-based phenomenon. This kind of differential growth pattern can be observed in almost all large-scale testing results that track students longitudinally (such as NAPLAN).

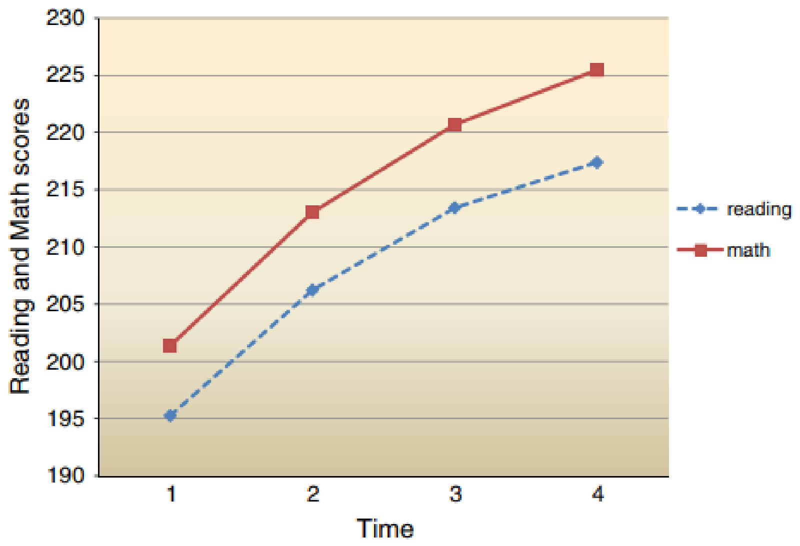

Fig. 2. Mean growth curve of reading and mathematics achievement plotted against time. Note. Using four-wave longitudinal data, Times 1, 2, 3, and 4 respectively indicate test scores in the 4th, 5th, 6th, and 7th grades.

Figure 2. Age-based differential growth (Shin, Davison, Long, Chan, & Heistad, 2013).

The cognitive-based phenomenon can also manifest as a *reduction in high performing students' growth in comparison to low performing students,* as high performing students are also those who have a ZPD at the higher end of the latent continuum. This type of differential growth pattern has direct implications for the analysis of growth trends within the data reported in this study. To illustrate, the 2016 REAP growth data for reading comprehension is graphed by year levels (panels), showing increased cognitive ability on the y-axis (Figure 3), with high ability students showing reduced growth compared to low ability students.

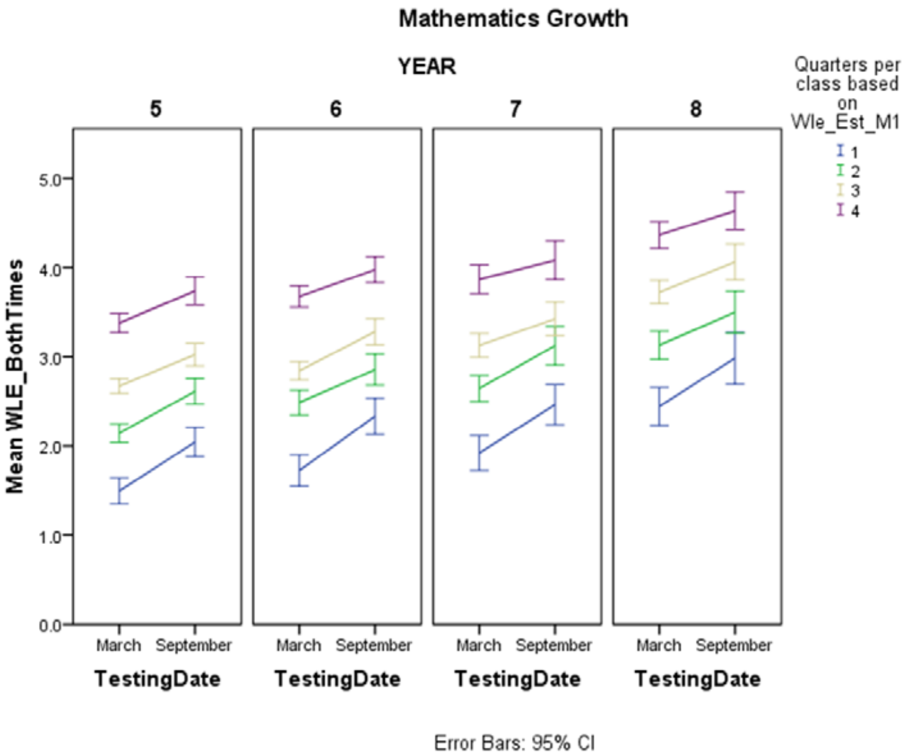

Figure 3. Reduced high performing student growth (purple lines) compared with low performing student growth (blue lines) as a possible function of cognitive differential growth (REAP data).

2. Time series-based differential growth

Time series-based differential growth describes a situation where students do not achieve higher levels of cognitive understanding across years. Data is not longitudinal as the same students are not tracked, but the achievement of students at particular grades or ages is measured across calendar years, such as in PISA, PIRLS, TIMMS or NAPLAN. Many of these time series analyses show that students are not growing over time (see, for example, Figure 4, where NAPLAN results stagnate from 2008 to 2017). When sub-populations of students are considered, such as high and low performing, differences can be seen in the growth across years, which can be considered a time series-based differential growth effect. Figure 5 demonstrates increased growth for students within the 10th percentile compared with students in the 90th percentile on NAEP testing data (Shin et al., 2013). This difference in growth may be related to differences in teaching, testing or educational advances that favour the low performing students.

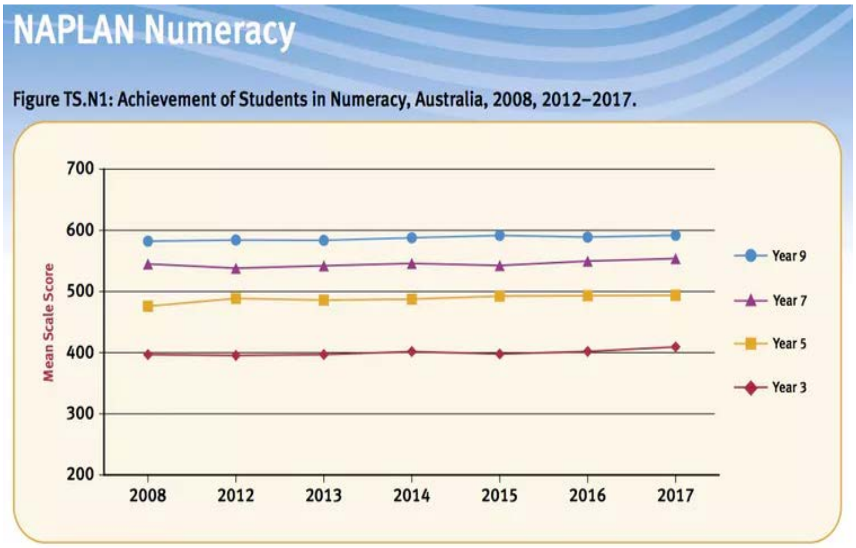

Figure 4. Time series-based differential growth: National Numeracy Trends 2008–2017 (ACARA, 2017).

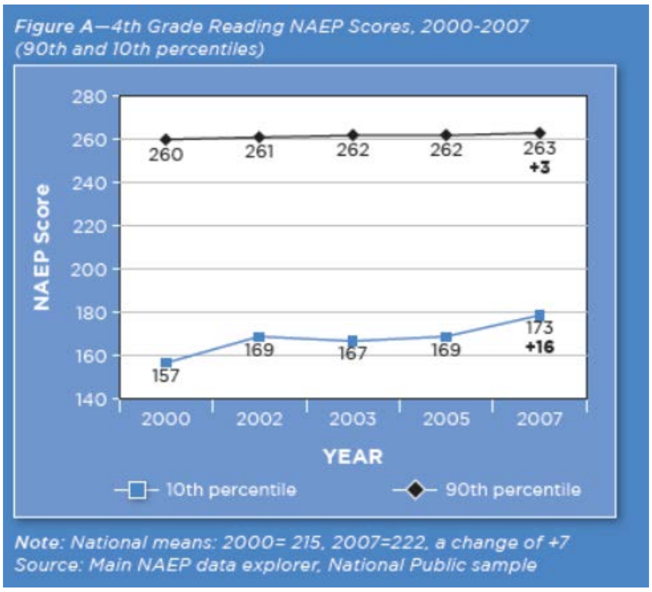

Figure 5. Time series-based differential growth: NAEP data 2000–2007 (Loveless, 2008).

