Melbourne Graduate
School of Education
Assessment Research
Centre

Achieving Reading Comprehension 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., Nibali., N., Griffin., P., Graham., L., Arnup., J., English., N., Alom, BM., & Zhang., Z. (2018). *Achieving reading comprehension 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, in which teachers in 58 schools were asked to assess year 5 to 8 students in March/April (T1) and then in September/October (T2) to determine their growth in reading comprehension. 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).

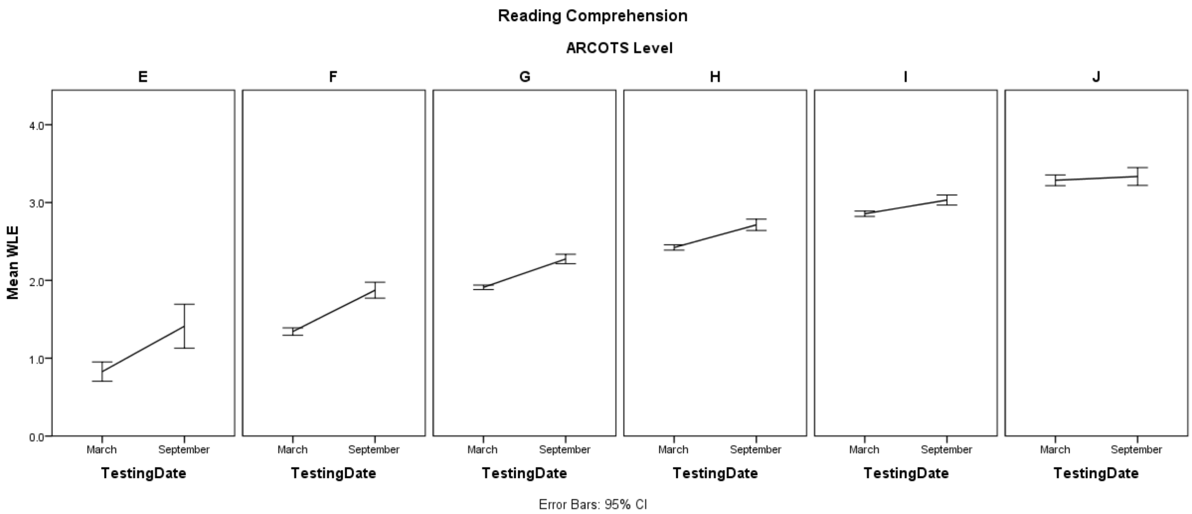

Figure 1. Cognitive-based differential growth (REAP data).

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).

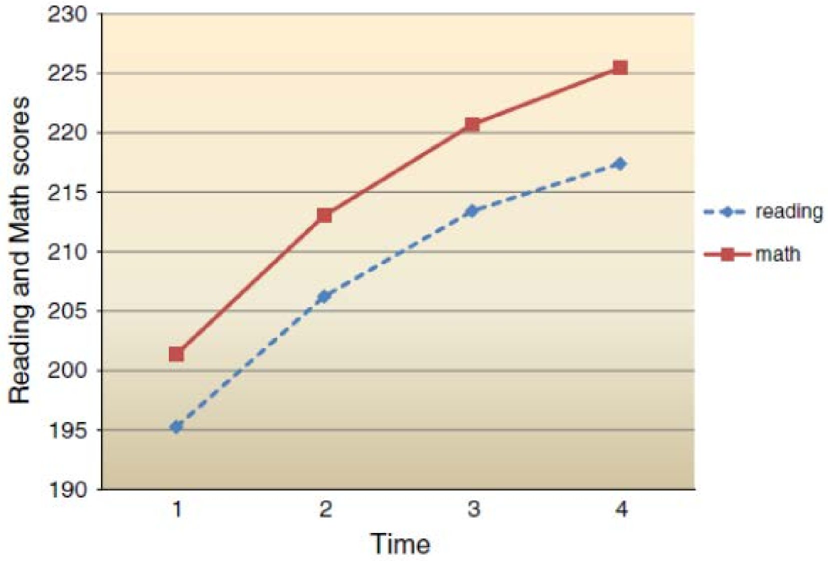

Fig. 2. Mean growth curve of reading and mathematics achievement plotted against time. Note. Using four-wave longitudinal data, Times 1, 2, 3, and 4 respectively indicate test scores in the 4th, 5th, 6th, and 7th grades.

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.

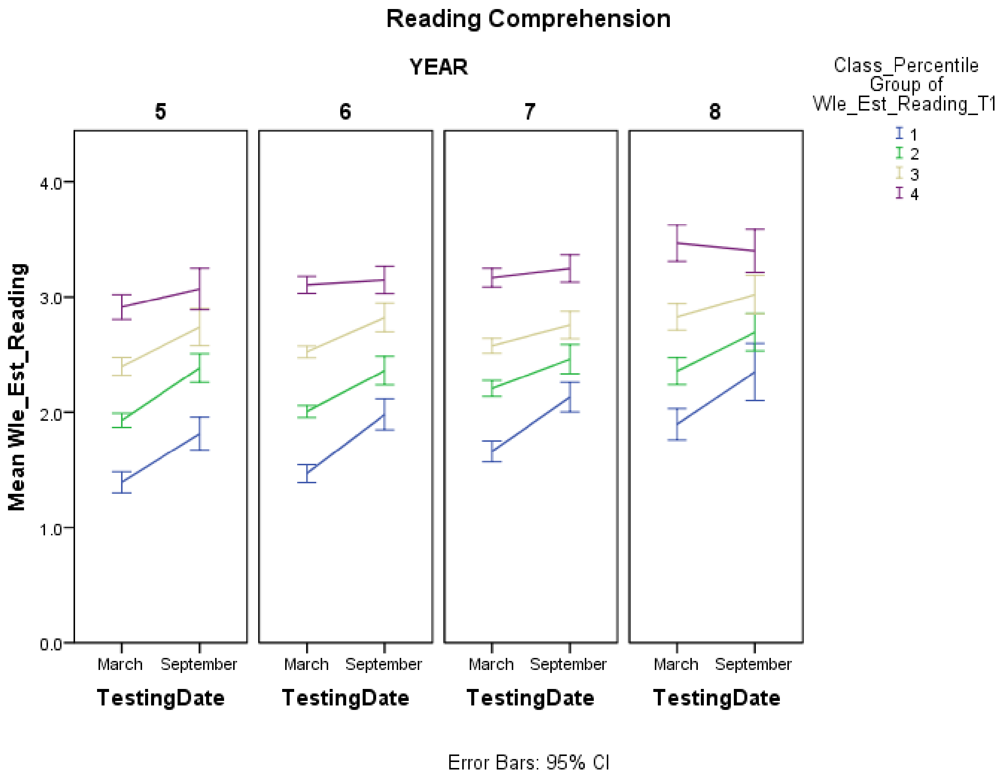

Figure 3. Reduced high performing student growth (purple lines) compared with low performing student growth (blue lines) as a possible function of cognitive differential growth (REAP data).