3. *Testing-based differential growth*

Differential growth may be caused or explained by issues with either testing or interpreting growth measurements. In terms of testing, the potential issue is often referred to as a “testing ceiling”, which is the proposition that students of high ability “top the test” or, in other words, answer most or all items correctly. When this occurs, students cannot show progress on a second or subsequent test. Most large-scale or other testing systems know this nowadays, so items are included beyond which the students can reasonably be expected to answer. This provides opportunity for highly able students to display growth, if a growth measure is required from the data. However, even if there are no students who top the test, there are still many reasons why testing issues may themselves cause differential growth patterns, including because they exhibit the cognitive differential growth pattern described above.

Regression to the mean is one problem that can occur when analysing student growth over time. Regression to the mean refers to the tendency for students achieving at the tail end of an assessment (students that are achieving very high or very low) to receive a score closer to the mean on the second assessment. Regression to the mean occurs if scores are not perfectly correlated (Lohman & Korb, 2006). The observed score of a student on a test is a measure of the true ability of the student on the latent construct, plus any testing or measurement errors. The student's true ability is not and cannot be directly measured. Probability of over-estimation of the true score increases as the student's observed score reaches the higher end of the scale and the probability of under-estimation of the true score increases as the student's observed score reaches the lower end of the assessment. Therefore, on the second testing measure, a student's ability can regress towards the mean.

There are numerous examples that can be used to explain regression to the mean. Figure 6 shows baseball batter averages and changes over two years. When players are divided into quartiles (minimum 300 bats), the regression to the mean is obvious between years. The pattern is anticipated and sports statisticians use the calculation to predict future batting averages.

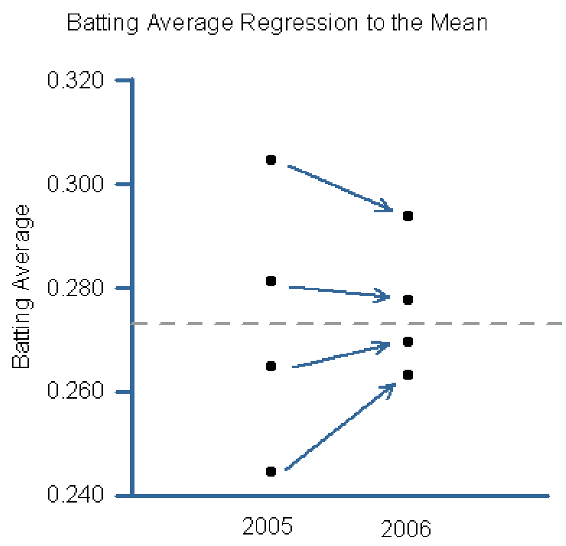

Figure 6. Regression to the mean differential growth: Batter averages for baseball (Studeman, 2007).

At least two contributing factors to regression to the mean can be avoided:

1) Testing errors, which are generally higher for students at either tail of the normal distribution, contribute to regression to the mean as increased error at each tail of the distribution increases the likelihood that ability levels have been over- or under-estimated. Testing errors can be reduced by using computer adaptive or quasi-computer adaptive testing or various forms of test administration and/or equating methods that result in students sitting an assessment with greater testing information and reduced error levels at each tail of proficiency (for example, the Assessment Research Centre's Online

Testing System [ARCOTS] described by Griffin (2014), which asks the teacher to choose the appropriate test form for each student in a class following which tests are equated, allowing test information to be maximised for each student and measurement errors at the tails of the student distribution to be reduced).

2) Luck. There is an element of luck and/or chance involved in testing, which cannot and should not be ignored. When students complete a test, they may be more or less familiar with the content, context or type of question offered. This is an inbuilt predictive validity issue in testing, which can be avoided to some degree by maintaining testing confidentiality, increasing the variation of items covering the latent construct and avoiding testing topics strictly related to concepts that advantage sub-populations of students (avoiding bias). Similarly, student anxiety or unfamiliarity with the testing format can result in under-estimation of the true score during the first assessment. It is possible to avoid or reduce some of these factors; however, it is pertinent to acknowledge that students can still have a good day or a bad day.

The effect of guessing is another issue in measuring student growth, which is especially pertinent when students are completing multiple-choice-based assessments. Students at the lower end of the latent continuum have a higher propensity (or need) to "guess" the response to items they do not know the answer to. Similarly, responses to difficult items are more likely to be guessed than responses to easier items. Experts agree that traditional methods favour the low ability student: "Generally speaking, only lower ability people have the opportunity to do much random guessing; clearly the farther up the ability continuum, the less the opportunity to guess" (Waller, 1974, p. 40). The logical consequence is that low ability student estimates are inflated if corrections for guessing do not take place. The relationship is not linear with high ability students (Andrich, Marais, & Humphry, 2012); therefore, the effect of guessing may contribute to differential growth pattern between high and low ability students.

4. Society influences on differential growth

Other than cognitive- and testing-based differential growth patterns, many researchers and testing agencies focus on society or personal variables and study the differences between growths of sub-groups of students. This can be more or less accurate depending on the analysis method and the structure of the data. Different types of "gap" in student progression have been described:

<u>Achievement gap</u> – gap in achievement of sub-groups of students in a particular year level. Chubb and Loveless (2002) looked at the "achievement gap" of the lowest and highest achieving students in a year. Other achievement gaps exist, for example, depending on gender, social economic status and presence of learning difficulty. The term achievement gap commonly refers to the disparity in achievement between groups of students. Gaps in achievement can be measured in terms of various factors, such as gender, ethnocultural background, socio-economic status, special education needs, language proficiency or

number of credits accumulated by the end of a particular grade. Achievement gaps can also be defined according to combinations of these factors, such as gender and special education needs, gender and socio-economic status or ethnocultural background and credit accumulation by year and grade.

Racial differences gap – gap in achievement seen between races; that is, in the United States between African American/Hispanic vs Caucasian Americans, and in Australia between Indigenous and non-Indigenous (for example in Song, Perry, & McConney, 2014).

Excellence gap – gap in achievement between groups at high levels. Sub-groups may include race, socio-economic status (SES) and English as additional language (EAL) students. An excellence gap is simply the difference in percentage of students scoring at advanced levels. For example, in 2015, 22% of Asian students scored advanced on the 4th grade NAEP mathematics test compared to 10% of Caucasian students (Plucker, Burroughs, & Song, 2010). Excellence gap refers to the disparity in the percentage of lower income versus higher income students who reach advanced levels of academic performance. The "gap" appears in elementary school and continues as students move through middle school, high school, college and beyond (Plucker et al., 2010). Only a limited amount of research has been conducted on achievement gaps among students who perform at advanced levels, but existing research shows that the educational system systematically short-changes certain populations of students capable of reaching high levels of academic performance. Much of this research focuses on the gaps among White and Black students; similar gaps involving Hispanic, free lunch-eligible and English language learning students are largely ignored (Plucker et al., 2010).

Learning gap – often used to refer to the gap between a student's actual achievement and their potential for achievement, such as in the *Learning for All* document prepared by the Ontario Public Service (2013) for the Ministry of Education in Canada.

5. *Transition-based differential growth*

Two major transition-based differential growth patterns exist within the assessment measurement domain. The first describes a reduction in students' ability during the transition from primary to secondary school or from secondary to tertiary school. This differential growth pattern has multiple potential causes such as "teacher content knowledge and pedagogy, curriculum sequencing from primary to secondary school, communication between primary and secondary schools, socioeconomic factors, family support, social adjustment, and students' self-efficacy" (Hopwood, Hay, & Dyment, 2017, p. 48).

The second is related to a teaching-based differential growth pattern, where students at the boundaries of schooling receive different instruction based on their ability level. An example of this phenomenon can be seen in the data collected by this study, in which Grade 6 high ability students progress at a lower rate than high ability Grade 5 students. The theoretical basis is that teachers are either not as able or not as willing to stretch the highly able students at the end of the primary schooling years, or that their focus is on the lower achieving students in the year prior to transition to ensure they do not get left behind. The first year of secondary schooling may share a similar transition effect; teachers may attempt to bring students up to a certain standard so that they can ensure all students in the class are prepared for the next part of the curriculum. It is possible that some countries intentionally build redundancy into the curriculum across the transition years. If true, this would increase the likelihood of a transition-based differential growth pattern in the first year of secondary school.

6. ***Teaching-based differential growth***

A teaching-based differential growth pattern is that which is caused by teachers targeting their approaches to students who are at the lower or middle ranges of classroom ability. This growth pattern is the focus of the REAP project. As Emeritus Professor Patrick Griffin stated in a submission to the Senate NAPLAN inquiry, “It is possible that teachers are not concerned about high-performing students because they are doing well regardless. Their energies and efforts are focused on students who are at risk of failure” (Hare, 2013).

Measuring Growth

Growth modelling

Once data is collected, a decision remains about how to model the growth of student achievement over time. Linear modelling fits a straight line to the data points. For multiple points of data, piecewise linear modelling can also be used. Other ways to model achievement data include using quadratic or log models. Shanley (2016) tested nine different growth models on mathematics achievement data collected in the Early Childhood Longitudinal Study over eight waves of data collection. The assessment used was vertically scaled, which allowed achievement at each point to be directly compared. The authors found that Model 9 had the best fit of the data – this included a quadratic component between Kindergarten and Grade 3, a linear component between Grades 3 and 5 and a linear component between Grades 5 and 8. Other models the authors found to have a good fit included Model 4 (quadratic model, including summer lag) and Model 6 (linear component between Kindergarten and Grade 1, and quadratic component between Grades 1 and 8). Other researchers have found that a curvilinear modelling with decelerating growth fits mathematics growth best (Stevens, Schulte, Elliott, Nese, & Tindal, 2015) or a transformed log model (Shin et al., 2013).

The emphasis of the REAP project was not on how to measure differential growth patterns but on how to avoid them. The method of measurement is described in the following section.

Method

Participants

This report presents the findings associated with data captured during the 2016 school year, where 43 schools (primary, secondary or P–12) participated in the REAP project. Numbers of participating students in each grade who met the following requirements are shown in Table 1. The total number of students included was 1,685.

Student inclusion was dependent on:

1. Completion of an appropriately targeted mathematics pre-test at the end of Term 1 or beginning of Term 2 (March/April).

2. Completion of an appropriately targeted mathematics post-test at the end of Term 3 or beginning of Term 4 (September/October).

3. Sufficient data to link students' mathematical performance with their respective mathematics teacher. This was not possible in some circumstances where students were moved into a different class for mathematics instruction and the other teacher was not a participant of the REAP project.

Measurements

Content tests

An appropriately designed assessment system was provided by the Assessment Research Centre at the University of Melbourne (Australia). Students' content ability in mathematics or reading comprehension was tested using ARCOTS. The assessments were delivered online, together with an integrated reporting system. Teachers were tasked with the role of targeting the appropriate test to each student in their class. Teachers selected from a series of eight tests, which were colour coded in order of difficulty and complexity of the material and questions: red, orange, yellow, lime, green, aqua, blue and purple, with red being the easiest. Teachers were given access to view each test prior to allocating tests to students. Teachers were expected to use a variety of test colours per class for the same year level. As an example, a Year 6 teacher may have needed to administer the orange test to some students and the yellow, lime or green test to others. Teachers received PD through a series of online guides informing how to target and administer the tests. The online guides explained how to receive the most accurate report for each student; teachers aimed to administer a test where the student could answer approximately 50% of questions correctly, maximising test information for the student.

The tests in each learning area varied in content and complexity. In reading comprehension, for example, the content was the written passage, and the items based on the passage differed in complexity of the skills assessed. For example, the same passage may have had one associated question requiring students to locate information directly stated in the passage and another question requiring students to identify possible reasons for a character's motivation (a more complex skill). These are questions of different complexity on the same content. Each test had questions drawing on a range of content with varying levels of complexity. There was overlap in both content and complexity between one test and the next. This allowed the psychometric scaling of the set of tests to place students on the same (logit) scale regardless of the coloured test they sat. Using this method of test targeting and equating, students' ability estimates could be calculated accurately across grade levels.

Reports were available for students who received an assessment measure that was within the acceptable range of appropriate targeting (generally around 30–80% of questions answered correctly). As the student reached either end of the usable proportion of the assessment, it was considered that the testing errors for the student were too high to accurately determine the student's capacity. Students who answered fewer than 30% of questions correctly were administered an easier test, and students who answered more than 80% correctly completed a harder test. Teachers were given the option of re-testing these students with a more appropriately targeted test. Only appropriately targeted test scores were used in the analysis for this study in order to keep testing errors low and increase the validity of the results presented.

Student results were not reported to teachers as a grade or score but as a written description of the skill level at which the students were ready to learn (i.e., their ZPD). The written descriptions were presented as a progression of skills from low (Level A) to high (Level M for reading comprehension and Level L in mathematics). Progressions were not based on what should have been taught at any given grade level; instead, they were derived from the Centre's research on how students learn and from large empirical data sets obtained from hundreds of thousands of students (M. Pavlovic, personal correspondence, 2017). Progression or ZPD levels were not used in this study to investigate student ability levels; rather, their weighted likelihood estimates (WLE), calculated from data provided by ARCOTS based on Rasch analysis (Warm, 1989) were used to model with teacher covariates.

Teachers' strategies

Teaching strategies were collected via a mixed methods approach, where teachers responded to a short questionnaire after completing each of the eight PD modules to gain access to the following module. Items in the questionnaires were either multiple-choice, short or long answer. Because of the rich and varied classroom approaches to teaching mathematics, and limitations of the data collected, responses were collated and composite variables were created.

As observations of teachers were not performed, results were limited to self-reported measures of strategies that teachers believed they were implementing. For example, responses to multiple-choice items such as "Which of the following strategies are you using for your high capacity students? (You may select more than one response if required)" were limited by the following:

a) Honesty in reporting. Teachers may not have been using the strategy/s they had listed.

b) Understanding of strategies. Teachers who responded that they were "providing more choice in terms of student learning" may have had different understandings of the meaning of "more choice" (although definitions were provided in the PD modules).

c) Implementation differences. Teachers likely had differing levels of implementation of strategies.

Responses were generally not limited to the options presented; in the above example, there was a category for "Are there any main strategies that you are using that are not on this list?" with an open text box provided so teachers could list additional strategy/s.

The creation of composite variables allowed researchers to combine responses to the multiple-choice questions, such as in the preceding example, with qualitative open responses to questions like "What actions have you taken to address the issue of the flatline of the high capacity students?". This gave an overall judgement of which strategies teachers were using so they could be accurately coded.

Strategies that did not relate to positive growth in mathematics, when strategies were modelled separately to gauge effectiveness, are not included in this report. There is no guarantee that a teaching strategy that was not found to promote growth in this study does not in fact promote student growth. The limitations of coding the teacher responses also present an issue with the reported "effect sizes" of each strategy. The rate at which each strategy affects growth, or the value of the Level 2 predictor, should be evaluated with caution. It is possible that in a different study, the order or magnitude of the effect of each strategy might change.

Modelling

Multi-level modelling was completed using MLwiN software. The models account for the nested structure of students within classes, with students modelled on Level 1 and teachers on Level 2 as shown below. Student growth as a variable in itself is not included in the model; rather, the model aims to predict students' post-test (T2), while controlling for their pre-test (T1), and estimating the value of the teacher covariate (teacher strategy).

There were five models used to describe the effect of various teaching strategies and other categorical variables on student growth. All models involve predicting students' T2 scores.

Model 1:	Unconditional model:	Variance at level 1 and 2 decomposed, ICC's generated.
Model 2:	Student T1 model:	T1 WLEs are used to predict T2.
Model 3:	Cohort predictors:	Added to Model 2 are school type, school grade and student quartile (based on class rank) covariates (level 2).
Model 4:	Teaching strategy predictors:	Added to Model 2 are the teaching strategy covariates (level 2), modelled one by one as in equation (3). Each covariate was modelled separately.
Model 5:	Full Model teaching predictors:	Added to Model 2 are the full-set of teaching strategy covariates (level 2), modelled together as in equation (3).