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.

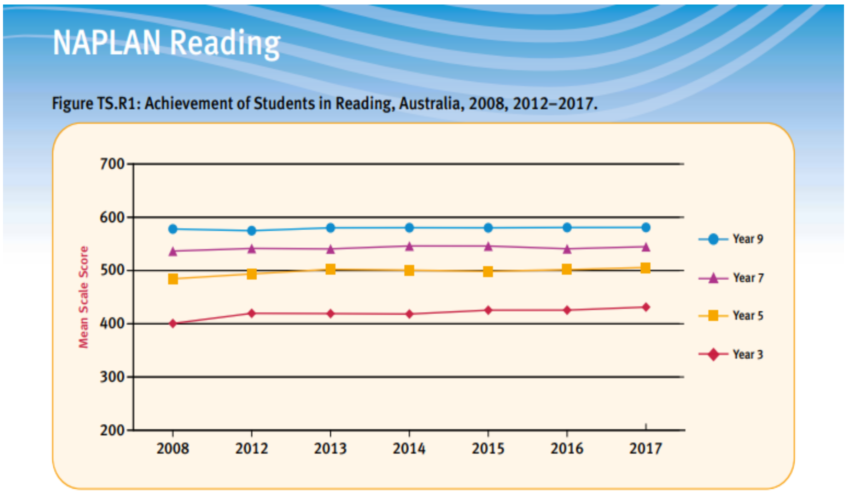

Figure 4. Time series-based differential growth: National Numeracy Trends 2008–2017 (ACARA, 2017).

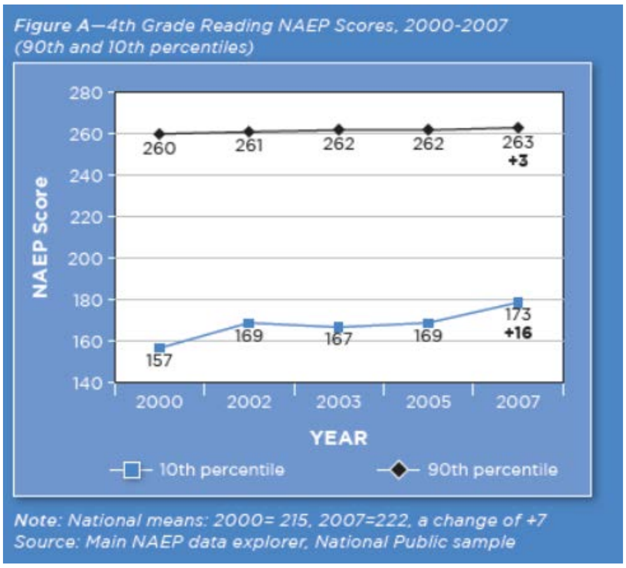

Figure 5. Time series-based differential growth: NAEP data 2000–2007 (Loveless, 2008).

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 are designed to take this into account, so items are included beyond those 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 at time 1 and time 2 are not perfectly correlated (Lohman & Korb, 2006). The observed score of a student on a test is an estimate 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 estimate 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 were 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.

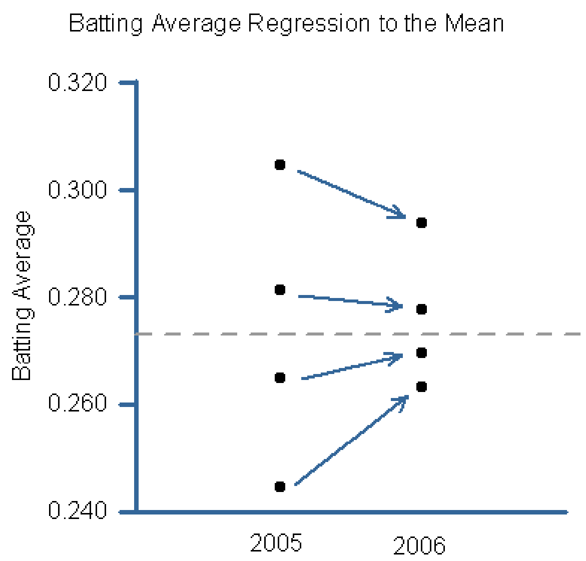

Figure 6. Regression to the mean differential growth: Batter averages for baseball (Studeman, 2007).

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) 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. 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 have been described 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 in some countries, curriculum designers 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 1,322.

Student inclusion was dependent on:

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

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

3. Sufficient data to link students' reading comprehension performance with their respective reading/English teacher. This was not possible in some circumstances where students were moved into a different class for reading 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. 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.

Only 37 teachers who had answered enough questions to allow accurate coding of their teaching practices were able to be included in the analysis. While the entire cohort of student (N= 1322) was used to determine the effect of school year and quartile on growth, smaller subsets were used to identify evidence for success of particular teaching practices (N = 721 for eight of the practices tests and N= 365 for one of the teaching practices). This is because not all teachers answered enough questions for researches to accurately designate 'presence' or 'absence' of particular teaching strategies. If researchers were not confident whether to assign a teacher as using a practice or not, that teacher and their students were not included in the analysis.

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 in the equations presented. 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*

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

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

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) reading comprehension ability estimates and differences: All students

Grade Level		T1 WLE Estimate	T2 WLE estimate	Mean Difference
5	Mean (std. dev) N	2.107 (0.642) 268	2.464 (0.763) 268	+ 0.357
6	Mean (std. dev) N	2.335 (0.675) 391	2.621 (0.757) 391	+ 0.268
7	Mean (std. dev) N	2.408 (0.682) 480	2.651 (0.787) 480	+ 0.243
8	Mean (std. dev) N	2.639 (0.725) 183	2.868 (0.747) 183	+ 0.229
All	Mean (std. dev) N	2.357 (0.695) 1322	2.634 (0.776) 1322	+ 0.277

Note: Standard deviation in parenthesis.

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

Grade Level		T1 WLE estimate	T2 WLE estimate	Mean Difference
5	Mean (std. dev) N	2.913 (0.395) 54	3.069 (0.661) 54	+ 0.156
6	Mean (std. dev) N	3.106 (0.383) 107	3.149 (0.618) 107	+ 0.043
7	Mean (std. dev) N	3.168 (0.436) 114	3.247 (0.634) 114	+ 0.079
8	Mean (std. dev) N	3.467 (0.526) 45	3.399 (0.619) 45	- 0.0675
All	Mean (std. dev) N	3.146 (0.453) 320	3.206 (0.636) 320	+ 0.059

Note: Standard deviation in parenthesis.