Modelling students' T2 while controlling for T1 allows the impact of teaching strategies to be measured independently of this relationship. The equations used for the modelling are equation (1) for level 1 and equation (2) for level 2. The equations together can be expressed as in (3) fully combined in (4). The 'combined equation' presented below is a representation of how the data will appear in the results table, with each term representing a column in the tables.

Modelling equations

General equation

Level 1 (Student): $T2\ Wle\ estimate_{ij} = \beta_{0j} + \beta_{1j}.T1\ Wle\ estimate_{ij} + E_{ij}$ (1)

for students i = 1,……,n

Level 2 (Class): $\beta_{0j} = \gamma_{00} + \sum_{k=1}^{K} r_{ok}\, Teaching\ Strategy_{kj} + \mu_{00}$ (2)

$\beta_{1j} = \gamma_{10}$

for classes j = 1,……..,J

Teaching strategy denoted by X

Type of strategy is k where K is maximum number of teaching strategies

Which can be expressed as (3)

$$T2\ Wle\ estimate_{ij} = \gamma_{00} + \sum_{k=1}^{K} r_{ok}\, X_{kj} + \mu_{00} + \gamma_{10}.T1\ Wle\ estimate_{ij} + E_{ij}$$

and, $T2\ Wle\ estimate_{ij} = \gamma_{00} + \gamma_{10}.T1\ Wle\ estimate_{ij} + \sum_{k=1}^{K} r_{ok}\, X_{kj} + E_{ij} + \mu_{00}$

If: $\gamma_{00} = \beta_0$

$\gamma_{10} = \beta_1$

$\gamma_{0k} = \beta_k$

$e_{ij} = E_{ij} + \mu_{00}$

This gives rise to the *Combined equation*

$T2\ Wle\ estimate_{ij} = \beta_0 + \beta_1 T1\ Wle\ estimate_{ij} + \sum_{k=1}^{K} \beta_k\, X_{kj} + e_{ij}$ (4)

$e_{ij} \sim N(0, \sigma_e^2)$

Effect sizes have been calculated as proportion reduction in variance (PRV), as outlined by Peugh (2010) where: PRV = $(Variance_{NoPredictor} - Variance_{Predictor}) \div Variance_{NoPredictor}$

While Peugh acknowledges that "Effect sizes in MLM analyses are not as straightforward [as in ANOVA and multiple regression analysis] and currently no consensus exists as to the effect sizes that are most appropriate" and "The MLM effect sizes ……are not comparable in the same sense as a *d* or η^2" pg. 97. The effect sizes for each separate teaching strategy have been included to represent the magnitude of change determined in this study based on the use of particular teaching strategies. Importantly the magnitude of the effect of using each strategy is affected by the quality of implementation and this was not quantified in this study. Therefore, the order of effect sizes should be taken with caution and teachers are recommended to use the strategies that fit their practice as is explained in the teacher recommendations section.

Results

Student Results

Mean student WLE estimates were calculated for each grade and compared. Results are presented separately for all students (Table 1) and high capacity students (Table 2). The mean student differences between the pre-test (T1) and the post-test (T2) are also shown. There was decreased growth for students in Grades 7 and 8 compared with students in Grades 5 and 6, before T1 was controlled for by modelling. The mean growths of high capacity students from each grade level (Table 2) were lower than the comparative mean growths of each grade level (Table 1).

Table 1. Mean Time 1 (T1) and Time 2 (T2) mathematical ability estimates and differences: All students

Grade Level		T1 WLE Estimate	T2 WLE Estimate	Mean Difference
5	Mean (std. dev) N	2.371 (0.927) 551	2.811 (1.074) 551	+ 0.439
6	Mean (std. dev) N	2.764 (0.998) 475	3.185 (1.074) 475	+ 0.421
7	Mean (std. dev) N	2.906 (1.036) 360	3.284 (1.158) 360	+ 0.378
8	Mean (std. dev) N	3.452 (0.997) 299	3.826 (1.192) 299	+ 0.374
All	Mean (std. dev) N	2.788 (1.051) 1685	3.198 (1.160) 1685	+ 0.410

Note: Standard deviation in parenthesis.

Table 2. Mean Time 1 (T1) and Time 2 (T2) mathematical ability estimates and differences: High capacity students

Grade Level		T1 WLE Estimate	T2 WLE Estimate	Mean Difference
5	Mean (std. dev) N	3.376 (0.568) 113	3.734 (0.848) 113	+ 0.357
6	Mean (std. dev) N	3.676 (0.703) 134	3.977 (0.833) 134	+ 0.301
7	Mean (std. dev) N	3.868 (0.767) 89	4.082 (1.021) 89	+ 0.215
8	Mean (std. dev) N	4.378 (0.653) 77	4.652 (0.928) 77	+ 0.275
All	Mean (std. dev) N	3.766 (0.754) 413	4.059 (0.949) 413	+ 0.293

Note: Standard deviation in parenthesis.

It is noted that the mean growth of students in Grade 8 was similar to that of the high capacity students in Grade 5, and the starting mean WLEs (T1) were also similar (Tables 1 and 2), inspiring the graphical representation of the relationship between the mean T1 WLE per group and the growth of each group (Figure 7). The line of best fit is based on the relationship between the means of each grade's T1 and growth, with all students represented by the blue dots. High capacity student growth means (red dots) are overlayed and do not form the basis of the line of best fit.

If the relationship between T1 and growth were completely linear, it could be said that either (a) the only differential growth pattern existing in the data is based on a cognitive-based pattern (the higher the initial score the least likely a student is to grow), (b) the teaching-based differential growth pattern is equal across all grades, where teachers at Grades 5, 6, 7 and 8 are equally unable to target for high capacity students or (c) both differential growth patterns exist in the data.

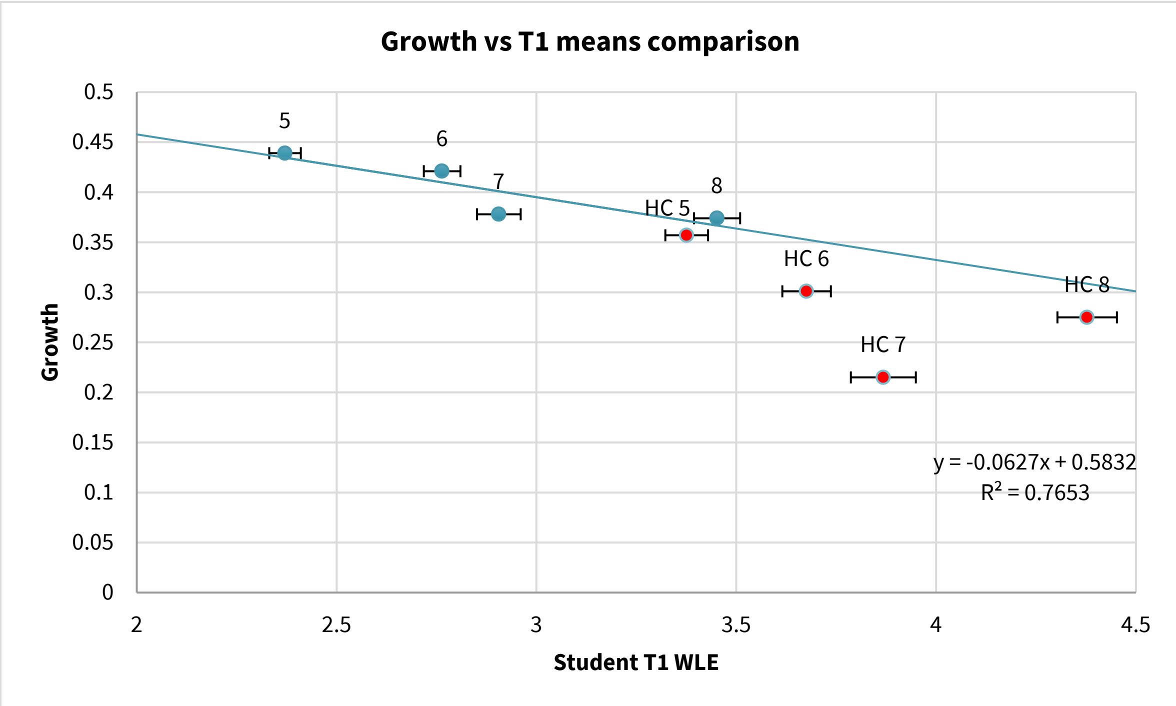

Figure 7. Graphical representation of mean students' mathematics T1 vs growth for specific groups of students.

Notes: HC refers to high capacity students. Linear trendline based on mean T1 WLEs for Grades 5, 6, 7 and 8 "All students" trendline extrapolated for remainder of chart. Error bars are standard errors, based on standard deviations and sample sizes for T1 means.

The high capacity students in Grade 7 appear to be the group most disadvantaged by teaching practices. These students' mean growth is farthest below the linear line, which represents the general relationship between growth and T1 (all students). While all groups of high capacity students are below the best fit line for growth, high capacity Grade 5 and high capacity Grade 8 students almost reach the line. It should be noted that there is no particular reason to assume the relationship between T1 and growth is linear; however, given the data produced by this study, it does appear to be. The strong relationship between T1 and growth implies that some of the differential growth pattern seen in high capacity students can be explained by a cognitive- or testing-based flatline rather than a teaching one (refer to cognitive- and testing-based differential growth pattern explanations in the Introduction). However, the reduced growth in high capacity Grade 7 students, and perhaps Grade 6 students, is more likely to be symptomatic of either a transition-based or a teaching-based differential growth pattern.

Relationship between Covariates and Student Growth

The relationship between student (T1) and class (school type, school year, quartile and teacher strategy) covariates is shown in Table 3 (all students) and Table 4 (high capacity students).

While the entire set of students (n = 1,685) is modelled for most of the analyses, a separate set of models is shown for the sub-set of students (n = 1,130) whose teachers responded to certain questions crucial to categorising the teacher's strategies. Results from both sets of models are compared with the appropriate -2*loglikelihood value to determine the significance of the model.

For the full cohort of students, with every 1 standard deviation (1.051 logits – Table 1) increase in T1 WLE estimate, students' T2 WLE estimate was predicted to increase by 0.944 logits. Students' T1 WLE estimate was a significant predictor of students' T2 WLE estimate, $z = 67.43$, $p < 0.00001$ (Table 3).

For high capacity students tested, with every 1 standard deviation (0.754 logits) increase in T1 WLE estimate, students' T2 WLE estimate was predicted to increase by 0.969 logits (Table 4).

School type and grade were not significant predictors of students' T2 (Table 3) beyond what students' T1 score predicted, except for high capacity Grade 7 students (Table 4) who were significantly disadvantaged in terms of achieving growth (also reflected in Figure 7). The quartile the students belonged to (Q1 = low ability, Q2 = low/middle ability, Q3 = middle/high ability and Q4 = top 25%, high capacity students) was a negative predictor of T2 score over and above that which the T1 score predicted (as was shown before modelling in Figure 7).

Teacher strategies that predicted a greater T2 score (after controlling for T1) are listed in Table 5, with the effect size of each strategy and significance noted for Model 4. In Model 5, using an online curriculum was identified as being the most contributory teaching strategy to gaining student growth for both 'all students' and 'high capacity students' (Table 4).

Table 3. Predictors of students' Time 2 test: All students

Parameter	Intercept β_0	T1 Student Predictor β_1	Level 2 Predictor β_k	$E_{ij} \sim N(0,\sigma_e^2)\ \sigma_e^2$	-2*loglikelihood n = 1685 (Δ,df)	-2*loglikelihood n = 1130 (Δ,df)
Model 1						
Unconditional (n = 1685)	3.198(0.028)	-	-	1.345(0.046)	5281.698	
Unconditional (n = 1130)	3.213(0.036)	-	-	1.452(0.061)		3628.077
Model 2						
Student T1 (n = 1685)	0.565(0.042)	**0.944(0.014)*** **	-	0.361(0.102)	3065.821(2216,1)***	
Student T1 (n = 1130)	0.556(0.049)	**0.956(0.016)*** **	-	0.360(0.015)		**2051.024(1577,1)*** **
Model 3: Cohort predictors						
School Type (Primary)	0.565(0.042)	**0.947(0.015)*** **		0.361(0.012)	3065.287(+0.53,1)^{ns}	
Secondary			-0.023(0.031)^{ns}			
School Year (Grade 5)	0.568(0.044)	**0.946(0.015)*** **		0.361(0.012)	3064.994(0.83,3)^{ns}	
Grade 6			0.003(0.038)^{ns}			
Year 7			-0.032(0.041)^{ns}			
Year 8			-0.007(0.046)^{ns}			
Quartile (Q1)	0.549(0.045)	**1.006(0.018)*** **		0.356(0.012)	**3039.193(26.6,3)*** **	
Q2			**-0.138(0.043)*** **			
Q3			**-0.195(0.047)*** **			
Q4			**-0.278(0.055)*** **			
Model 4: Teaching predictors						
Online Curriculum	0.534(0.049)	**0.959(0.016)*** **	**0.179(0.066)** **	0.357(0.015)		**2043.618(-7.41,1)** **
Curriculum Extension	0.530(0.050)	**0.958(0.016)*** **	**0.089(0.042)* **	0.358(0.015)		**2046.626(-4.40,1)* **
Targeted Interventions	0.460(0.062)	**0.958(0.016)*** **	**0.050(0.021)** **	0.358(0.015)		**2045.017(-6.01,1)* **
Like-Ability Peers	0.520(0.043)	**0.942(0.014)*** **	**0.101(0.029)*** **	0.359(0.012)	**3054.036(11.8,1)*** **	
Appropriate Goals	0.528(0.043)	**0.944(0.014)*** **	**0.102(0.030)*** **	0.359(0.012)	**3054.576(11.2,1)*** **	
Students are Challenged	0.539(0.043)	**0.942(0.014)*** **	**0.080(0.030)** **	0.360(0.012)	**3058.750(7.07,1)** **	
Model 5: Full Model						
Teaching predictors	0.424(0.065)	**0.957(0.016)*** **		0.354(0.015)		**2032.161(18.86,1)*** **
Online Curriculum			**0.146(0.069)* **			
Curriculum Extension			0.065(0.046)^{ns}			
Targeted Interventions			0.007(0.026)^{ns}			
Like-Ability Peers			0.061(0.048)^{ns}			
Appropriate Goals			0.064(0.040)^{ns}			
Students are challenged			0.018(0.045)^{ns}			

Notes. *Estimates unstandardised beta values (β); unstandardised errors are in parentheses; statistically significant results emboldened and underlined; ns = not statistically significant, *p < .05, **p < .01, ***p < .001. In Model 3, cohort predictors are in comparison to reference category in parentheses.*

Table 4. Predictors of students' Time 2 test: High capacity students

Parameter	Intercept β_0	T1 Student Predictor β_1	Level 2 Predictor β_k	$E_{ij} \sim N(0, \sigma_e^2)$ c	-2*loglikelihood n = 413 (Δ,df)	-2*loglikelihood n = 279 (Δ,df)
Model 1						
Unconditional (n = 413)	4.059(0.047)	-	-	0.898(0.062)	1127.381	
Unconditional (n = 279)	4.124(0.057)	-	-	0.913(0.077)		766.434
Model 2						
Student T1 (n = 413)	0.411(0.152)	**0.969(0.040)*** **	-	0.366(0.025)	**756.961(370,1)*** **	
Student T1 (n = 279)	0.475(0.176)	**0.966(0.046)*** **	-	0.350(0.030)		**499.137(267,1)*** **
Model 3: Cohort predictors						
School Type (Primary)	0.372(0.155)	**0.987(0.042)*** **		0.365(0.025)	755.582(1.38,1)^{ns}	
Secondary			-0.077(0.065)			
School Year (Grade 5)	0.409(0.160)	**0.985(0.044)*** **		0.364(0.025)	754.599(2.36,3)^{ns}	
Grade 6			-0.052(0.078)			
Year 7			**-0.135(0.088)***			
Year 8			-0.067(0.100)			
Model 4: Teaching predictors						
Online Curriculum	0.364(0.174)	**0.985(0.045)*** **	**0.484(0.129)*** **	0.334(0.028)		**485.455(13.68,1)*** **
Curriculum Extension	0.386(0.177)	**0.976(0.045)*** **	**0.219(0.083)** **	0.342(0.029)		**492.248(6.89,1)** **
Targeted Interventions	0.311(0.196)	**0.974(0.046)*** **	**0.075(0.040)***	0.346(0.029)		495.695(3.44,1)^{ns}
Like-Ability Peers	0.374(0.152)	**0.960(0.039)*** **	**0.134(0.059)***	0.362(0.025)	**751.909(5.05,1)***	
Appropriate Goals(34c)	0.340(0.153)	**0.971(0.039)*** **	**0.172(0.061)** **	0.359(0.025)	**749.156(7.80,1)** **	
Students are Challenged (34a)	0.389(0.152)	**0.964(0.040)*** **	**0.099(0.061)***	0.364(0.025)	754.319(2.64,1)^{ns}	
Model 5: Full Model						
Teaching predictors	0.233(0.192)	**0.996(0.045) *** **		0.325(0.028)		**478.307(20.83,6)** **
Online Curriculum			**0.456(0.135)*** **			
Curriculum Extension			**0.150(0.088)***			
Targeted Interventions			0.005(0.049)			
Like-Ability Peers			-0.030(0.093)			
Appropriate Goals			0.114(0.078)			
Students are challenged			0.015(0.086)			