A graphical representation of the relationship between the mean T1 initial reading comprehension WLE per group and the growth of each group was created (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- or testing-based pattern (the higher the initial score, the least likely a student is to grow), or (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. With the exception of possibly the Grade 6 high capacity students, the high capacity students do not fall significantly below the best fit line for growth. This indicates that the differential growth pattern may almost be completely explained by either a cognitive- or testing-based flatline (refer to cognitive- and testing-based differential growth pattern explanations in Introduction).

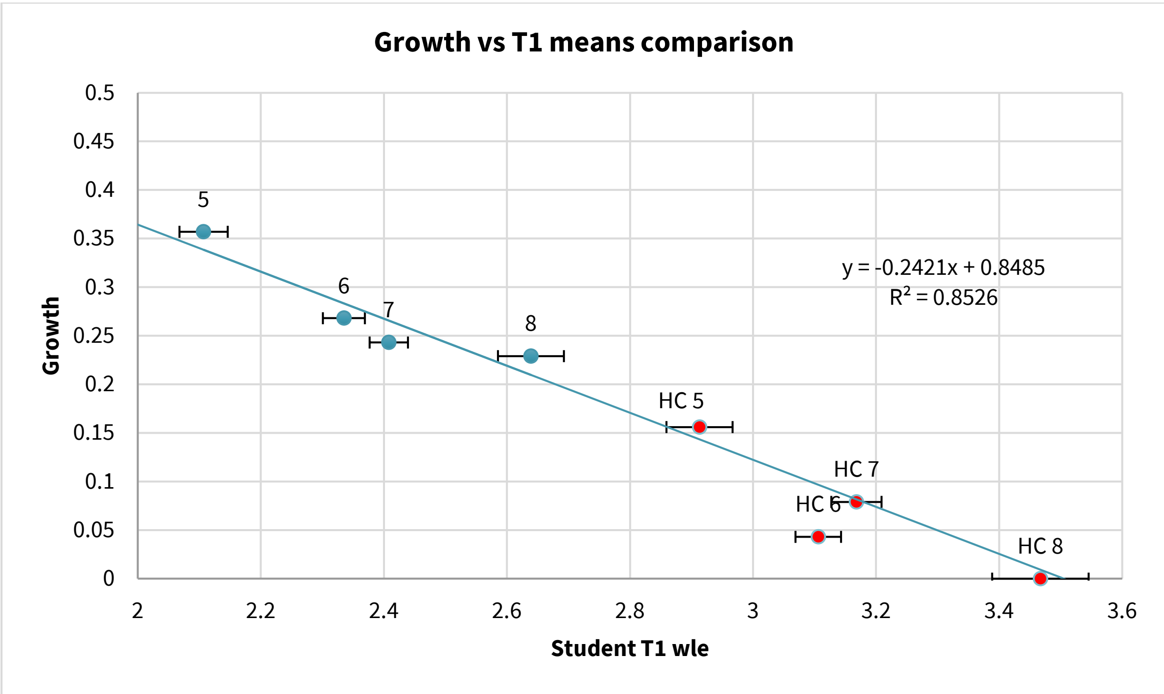

Figure 7. Graphical representation of mean students' reading comprehension T1 vs growth for specific groups of students.

Notes: High capacity 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.

This pattern can also be identified when the growth of each quarter of students on two progression levels are mapped (Figure 8). Within the same cognitive level, students in higher ability quarters are not growing as much as students in lower ability quarters (represented by the slope of the lines). Even within the same level or band of reading comprehension understanding, there is a range of ability levels. The students higher up on the y-axis (higher T1) do not progress as much as the students lower down. Students in the lowest ability quarter progress the most (Figure 8). This appears as though the data could be explained by a regression to the mean effect, or a cognitive reduction in growth effect or a teaching-based growth pattern where the high capacity students are not being catered to according to their ability. As all patterns can simultaneously occur in the data, the cause of the differential growth pattern cannot be elucidated.

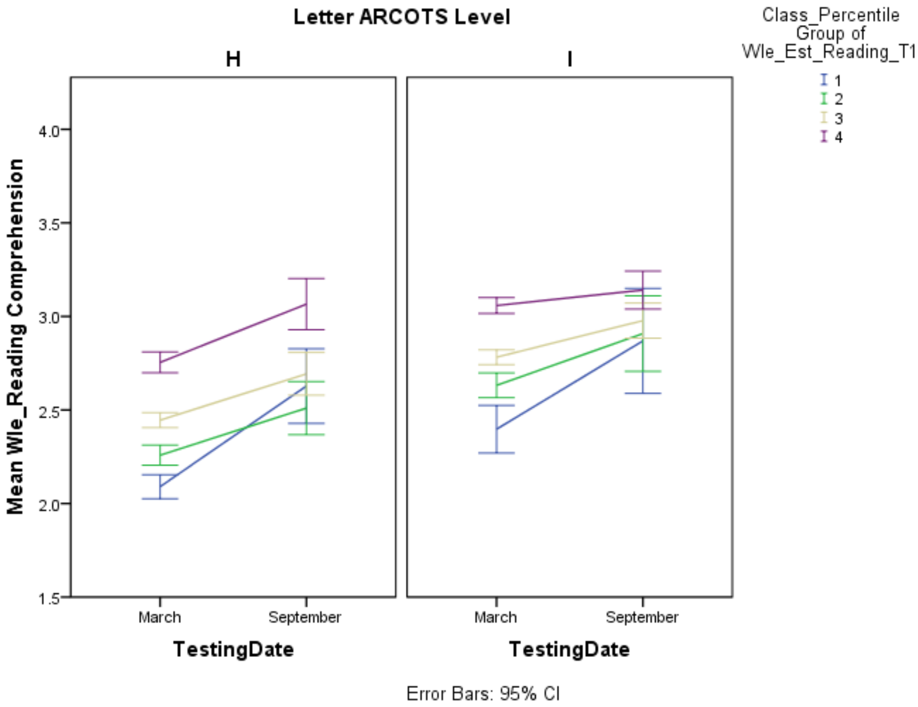

Figure 8. Mean students' reading comprehension T1 and T2 for two ARCOTS progression levels.

Notes: Students from all grades who are designated either of these levels are graphed. Level H, n = 350 students and Level I, n = 329 students. Students are separated into ability quarters per class. Blue lines are students in the lower ability groups per class (lowest 25%), green lines are students in low/middle 25%, brown lines are students in middle/high 25% and purple lines are high capacity students (top 25%).

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 = 1,322) is modelled for some of the analyses, a separate set of models is shown for the sub-set of students, n = 721 and n = 365, whose teachers responded to certain questions crucial to categorising the teacher's strategies. Results from each set of models are compared with the appropriate -2*log likelihood value to determine the significance of the model.

For the full cohort of students, with every 1 standard deviation (0.695 logits; Table 1) increase in T1 WLE estimate, students' T2 WLE estimate was predicted to increase by 0.751 logits. Students' T1 WLE estimate was a significant predictor of students' T2 WLE estimate, $p < 0.001$ (Table 3).