Notes. *Estimates unstandardised beta values (β); unstandardised errors are in parentheses; statistically significant results emboldened and underlined; ns = not statistically significant, *p < .05, **p < .01, ***p < .001. In Model 3, cohort predictors are in comparison to reference category in parentheses.*

All strategies that were effective for high capacity students (Table 3) were also effective for all students (Table 4). This lack of difference indicates that the strategies that support high capacity students are simply good teaching practices that extend all students based on their individual level of need (ZPD). Of importance, strategies involving appropriate curriculum based on learning needs (online or other) had a greater effect on growth for high capacity students compared to all students. This result defines the reason that high capacity students are not growing to the same extent as other students in Australian classrooms: their learning needs based on their ZPD are largely not currently being met. Results are summarised in table 5, where effect sizes as PRV are also shown.

Table 5. Summary of teaching practices related to growth

Teaching Strategy	% Using Practice	Growth for All	Growth for HC
Students use online self-paced curriculum resources (online curriculum programs)[#]	8 %	0.179(0.066)** 7.7% variance explained	0.484(0.129)*** 36.9 % variance explained
Use of self- or teacher-paced curriculum extension activities/tasks/programs (curriculum extension)	23 %	0.089(0.042)* 3.8 % variance explained	0.219(0.083)** 17.4 % variance explained
Teacher acted to specifically provide targeted strategies for HC students. Includes: • using assessment evidence to select interventions for individual students; • modifying content of teaching and learning activities for high capacity students; • providing more choice for high capacity students; • other specific strategies designed to support the growth of high capacity students.	0 – 8% 1 – 27 % 2 – 45 % 3 – 20 %	0.050(0.021)** 3.8 % variance explained	0.075(0.040)* 8.7 % variance explained
Teacher provided opportunities for students to work with others at their same level (like-ability interactions)	51 %	0.101(0.029)*** 7.7 % variance explained	0.134(0.059)* 12.9 % variance explained
Appropriate goals for students – Teacher opinion	36 %	0.102(0.030)*** 7.7 % variance explained	0.172(0.061)** 22.5 % variance explained
Students are challenged – Teacher opinion	40 %	0.080(0.030)** 3.8 % variance explained	0.099(0.061)* 6.45 % variance explained

[#]Note: No secondary teachers in the study reported use of online curriculum as a teaching strategy; *$p < .05$, **$p < .01$, ***$p < .001$

The PVC for the full model for all students (all teaching strategies modelled together, model 5) was 0.23. Therefore, the teaching strategies examined make up 23% of the variance which explains the students' results. This indicates that there are other classroom level predictors that have not been accounted for in this study, which make up the remaining 77 % of the variance explained by the teacher.

The PVC for the HC student sample reached 56%, with 44% of classroom level effects unaccounted for. The teacher covariates explained more of the variance for the HC student cohort as was expected with strategies intended to cater for HC students modelled.

Variance between and within classes

To determine the possible overall impact of classroom factors (teaching strategies) on students' growth, variation within and between classes were identified. Variation is caused by within-class variance; therefore, slopes were fixed to interpret the effect of each covariant on within-class variation (this is relevant as the teacher covariates vary for each class [Level 2]). For all students, 92% of the variance in student growth was caused by student differences, whereas 8.3% of the variance in growth was attributed to class-level effects. For high capacity students, 91.5% of the variance in student growth was caused by student differences, whereas 8.5% of the variance in growth was attributed to class-level effects. Individual student differences in learning accounted for most of the variation in student growth per class, but the teacher still had a major impact with over 8% of the variance in student growth attributed to classroom effects, most notably by the teacher's practices.

All students

Within-class variance:

$e_{ij} \sim N(0, \sigma_e^2)$ σ_e^2 = 0.333(0.012)

Between-class variance:

$\mu_{ij} \sim N\left(0, \sigma_{\mu 0}^2\right)$ $\sigma_{\mu 0}^2$ = 0.030(0.007)

Variance partition coefficient (VPC) = 0.030/ (0.030+0.333) = 0.0826 or 8.3% of the variance in growth was attributed to class-level effects.

92% of variance in growth was caused by student differences within class.

High capacity students

Within-class variance:

$e_{ij} \sim N(0, \sigma_e^2)$ σ_e^2 = 0.335(0.026)

Between-class variance:

$\mu_{ij} \sim N\left(0, \sigma_{\mu 0}^2\right)$ $\sigma_{\mu 0}^2$ = 0.031(0.016)

VPC = 0.031/ (0.031+0.335) = 0.0846 or 8.5% of the variance was attributed to class-level effects.

91.5% of variance was caused by student differences within class.

Teacher Recommendations

The information presented here is not a "one size fits all" approach to teaching; instead, it recommends that teachers choose one or two practices at a time to trial. The practice must fit the personality, philosophy and circumstances of the teacher. For example, the third practice (curriculum extension) may be difficult to implement in a school that has limited access to curriculum resources beyond the textbook.

Targeted Strategies

We define targeted strategies as "teachers specifically acting to increase a student's achievement, based on assessment". Targeted teaching strategies resulted in growth for all students, not just for high capacity students (high capacity students are defined as the top 25% of students in a particular cohort). We also found that the more targeted strategies the teachers reported implementing for high capacity students, the higher students' growth.

In the REAP project, targeted strategies were evidenced by

- using assessment evidence to select interventions for individual students
- modifying content of teaching and learning activities for high capacity students
- providing more choice for high capacity students
- other specific strategies designed to support the growth of high capacity students.

To be able to select appropriate strategies and learning activities, teachers need to know what their students can do and what they are ready to learn next. Teachers can use regular formative assessment to help identify the point of need for each student (the student's ZPD). Using this ongoing assessment, teachers can then manipulate the learning environment and scaffold learning for every student, regardless of the student's development or intellectual capacity (Griffin, 2007). In other words, teachers target their teaching and learning activities based on students' point of readiness to learn. Similarly, Tomlinson (1997) suggested that teaching for highly abled students should be paced in response to students' individual needs.

Evidence, in the form of ongoing assessment, is then used to monitor the effectiveness of the targeted strategies. If assessment evidence reveals that a particular strategy is not working, teachers may then need to change their approach. It can be helpful to talk to colleagues (at and above the student's year level) about what strategies have worked for them, as well as reading education research on best practice instructional strategies (e.g., Hattie, 2009; Petty, 2014. Table 6 shows how teachers in this study used targeted strategies to support the growth of high capacity students.