For high capacity students tested, with every 1 standard deviation (0.453 logits) increase in T1 WLE estimate, students' T2 WLE estimate was predicted to increase by 0.514 logits (Table 4).

School type, grade and quartile were not significant predictors of students' T2 beyond what students' T1 score predicted (Table 3). The finding that student quartile was not a predictor for growth (beyond T1) can also be viewed in Figure 7. This finding was not consistent with the mathematics data (Harding et al., 2018), where student quartile did have a significant effect on student growth beyond students' T1. Importantly this study consisted of teachers who were using a variety of teaching strategies to support high capacity students, therefore the differential growth pattern or significance of growth differences between quartiles should have been somewhat ameliorated or completely ameliorated for this sample of students. The hope for the study was that the differential growth pattern or flat-line would vanish. Independent of the causes or patterns in differential growth for high capacity students, of interest to this study are the teacher covariates that can improve growth for the students.

Teacher strategies that predicted a greater T2 score (after controlling for T1) are listed in Table 4, with the effect size of each strategy and significance noted for Model 4. In Model 5, modifying teaching for individual students and teaching students SRL behaviours were identified as being the greatest teaching strategy to enhance student growth for all students (Table 4). While the largest significant predictor for student growth was to include high order thinking skills in reading comprehension instruction (Model 4), this covariate could not be included in the final model (Model 5) as the sample size was low for teachers who responded to the questions allowing coding for this variable (n = 365).

Modifying whole-class instruction and selecting or modifying tasks given based on students' ZPD were both significant predictors of student growth based on individual modelling of teacher covariates (Table 4, Model 4). While including high order thinking skills in reading comprehension instruction had the same effect size as that for the whole cohort (compare Table 3, Model 4 and Table 4, Model 4), the covariate did not reach significance in the high capacity sample due to larger error. Selecting and modifying tasks given to students at their ZPD and teaching students how to regulate their own learning were significant predictors of growth for high capacity students in the full model (Table 4, Model 5). The effect sizes of significant teacher covariates were greater for high capacity students than for all students (compare Tables 4 and 3); this could be due the emphasis on high capacity student during the project. This pattern was also observed in the mathematics data (Harding et al., 2018).

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

Parameter	Intercept β_0	T1 Student Predictor β_1	Level 2 Predictor β_k	$E_{ij} \sim N(0, \sigma_e^2)$ σ_e^2	-2*log likelihood n = 1,322	-2*log likelihood n = 721 (Δ,df)	-2*log likelihood n = 365 (Δ,df)
Model 1: Null							
Unconditional (n = 1,322)	2.634(0.021)	-	-	0.601(0.023)	3079.330		
Unconditional (n = 721)	2.642(0.028)	-	-	0.581(0.031)		1655.200	
Unconditional (n = 365)	2.527(0.042)	-	-	0.634(0.047)			869.446
Model 2: Student							
Student T1 (n = 1,322)	0.865(0.056)	**0.751(0.023)*** **	-	0.329(0.013)	**2282.617(797,1)*** **		
Student T1 (n = 721)	0.888(0.073)	**0.749(0.030)*** **	-	0.310(0.016)		**1202.057(454,1)*** **	
Student T1 (n = 365)	0.747(0.099)	**0.789(0.042)*** **	-	0.322(0.024)			**621.758(248,1)*** **
Model 3: Cohort predictors							
School Type (Primary)							
Secondary	0.869(0.056)	**0.753(0.023)*** **	-0.019(0.032) ns	0.329(0.013)	2282.251(0.37,1) ns		
School Year (Grade 5)							
Grade 6	0.880(0.060)	**0.752(0.023)*** **	-0.013(0.046) ns	0.329(0.013)	2281.395(1.22,3) ns		
Year 7			-0.039(0.044) ns				
Year 8			0.005(0.056) ns				
Quartile (Q1)							
Q2	0.839(0.069)	**0.766(0.039)*** **	-0.005(0.049) ns	0.329(0.013)	2281.555(1.06,3) ns		
Q3			0.009(0.059) ns				
Q4			-0.042(0.076) ns				
Model 4: Teaching predictors (37 teacher)							
Using assessment to:							
Modify teaching for individuals	0.846(0.078)	**0.749(0.030)*** **	0.063(0.044) ns	0.309(0.010)		1200.008(2.05,1) ns	
Modify whole-class instruction	0.873(0.076)	**0.751(0.030)*** **	0.031(0.043) ns	0.310(0.016)		1201.528(0.53,1) ns	
Select or modified tasks given	0.821(0.082)	**0.757(0.030)*** **	**0.076(0.044)* **	0.309(0.016)		1199.069(2.99,1) ns	
Use evidence to identify student ZPD	0.848(0.081)	**0.749(0.030)*** **	0.053(0.047) ns	0.310(0.016)		1200.782(1.28,1) ns	
Evaluate impact of intervention	0.872(0.076)	**0.750(0.030)*** **	0.033(0.042) ns	0.310(0.016)		1201.443(0.61,1) ns	
Collaborate with colleagues to develop assessments	0.863(0.079)	**0.750(0.030)*** **	0.034(0.043) ns	0.310(0.016)		1201.422(0.63,1) ns	
Include teaching of SRL skills	0.846(0.076)	**0.754(0.030)*** **	**0.081(0.043)* **	0.309(0.016)		1198.482(3.58,1) ns	
Include teaching of collaborative learning skills	0.896(0.077)	**0.748(0.030)*** **	-0.014(0.043) ns	0.310(0.016)		1201.957(0.10,1) ns	
High order thinking	0.655(0.110)	**0.796(0.042)*** **	**0.116(0.063)* **	0.319(0.024)			618.313(3.45,1) ns
Model 5: Full model							
Using assessment to:	0.701(0.103)	**0.752(0.030)*** **		0.303(0.016)		**1184.951(17.11,8)* **	
Modified teaching for individuals			**0.084(0.046)* **				
Modify whole-class instruction			-0.005(0.051)				
Select or modified tasks given			0.073(0.049)				
Use evidence to identify student ZPD			0.059(0.048)				
Evaluate impact of intervention			0.032(0.047)				
Collaborate with colleagues to develop assessments			0.040(0.055)				
Include teaching of SRL skills			**0.166(0.065)** **				
Include teaching of collaborative learning skills			**-0.178(0.068)** **				