Table 6. Sample teacher descriptions of targeted interventions used

Type of Targeted Intervention	Sample Teacher Responses for Supporting the Growth of High Capacity Students
Use evidence (assessment) of learning to select interventions for individual students	"Targeted learning, regular testing with regular feedback and specific goal setting". "Understand who they [high capacity students] are, track their progress, improve self-regulated learning skills, ensure work is appropriate/engaging/challenging". "Differentiate the curriculum and encourage them [high capacity students] to participate in other activities that may support their learning that is offered in the school or programs they could work on at home". "Identify the high capacity students, identify the current interventions teachers are using and assess their effectiveness, increase professional development for teachers in catering to high capacity students".
Modify content for high capacity students	"Creating challenging extension problems. Grouping [high capacity] students together. Increasing the amount of difficult problems that require students to do research and think in other ways". "Detailed lesson planning, scope for problem solving and collaborative group work". "High capacity students were given access to an adaptive learning program to work through content knowledge from the year level above them (and beyond)". "Differentiation of the curriculum to provide more challenging content and learning tasks".
Provide choice for high capacity students	"Allowing students to inquire/self-direct their learning by selecting what they want to know more about and further developing skills by further inquiring into concepts of interest". "[I] tried to change the way I give content information and skill development. I am trying to increase student voice and agency in my class and allowing some choice in areas where that is possible". "Students have been given dedicated time, during the year, to work on Number and Algebra goals of interest to them". "Use of open ended, real-world problems. Getting students to create their own real-life problems".

Setting Appropriate Learning Goals

Our data shows that setting learning goals based on students' capacity was related to growth in mathematics. Goal setting is the process of establishing an outcome (a goal) to serve as the aim of one's actions (Turkay, 2014). Setting goals makes the direction of learning clear to the student and the teacher.

Research suggests that setting learning goals increases students' motivation and achievement levels. Students achieve more if they are given specific learning goals to achieve, rather than simply being told to try their best (Locke & Latham, 2002).

Setting *appropriate* learning goals consists of several components. First, teachers need to use assessment to find out the student's skill level and what they can do well (their ZPD). Next, the student's achievement should be placed on a learning progression. Then, a learning goal is set that is just out of reach for the student. In other words, learning goals should be set based on the student's capacity and point of need.

Students can also set and document their own goals. When students set their own goals, they have more ownership and take more responsibility for progress towards their goals (Turkay, 2014). High capacity students in particular may relish the opportunity to set their own goal in mathematics, beyond the assigned curriculum level.

Curriculum Extension Resources

Results from the project show that teachers who used curriculum extension resources had higher student growth in mathematics. Curriculum extension resources refer to any resource that provides students with access to more advanced parts of the curriculum. Curriculum extension resources may include

- mathematics textbooks (at or above the student's grade level);
- other instructional materials such as workbooks, videos or games;
- mathematics problems designed specifically for high performing students;
- mathematics problem-solving questions, for example, from past mathematics competition papers.[1]

Students need access to extension material to be able to grow academically. After assessment of student skill levels, teachers can choose extension material based on what skills students have mastered and what comes next in the curriculum.

When using curriculum resources to extend high capacity students, it may be helpful to consider the following:

- Do I want students to move through the curriculum at a faster pace?

[1] For example, the University of Melbourne School Mathematics Competition (www.mathscomp.ms.unimelb.edu.au/)

- Do I want students to minimise time spent on work already mastered?
- Do I want to provide students with greater breadth of work?
- Do I want to provide students with more depth of work?
- Do I want to provide access to advanced work beyond the student's chronological age or grade level?

Our data also shows that primary teachers reported using *online* curriculum resources to extend their students, which also resulted in student growth. We refer to online curriculum extension as providing opportunities for students to access advanced parts of the curriculum via online resources. This finding was seen in overall class growth levels, as well as growth of high capacity students. It should be noted that secondary school teachers did not explicitly report using online curriculum resources to extend their students.

With online resources now easily accessible to most teachers via the internet, educators can find additional challenging mathematics resources beyond those provided in a textbook or workbook. Online tools are often used when differentiating work for students at different levels of achievement (Park & Datnow, 2017).

Naturally, teachers will need to spend some time familiarising themselves with the online curriculum resource before including it in their lessons. After investigating the online tool's ability to explain complex concepts, the resource could be used to extend and reinforce the curriculum.

Online curriculum resources should not be a replacement for direct teaching. Instead, online resources can be drawn on when the teacher has assessed what students can do and has identified what students are ready to learn next (the student's ZPD). When formative assessment of students is complete, and appropriate goals have been established, online resources can help teachers plan and provide academic intervention for students. For example, teachers could use the Maths300 website (www.maths300.com) for students who need to work on mathematics proficiencies (understanding, fluency, reasoning, modelling and problem-solving), ensuring that the task selected matches the ZPD of the student. Online curriculum resources can also be used to consolidate mathematics knowledge, for example through extra practice, and can increase engagement of students. Another benefit of using online mathematics platforms is that students may be able to interact with students at the same ability level or higher ability level. The importance of like-ability interaction is covered in further detail in the following section.

Like-Ability Interactions

We examined the mathematics growth of students in classes that provided for students to work in groups, pairings or environments with students of a similar or higher ability. We found that teachers who reported providing like-ability interactions for students had, on average, higher student growth compared to those teachers who did not report using this strategy. Providing like-ability interactions resulted in an even larger student growth when we analysed the growth of high capacity students only. In other words, having access to peers of the same or greater ability is important for increasing the growth of all students and even more so for high capacity students.

High capacity students need interaction with peers of a similar or more advanced ability or achievement level. Interaction with like or more advanced students increases engagement, effort and satisfaction in the activity. High capacity students are one of approximately five/six students in a class and may have limited access to like-ability peers, let alone peers of a higher ability. High capacity students may go to the teacher for this type of interaction, but it is often the case that the classroom teacher is busy helping other students.

Gross (2006) suggested that high achieving students prefer the company of peers at similar stages of intellectual and emotional development. Furthermore, Gross suggested that behavioural and emotional problems can result if high capacity students do not have access to similar ability peers. In mixed-ability classrooms, high capacity students may hide their capacity to be accepted by their classmates, particularly students who may be at the very top of the class. In fact, in a review of the available research, Rogers (1998) stated that low *and* high ability students benefit from more social interactions when they are grouped in a class with like-ability peers.

Some ideas for promoting like-ability interactions for high capacity students include

- grouping students of similar ability or achievement levels within classes;
- allowing high capacity students to interact with other like-ability peers from other classes in the same year;
- providing opportunities for high capacity students to interact at lunchtime or afterschool (i.e. a mathematics “club”);
- allowing high capacity students to visit other year levels for certain topics or assignments;
- arranging (face-to-face or online) interaction with other high capacity students from other schools;
- promoting external events and activities where students can meet like-ability students from other schools (e.g. MAV Mathematics Games Days: www.mav.vic.edu.au/student-activities/student-games-days.html).

A word of caution about grouping students

Grouping students by “streaming” students in terms of their ability has long been considered a controversial practice in schools (Gamoran, 2011). Some of the controversy is based on the practice of taking high ability students and streaming them into a separate class, leaving the remaining classes homogeneous in terms of ability. Some researchers have argued that ability grouping impacts students' self-esteem or creates a "self-fulfilling prophecy"; that is, students in a low group only achieve to a low level as they are perceived to perform to such a level (Clarke, 2003). On the other hand, students in high ability groups may feel overwhelming pressure to perform and struggle with the advanced pace of the curriculum (Boaler, Wiliam, & Brown, 2000).

A recent review by Steenbergen-Hu, Makel, and Olszewski-Kubilius (2016) suggested that within-class grouping (groupings within a class), cross-grade subject grouping (grouping students across a year level to learn a specific subject) and special groupings for the gifted *does* benefit student achievement. They found limited evidence for the use of between-class ability grouping (grouping students across a year level based on prior achievement or ability levels, often referred to as streaming).

Rogers (2001) argued that grouping is helpful for gifted students, and that there are many common misconceptions or "myths" about the effects of grouping based on ability, such as "It removes the role-models at-risk students need to succeed and behave" (p. 76).

In the current study, all but two teachers reported using grouping in some way. The majority of teachers reported using groups based on ability and other times grouping for interest or friendship. Teachers also reported grouping students across classes (with some groups taught by other teachers) and grouping within classes (with all groups taught by the same teacher). This indicates that, at least in the REAP project, teachers are using grouping in some form in their practice.

Our advice is that *student groupings should be flexible and adaptable.* Student grouping is often used as part of differentiated instruction, but it is not fixed or based on ability; rather, the groups are used to support individual needs of students in terms of skill or achievement level (Park & Datnow, 2017).

One reason for using flexible grouping is that student abilities can differ within different topics. For example, a student may demonstrate advanced ability in algebra and benefit from being grouped with others of similar high ability, but that same student may struggle with chance and data and have their needs met in a mixed-ability class. Additionally, flexible groupings allow students to move between groups and do not categorise a student in the low group for the whole year and beyond; this is one of the main criticisms of ability grouping (Clarke, 2003).

Summary

- Four evidence-based strategies were identified that support the growth of high capacity students in mathematics: academic intervention, setting appropriate goals, online curriculum extension, like-ability interactions.
- Increased growth for high capacity students was found in classrooms where teachers reported implementing academic interventions to support high capacity students. The following aspects of academic interventions were discussed: using evidence/assessment to select and evaluate learning interventions, modifying the content of student work, and providing high capacity students with more choice in their learning.
- Setting appropriate goals based on students' capacity was also found to be related to higher student growth in mathematics. Teachers can document goals formally, such as in an individual learning plan. Students can also be involved in creating and monitoring their goals.
- In mathematics, providing opportunities for online curriculum extension was related to higher growth in mathematics.
- Students demonstrated higher growth when provided with opportunities to interact with like-ability or higher ability peers. Flexible and adaptable groupings should be used.

Teacher Reflection

Individually or within your professional learning team, reflect on the information discussed in this report.

1. Of the evidence-based teaching practices discussed, choose one strategy that you could implement in your teaching of mathematics. Make a plan (when, where, how, what resources do you need, etc.) for how you could trial this strategy. Reflect on the success of the strategy after having implemented it for three to four weeks.
2. Reflecting on targeted strategies, do you, or how could you, modify the content of teaching and learning activities for high capacity students in your mathematics classes? Discuss with at least two colleagues how they modify work for high capacity students.
3. Why is it important to have goals for student learning? Do students at your school have access to like-ability peers? Brainstorm ways to increase the number of interactions high capacity students have with similar or higher ability peers.
4. How do you use grouping in mathematics? Evaluate your approach based on the research presented.
5. How did reading this report change or impact your perspective on teaching strategies for high capacity students?

References

Andrich, D., Marais, I., & Humphry. S. M. (2012). Using a theorem by Andersen and the dichotomous Rasch model to assess the presence of random guessing in multiple choice items. *Journal of Educational and Behavioural Statistics, 37*(3), 417–442.

Australian Curriculum, Assessment and Reporting Authority (2017). *NAPLAN achievement in reading, writing, language conventions and numeracy: National report for 2017.* Sydney, Australia: ACARA.

Boaler, J., Wiliam, D., & Brown, M. (2000). Students' experiences of ability grouping: Disaffection, polarisation and the construction of failure. *British Educational Research Journal, 26*(5), 631–649.

Chubb, J. E., & Loveless, T., (2002). *Bridging the achievement gap*. Washington, DC: Brookings Institution Press.

Clarke, D. (2003). *Challenging and engaging students in worthwhile mathematics in the middle years*. Paper presented at the Mathematics Association of Victoria Annual Conference: Making Mathematicians, Melbourne. Retrieved from http://websitewww.mav.vic.edu.au/files/conferences/2003/Clarke_paper.pdf

Gamoran, A. (2011). Designing instruction and grouping students to enhance the learning of all: New hope or false promise? In M. T. Hallinan (Ed.), *Frontiers in sociology of education* (pp. 111–126). Dordrecht, Netherlands: Springer.

Griffin, P. (2007). The comfort of competence and the uncertainty of assessment. *Studies in Educational Evaluation, 33*(1), 87–99. doi:10.1016/j.stueduc.2007.01.007

Griffin, P. (2014). *Assessment for teaching*. Cambridge, England: Cambridge University Press.

Gross, M. U. (2006). To group or not to group – is THAT the question? In C. Smith (Ed.), *Including the gifted and talented: Making inclusion work for more gifted and able learners* (pp. 119–137). London, England: Routledge.

Hattie, J. (2009). *Visible learning: A synthesis of meta-analyses in education*. Oxford, England: Routledge.

Hare, J, (2013, June 25). Study shows top students flatlining. *The Australian*. Retrieved from https://www.theaustralian.com.au/national-affairs/education/study-shows-top-students-flatlining/news-story/70c16336df76a4ed332ed1d74aad3f88?sv=a6f35e8808d981551edeac9289c7240

Hopwood, B., Hay, I., & Dyment, J. (2017). Students' reading achievement during the transition from primary to secondary school. *Australian Journal of Language and Literacy, 40*(1), 2017.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*(9), 705–717. doi:10.1037/0003-066X.57.9.705

Lohman, D., & Korb, K. (2006). Gifted today but not tomorrow? Longitudinal changes in ability and achievement during elementary school. *Journal for the Education of the Gifted, 29*(4), 451–484.

Loveless, T. (2008). High-achieving students in the era of NCLB: Analysis of NAEP data (Part 1). Retrieved from https://files.eric.ed.gov/fulltext/ED501703.pdf. Washington, DC: Thomas E. Fordham Institute.

Ontario Public Service. (2013). *Learning for all: A guide to effective assessment and instruction for all students kindergarten to grade 12*. Ministry of Education in Canada. Retrieved from http://www.edu.gov.on.ca/eng/general/elemsec/speced/LearningforAll2013.pdf

Park, V., & Datnow, A. (2017). Ability grouping and differentiated instruction in an era of data-driven decision making. *American Journal of Education, 123*(2), 281–306. doi:10.1086/689930

Petty, G. (2014). *Evidence based teaching: A practical approach*. Oxford, England: Oxford University Press.

Plucker, J. A., Burroughs, N., & Song, R. (2010). *Mind the (other) gap: The growing excellence gap in K-12 education*. Indiana University, Bloomington, IN: Center for Evaluation Policy.

Rogers, K. B. (1998). Using current research to make "good" decisions about grouping. *NASSP Bulletin, 82*(595), 38–46. doi:10.1177/019263659808259506

Rogers, K. B. (2001). Grouping the gifted: Myths and realities. *Gifted Education Communicator*, *32*(1), 20–25. Retrieved from http://www.thinkingahead.com.au/Documents/Grouping_the_Gifted_Rogers2001.pdf

Shanley, L. (2016). Evaluating longitudinal mathematics achievement growth: Modeling and measurement considerations for assessing academic progress. *Educational Researcher, 45*(6).

Shin, T., Davison, M. L., Long, J. D., Chan, C.-K., & Heistad, D. (2013). Exploring gains in reading and mathematics achievement among regular and exceptional students using growth curve modeling. *Learning and Individual Differences, 23*(Supplement C), 92–100. doi:10.1016/j.lindif.2012.10.002

Song, S., Perry, L., & McConney (2014). Explaining the achievement gap between Indigenous and non-Indigenous students: an analysis of PISA 2009 results for Australia and New Zealand. *Educational Research and Evaluation, 40*(3).

Steenbergen-Hu, S., Makel, M. C., & Olszewski-Kubilius, P. (2016). What one hundred years of research says about the effects of ability grouping and acceleration on K–12 students' academic achievement. *Review of Educational Research, 86*(4), 849–899. doi:10.3102/0034654316675417

Stevens, J. J., Schulte, A. C., Elliott, S. N., Nese, J. F. T., & Tindal, G. (2015). Growth and gaps in mathematics achievement of students with and without disabilities on a statewide achievement test. *Journal of School Psychology, 53*(1), 45-62. doi:10.1016/j.jsp.2014.11.001

Studeman, D. (2007, 4 January). *But I regress* [Web blog]. Retrieved from https://www.fangraphs.com/tht/but-i-regress/

Tomlinson, C. (1997). *The do's and don'ts of instruction: What it means to teach gifted learners well*. Retrieved from http://www.nagc.org/resources-publications/gifted-education-practices/what-it-means-teach-gifted-learners-well?id=659

Topsfield, J. (2013, 10 January). Results flatline for top students. *The Sydney Morning Herald*. Retrieved from https://www.smh.com.au/education/results-flatline-for-top-students-20130109-2cgud.html

Turkay, S. (2014). *Setting goals: Who, why, how?* Retrieved from https://hilt.harvard.edu/files/hilt/files/settinggoals.pdf

Waller, M. (1974, 19 April). *Removing the effects of random guessing from latent trait ability estimates*. Paper presented at the Annual Convention of the American Educational Research Association, Chicago, Illinois.

Warm T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54,* 427–450.