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	$E_{ij} \sim N(0,\sigma_e^2)$ σ_e^2	-2*loglikelihood n = 320	-2*loglikelihood n = 176 (Δ,df)	-2*loglikelihood n = 92 (Δ,df)
Model 1							
Unconditional (n = 320)	3.206(0.036)	-	-	0.404(0.032)	617.884		
Unconditional (n = 176)	3.178(0.046)	-	-	0.367(0.039)		323.003	
Unconditional (n = 92)	3.120(0.065)	-	-	0.386(0.057)			173.433
Model 2							
Student T1 (n = 320)	1.589(0.233)	**0.514(0.073)*** **	-	0.350(0.028)	572.051		
Student T1 (n = 176)	1.631(0.299)	**0.496(0.095)*** **	-	0.318(0.034)		**297.594(25.41,1)*** **	
Student T1 (n = 92)	1.666(0.373)	**0.482(0.122)*** **		0.330(0.049)			**159.064(14.37,1)*** **
Model 3: Cohort predictors							
School Type (Primary)	1.609(0.233)	**0.497(0.075)*** **		0.349(0.028)	571.203		
Secondary			0.063(0.068)^{ns}				
School Year (Grade 5)	1.616(0.241)	**0.499(0.078)*** **		0.349(0.028)	571.174		
Grade 6			-0.017(0.100)^{ns}				
Year 7			0.050 (0.100)^{ns}				
Year 8			0.054 (0.127)^{ns}				
Model 4: Teaching predictors (37 teachers)							
Using assessment to:							
Modify teaching for individuals	1.582(0.309)	**0.499(0.095)*** **	0.057(0.091)^{ns}	0.317(0.034)		297.207(0.39,1)^{ns}	
Modify whole-class instruction	1.394(0.394)	**0.543(0.094)*** **	**0.238(0.087)** **	0.305(0.032)		**290.274(7.32,1)** **	
Select or modified tasks given	1.127(0.328)	**0.595(0.097 *** **	**0.299(0.091)*** **	0.299(0.032)		**287.149(10.45,1)*** **	
Use evidence to identify student ZPD	1.614(0.315)	**0.497(0.095)*** **	0.017(0.097)^{ns}	0.318(0.034)		297.562(0.03,1)^{ns}	
Evaluate impact of intervention	1.643(0.314)	**0.476(0.095)*** **	0.119(0.086)^{ns}	0.314(0.033)		295.695(1.90,1)^{ns}	
Collaborate with colleagues to develop assessments	1.674(0.306)	**0.493(0.095)*** **	-0.057(0.087)^{ns}	0.317(0,034)		297.166(0.43,1)^{ns}	
Include teaching of SRL skills	1.558(0.305)	**0.508(0.095)*** **	0.096(0.088)^{ns}	0.315(0.034)		296.405(1.19,1)^{ns}	
Include teaching of collaborative learning skills	1.693(0.309)	**0.484(0.096)*** **	-0.068(0.089)^{ns}	0.317(0.034)		297.006(0.59,1)^{ns}	
High order thinking	1.821(0.410)	**0.457(0.125)*** **	0.116(0.131)^{ns}	0.327(0.048)			158.274(0.79,1)^{ns}
Model 5: Full model							
Using assessment to:	1.114(0.378)	**0.574(0.102)*** **		0.282(0.030)		**276.744(20.85,8)** **	
Modified teaching for individuals			0.055(0.093)^{ns}				
Modify whole-class instruction			0.127(0.098)^{ns}				
Select or modified tasks given			**0.227(0.100)***				
Use evidence to identify student ZPD			0.038(0.095)^{ns}				
Evaluate impact of intervention			0.087(0.096)^{ns}				
Collaborate with colleagues to develop assessments			-0.044(0.108)^{ns}				
Include teaching of SRL skills			**0.287(0.127)***				
Include teaching of collaborative learning skills			**-0.283(0.134)***				

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.*

Further research is required to determine the cause of the reduced growth pattern for high capacity students in reading comprehension. It could be interpreted as a result of teaching practices that target the lower ability quarters of the class, but there is evidence to suggest a strong cognitive- or testing-based differential growth pattern also; therefore, the lack of differentiated teaching strategies within a class may not be the primary issue (refer to the non-significant effect of quarter on growth, Table 3) visualised in Figure 7. Even though the cause of reduced growth for higher ability students remains unknown, the data does show that the higher the cognitive level on the ARCOTS reading progression, the less growth students achieve (Figure 7). While this could be due to a testing-based pattern, the high capacity Grade 6s are lower than the line predicted by the means of the grade cohorts (Figure 7). While this does not reach significance (Table 3), these students do have reduced growth compared to what is predicted by their T1 score. This might be due to a lack of appropriate task selection or modification for high capacity students, as this was a significant factor for high capacity student growth (Table 4).

Regardless if there is a systemic failure to teach high capacity students based on their ZPD, which was the hypothesis behind the 'flat-line effect' noted by previous studies, the aim of this study was to identify teaching practices that had evidence for growth for HC students. These strategies have been identified in table 4 and described in the teacher recommendations section of this report.

Teacher strategies that had evidence for success in catering for HC students are summarised in table 5 with effect sizes included as well as logit increases. With the exception of "Teaches use assessment data to modified whole class instruction", each strategy that was identified as supporting the growth of HC students, was also found to be successful in supporting the growth of "all students" (table 3 and 5). This is a promising finding as it supports the notion that quality teaching practices support all students. Importantly the effect of the quality teaching practices was larger in the high capacity student sample (table 4 and 5). This suggests that using teaching strategies involving differentiation for student ability are particularly important for HC students. 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 Teacher covariate modelled separately (model 4)**	**Growth for HC Teacher covariate modelled separately (model 4)**
Teaches use assessment data to modify teaching for individual students – self report	59%	NS	NS
Teaches use assessment data to modified whole class instruction – self report	30%	NS	0.238(0.087)** 37% variance explained
Teaches use assessment data to select or modify the learning tasks given to students – self report	58%	0.076(0.044)* 7.7% variance explained	0.299(0.091)*** 67% of variance explained
Teachers use evidence to identify student ZPD – teacher opinion	75%	NS	NS
Teachers evaluate the impact of the interventions they trialled with the students – teacher opinion	46%	NS	NS
Teachers collaborate with colleagues to develop assessments – self report	65%	NS	NS
Include teaching of self-regulated learning skills – self report	39%	0.081(0.043)* 7.7% variance explained	NS
Include teaching of collaborative learning skills – self report	42%	NS	NS
High order thinking: Teachers planned for extending students high order thinking in student groups according to the groups level of skill or using individual student learning plans instead of including vague plans, whole class plans, or no plans for High order skills. – reported planning	66%	0.116(0.063)* 20% of variance explained	NS[#]

Notes: No secondary teachers in the study reported use of online curriculum as a teaching strategy; *p < .05, **p < .01, ***p < .001
All variance changes were calculated from model 4; teaching strategies modelled separately.
[#]While the β_k for this strategy used in the HC cohort was the same as in the full cohort (0.116, refer to table 4), the large error size (0.131) resulted in the strategy failing to reach significance. The sample size for [#] was only 92 students (4 teachers).

The PVC for the full model for all students (all teaching strategies modelled together, model 5) was 0.616. Therefore, the teaching strategies examined make up 61.6% of the variance which explains the students' results. This indicates that there are other teaching predictors that have not been accounted for in this study, which make up the remaining 38.4 % of the variance explained by the teacher.

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, 95% of the variance in student growth was caused by student differences, whereas only 4.3% of the variance in growth was attributed to class-level effects. For high capacity students, 88.8% of the variance in student growth was caused by student differences, whereas 12% of the variance in growth was attributed to class-level effects. The increase from 4.3% to 12% explained by class-level effects could be due to the increased impact of teaching strategies that cater for HC students on

All students (n = 1322)

Within-class variance

$E_{ij} \sim N(0, \sigma_e^2)$ σ_e^2 = 0.315(0.013)

Between-class variance:

$\mu_{ij} \sim N\left(0, \sigma_{\mu 0}^2\right)$ $\sigma_{\mu 0}^2$ = 0.014(0.005)

(total variance $e_{ij} \sim N(0, \sigma_e^2)$ σ_e^2 = 0.329 as shown in table 3 column 4)

Variance partition coefficient (VPC) = 0.014/ (0.014+0.315) = 0.043 or 4.3% of the variance in growth was attributed to class-level effects.

VPC = 0.315/ (0.014+0.315) = 0.957 or 95% of the variance in growth was attributed to student differences within class.

High capacity students (n = 721)

Within-class variance:

$E_{ij} \sim N(0, \sigma_e^2)$ σ_e^2 = 0.309(0.028)

Between-class variance:

$\mu_{ij} \sim N\left(0, \sigma_{\mu 0}^2\right)$ $\sigma_{\mu 0}^2$ = 0.042(0.020)

(total variance $e_{ij} \sim N(0, \sigma_e^2)$ σ_e^2 = 0.350 as shown in table 4 column 4)

VPC = 0.042/ (0.042+0.309) = 0.1197 or 12% of the variance in growth was attributed to class-level effects.

VPC = 0.309/ (0.042+0.309) = 0.88 or 88% of the variance in growth was attributed to student differences within class.

Teacher Recommendations

The information presented here is not a “one size fits all” approach to teaching; the practice must fit the personality, philosophy and circumstances of the teacher.

Defining reading comprehension

Defining reading comprehension is complex given our concept of reading has expanded greatly with the expansion of technology and digital media. Reading requires multiple literacy skills to decode multiple text types from which information is extracted, understood and shared in a variety of ways and for a variety of purposes (Coiro, 2003). From an early age, children navigate print, diagrams, images, links, screen menus, video, audio and static as well as interactive content. Modern reading is multi-dimensional and nonlinear; it is a complex, individual, social and multi-resourced process (Duke & Pearson, 2008) that requires negotiation of multiple internal and external variables. Digital literacy, as the newest wave of communication, is at least as crucial as traditional literacies with both unique and overlapping comprehension skills required to enable the individual to “not only communicate, but to create, to manipulate, to design, to self-actualize” (Jones & Flannigan, 2006, p. 3). It is with these considerations and available literature that Nibali (doctoral dissertation; personal correspondence) defines reading comprehension as the interaction of several component processes that integrate information from a text with the individual’s background knowledge, and which are subject to a multitude of contextual constraints. Reading comprehension inherently involves inferential and evaluative thinking; it is not only used to communicate, but to create, to manipulate, to design and to self-actualise. Reading comprehension simultaneously extracts and constructs meaning through interaction and involvement with written and/or digital language.

Using assessment to select or modify tasks

This research demonstrated that teachers who use assessment data to select or modify tasks for students of different reading comprehension abilities increased their high capacity reading students’ growth (and the average growth of all students in the class).

Opportunity for students to learn advanced reading comprehension skills requires the selection of appropriate tasks that encourage the individual student’s progress. This opportunity to learn comes about by careful selection of activities based on student competence. Evidence, in the form of ongoing assessment, is used to select the tasks for students at specific stages along the reading comprehension developmental continuum. Reading assessments, such as the Progressive Achievement Tests in Reading (PAT-R), ARCOTS, NAPLAN and the VCAA’s On Demand Testing, can help teachers place students on a learning progression. Teachers also use their own expert judgement about a student and their abilities, collecting evidence to inform the location of the student on the construct. It is important, then, that teachers appropriately and correctly select tasks that match the student's point of readiness to learn, so that students can move through the developmental progression to the next level of reading comprehension competency.

Appropriate selection of reading questions and tasks for students of varying reading comprehension abilities requires that the teacher

1. has an established knowledge about, and understanding of, the stages and content of the developmental learning progression in reading comprehension;
2. has appropriately assessed each student's reading comprehension competence using a developmental approach; this may include recording what the student reads in their own time and other informal or formal methods of assessment, e.g. rubrics, observations or the assessments mentioned above;
3. can adapt practical strategies for addressing each level of competence while ensuring that all students can participate and contribute ideas and processes in the learning of reading comprehension.

For example, a Year 7 English teacher, Kalinda, explained that she needed to use assessments and tasks for students to demonstrate their ability to infer to a high and complex level. She designed a reading comprehension assessment that included "asking questions about character motivations where there's significant interpretation of the text required and there's quite subtle understanding working in multiple sections of the texts and having to make inferences about setting, character and story and plot to draw a conclusion". Kalinda was then able to select and modify reading comprehension tasks based on students' skill levels.

Example of how to select or modify tasks for students of varying levels of competence

It can be difficult for teachers to select or have access to a variety of different texts, targeted for a range of student ability, for all classes or activities. Instead, teachers could consider using a single text, like the example in Figure 9, and then choose assessment items/questions that cater for differing starting points. Analysing students' responses, the teacher can then allocate appropriate activities for individuals or groups of students to complete.

Read the passage below and then answer the following questions:

LOSING IT

The following is the conversation between two people discussing a recent incident at school that you've overheard.

JT: *"I think it's important to understand how this sort of behaviour arises. I mean, there must be some way to avoid it happening. Maybe we all contributed in some way."*

NP: *"I don't agree. You don't see everyone doing that sort of thing, do you?"*

JT: *"No, but we were all aware of how upset he gets about little things. It's only natural that he lashes out at some stage. It's a wonder no-one got injured."*

NP: *"Lash out? I hardly think we can call it 'lashing out'. I would think it more appropriately be described as 'losing control'."*

JT: *"Either way, at least the principal has meted out a reasonable punishment."*

NP: *"Absolutely. We can always rely on her to respond in a just way whatever the crisis."*

1. Which of the following is true about the two speakers in the conversation?
    a. They agreed on most things in the context of the matter
    b. They see the entire situation differently from one another
    c. They believe that an appropriate response to the incident has occurred
    d. They believe that the perpetrator deserves a severe punishment
2. Which of the following is most likely the incident that the students were discussing?
    a. A student playing 'spiderman' jumped off the slide and landed on another student
    b. A student yelling at the class and running from the room in anger
    c. A student kicking a football into a crowd of prep school students
    d. A student writing rumours about another student on the school wall
3. Which of the following best describes the view point of speaker JT, in comparison to speaker NP
    a. Considered
    b. Impetuous
    c. Reliable
    d. Spiteful

Figure 9. Example Reading comprehension item with questions.

After the student completes the questions, the teacher can look at selecting an appropriate activity for each student based on previous assessment data and the responses to the questions. It is important to consider not only the correct response in the original assessment, but also why a student may have selected a particular option and the level of skill that this implies. Appropriate activities might include those presented in Figure 10, ordered by increasing difficulty. The third activity caters for high capacity students who can infer to a high level from the text.

Activity 1: You are a reporter for the local newspaper. Write an article that covers the school incident. Use no more than 2 quotes from the conversation provided.

Activity 2: Write a fictional account of what happened from the viewpoint of the protagonist.

Activity 3: Compile a behaviour profile of the student in question and include a plan for managing this behaviour in the future.

Figure 10. Example Activities catering for different abilities.

The teacher may then finish the unit of work with a whole-class discussion, giving the students opportunity to share their understanding of the characters in the text with the rest of the class.

The following list includes other ways that teachers in the project reported selecting and modifying tasks based on assessment data:

- "Varying the complexity of the task according to student ability and need"
- "Select texts that stretch the comprehension skills of high capacity students – challenge their thinking not just decoding skills"
- '[Providing] open ended scaffolding so as not to limit students"
- "Include at least three levels in planning documents so the task suits high, medium and low ability students"
- "Increasing teacher conferencing/interviews to target and extend student ability"
- "Investigations/project work that includes open ended scaffolding"
- "Flexible groupings that change by content area/focus, with high capacity students working together"
- "Differentiated tasks – different entry and exit points for students . . . and open-ended tasks"

Modifying whole-class instruction based on student assessment data

Modifying whole-class instruction based on student assessment data was also found to be related to growth of high capacity students. This finding was a little surprising; however, it could be that teachers who change their whole-class instruction based on the assessed level of students in their class may not only be targeting their teaching at the student's point of need but also pitching their instruction at a higher level, raising expectations for all students. This strategy may be successful as "students learn more if you hold high but reasonable academic expectations for them" (Blair, Rupley, & Nichols, 2007, p. 436). It is possible that other teachers did not realise the range for which they needed to pitch lessons to include the high capacity students in their class.

Modifying whole-class instruction to a higher level includes extension for the high capacity students. The use of specific questioning is one way teachers can modify whole-class instruction to extend their students. Questioning can raise or lower the cognitive demands of a task, meaning that there are lower cognitive and higher cognitive questions that can be asked of the class. Lower cognitive questions are those that ask the student merely to recall verbatim, or in their own words, material previously read or taught by the teacher. Lower cognitive questions are also referred to in the literature as fact, closed, direct, recall and knowledge questions. Higher cognitive questions are defined as those that ask the student to mentally manipulate information previously learned to create an answer or to support an answer with logically reasoned evidence. Higher cognitive questions are also called open-ended, interpretive, evaluative, inquiry, inferential and synthesis questions. A list of question types based on lower order and higher order, open and closed, questions can be found in Table 5. This table was informed by the work of Hess et al. (2009, updated 2013), Webb (2002) and Bloom (1984). Examples of how these questions can be used can be found at the following links: Questioning and Bloom's Taxonomy of Questions.

Teachers should be conscious that the more complex mental operations required by students when higher cognitive questions are asked call for longer processing times. Providing students with more time to think and complete work is often needed for students to process at higher cognitive levels, but may not always be required for high capacity students (Peterson & Taylor, 2012; Tobin, 1987). Teacher knowledge of individual students is key to understanding the time required to complete a task.

As a further stimulus for teachers interested in improving their students' literacy levels, see the Department of Education & Training's Expert Literacy Videos (Expert Literacy Videos). The videos are presented by Professor Lorraine Graham and specifically address differentiation to meet as "many student needs in the classroom as possible".

Table 6. Examples of levels of questioning

Question Type	Examples
Type 1 (Low order, closed) Knowledge – retention of terminology, facts, conventions, methodologies, structures, principles list, define, describe, show, name, what, when	• Name all the characters in the story. • List five facts from the story. • When does the story take place? • Where does the story take place? • Which character appears first in the story? • How does the story end? • From what you read in the story, how would you describe what the main character looks like? • How would you describe the setting, using facts you read in the story?
Type 3 (High order, closed) Application – problem solving, usage of information in a new way apply, calculate, complete, show, solve, modify	• Think of a situation that occurred to a person in your story and decide whether you would have done the same thing or something different. What might you have done? • Give some examples of people who have had the same problems or have done the same kind of thing as the person in your story. • What are some things this character/s might do if he came to your school during reading? • What would you do if you could go to the place where the main character lives? • What would the main character do if he came to your house to visit? • If you had to cook a meal for the characters in the story, what would you cook? • If you met the main character in the story on the street, what would you talk about?
Type 2 (Low order, closed) Comprehension – grasping of meaning, translation, extrapolation, interpretation of facts, making comparisons summarise, compare and contrast, estimate, discuss	• What is the story about? • How did the main character feel at the beginning of the story? • How did the main character feel at the end of the story? • Think of a main event in the story. Why did it happen? • Why does the story have this title? • What is the main event in the story? • Explain why this picture may have been used. • What happened before/after this picture?
Type 4 (Low order, open) Analysis – making inferences and supporting them with evidence, identification of patterns separate, arrange, classify, explain	• What part of the story was the funniest, or the most exciting, or the saddest? • Tell what things happened in the story that couldn't have happened in real life. • Which of the things in the story were true and which were opinions; whose opinion? • Organise the story into parts – what would be a good title for each of the parts? • What could you do that was just like what the person in the story did? • Find five words in the story that begin with the same sound.
Type 5 (High order, open) Synthesis – derivation of abstract relations, prediction, generalisation, creation of new ideas integrate, modify, substitute, design, create, What if..., formulate, generalise, prepare	• How could you tell the story from an animal's point of view? • What would you put into a picture about the story? What would you add to the picture that was not in the story? • What would be the best way to tell this story if not in a book? • What is another possible ending to the story that is different from the one that the author wrote? • Pretend you are the main character. What might you write in your diary about this day? • How would you rewrite the story briefly, changing someone or something in it (for example, substitute a dog for a wolf in the Three Little Pigs). How would this change the meaning of the original story?
Type 6 (High order, open) Evaluation – judgement of validity, usage of a set of criteria to make conclusions, discrimination assess, rank, test, explain, discriminate, critique, support	• Was the main character in the story good or bad? Why? • Compare any two books you've read and tell which one you would recommend to your friend/s and why. • Compare two characters in the story. Tell which one you think is the bravest and why. • Which character in the story would you most like to spend the day with? Why? • Was this story worth the time it took to read? Why? • If you had the opportunity to go where this story takes place, would you want to go? Why/why not?

Planning for higher order thinking

Teachers who reported planning for extending students' high order thinking skills had higher growth for their students in reading comprehension. This may be because these teachers effectively targeted the level of thinking required, as reading comprehension develops towards the higher levels. Reading comprehension becomes increasingly complex, sophisticated and requires higher order skills as students progress in their education (Dymock & Nicholson, 2010; Griffith, Bauml, & Barksdale, 2015; McGee, Ward, Gibbons, & Harlow, 2004).

There are many different definitions among educators of what is meant by higher order skills and higher order thinking. Resnick (1987) defined higher order thinking skills as involving "a cluster of elaborative mental activities requiring nuanced judgement and analysis of complex situations according to multiple criteria" (p. 44). More recently, higher order thinking refers to "transfer", "critical thinking" and "problem solving": "transfer" refers to the student's ability to apply knowledge and skills to new contexts; "critical thinking" refers to the ability to reason, reflect and decide what to believe or do next; and "problem solving" refers to meeting a goal that cannot be met with a memorised solution or, in broader terms, is the application of non-automatic strategising (Brookhart, 2010, 2011).

Teachers are likely to be familiar with developmental models, thinking skills and learning taxonomies such as Bloom's revised taxonomy (remember, understand, apply, analyse, evaluate, create) (Anderson & Krathwohl, 2001) and Webb's (2002) depth of knowledge (recall and reproduction, application of skills, strategic thinking, extended thinking). These models share a common expectation in that the highest levels are about not only which or how much knowledge or skill is acquired, but how and for what purpose these capabilities are acquired and used. These models can also be used to plan activities at a high level.

For example, one strategy for planning for the development of higher order thinking skills is through careful consideration of the depth of knowledge required of students in any given task. The Depth of Knowledge Table (Table 6) can be used to inform planning of higher order skills and audit tasks for levels of thinking. This table was informed by the work of Hess et al. (2009) and Webb (2002). Including reading comprehension content and skills at the highest level (Level 4 – Extended Thinking) can help develop students' higher order thinking skills and help students set goals and monitor progress relative to more sophisticated comprehension skills.

Afflerbach, Cho, and Kim (2015) differentiated between basic and higher order reading comprehension skills. Basic skills include the ability to identify written words and match these from memory, and build an understanding of simple text. Higher order skills include the ability to manage constructive and integrated processes, make complex inferences using text information and prior knowledge, and deconstruct a text into idea units to grasp what the text says (Afflerbach, Cho, & Kim, 2015).

Table 7. Strategies for targeting depth of knowledge

DOK Level	Example of Activity Type and Expectation	Focus/Goal and Outcome	
Level 1 **Recall and Reproduction** *Recall a fact, term, principle or concept; perform a routine procedure.*	✓ Support ideas by reference to details in the text. ✓ Use a dictionary to find the meaning of words. ✓ Identify figurative language in a reading passage. ✓ Verbs used include identify, recall, use, recognise, calculate, measure, describe/explain.	**Focus/Goal** • Facts • Definitions • Routines • Procedures	**Outcome** • Right answer • Definitions • Details • Facts • Routines
Level 2 **Basic Application of Skills/Concepts** *Use information, conceptual knowledge; select appropriate procedures for a task; perform two or more steps with decision points along the way; solve routine problems; organise or display data; interpret or use simple graphs.*	✓ Use context cues to identify the meaning of unfamiliar words. ✓ Predict a logical outcome based on information in a reading selection. ✓ Identify and summarise the major events in a narrative. ✓ Verbs used include collect, classify, organise, estimate. ✓ Observe, explain describe, Interpret, collect, display/compare data, summarise.	**Focus/Goal** • Make connections • Explain how/why • Approximate • Categorise • Compare/contrast • Interpret • Infer • Determine whether fact or opinion	**Outcome** • Right answer • Apply skill or concept • Focus on relationship by comparing or cause–effect • Explain how/why • Make a decision – estimate • Interpret information
Level 3 **Strategic Thinking** *Reason or develop a plan to approach a problem; employ some decision-making and justification; solve abstract, complex or non-routine problems, complex. (DOK-3 problems often allow more than one possible answer.) Deep knowledge becomes more of a focus. Students are encouraged to go beyond the text; however, they are still required to show understanding of the ideas in the text.*	✓ Activities encourage students to explain, generalise or connect ideas. ✓ Activities involve reasoning and planning. ✓ Items may involve abstract theme identification and inference across an entire passage or students' application of prior knowledge. ✓ Items may involve superficial connections between texts. ✓ Verbs used include conclude, present evidence, argue a position, explain/ justify.	**Focus/Goal** • Identify and apply • Consider various contexts • Plan • Develop a strategy to explain potential resolution	**Outcome** • More than one answer is possible • Apply information in a new context • Reasoning is necessary • Planning is involved • Complex thinking • Abstract thinking • Support provided to explain/defend responses
Level 4 **Extended Thinking** *Perform investigations or apply concepts and skills to the real world that require time to research, problem solve, and process multiple conditions of the problem or task; perform non-routine manipulations across disciplines, content areas or multiple sources.*	✓ Activities require the student to o take information from at least one passage and apply this information to a new task; o develop hypotheses and perform complex analyses of the connections among texts; o extend their knowledge and skills; o synthesise knowledge to create new solution/ideas.	**Focus/Goal** • Connect information across all levels • Apply real world contexts • Decipher information and outcomes and find solutions	**Outcome** • More than one possible answer • Time required to process response – to generate • Apply information in a new context (generalise) • Complex reasoning and planning • Multiple steps • Cross-curricular • Real world connections

The low and high cognitive questions presented in Table 5 can also be used to target reading comprehension activities towards specific levels of thinking and skills, including higher order.

The following list includes strategies teachers used that were associated with the development of higher order thinking skills for students in their class.

- "Encourage asking thoughtful questions and digging deeper than initial responses – go beyond knee-jerk reactions"
- "Socratic questioning techniques"
- "Expanding/widening their knowledge . . . using different perspectives and empathy when reading texts"
- "Ensuring access to higher order thinking tools/skills/activities/graphic organisers etc., and challenging and engaging discussion prompts and questions during weekly Guided Reading Group sessions where students work with students of a similar high capacity or with the teacher"

Teaching of self-regulated learning skills

We found that teachers who reported teaching SRL skills enhanced the development of reading comprehension in their students. This finding is supported by a significant body of research that relates the skills and strategies associated with developing reading comprehension to those associated with SRL (Dermitzaki, Andreou, & Paraskeva, 2008; Minguela, Solé, & Pieschl, 2015; Pressley & Ghatala, 1990; Solheim, 2011; Zimmerman & Kitsantas, 2007).

SRL is a goal-driven process by which the learner monitors and regulates internal abilities (what they can do) and responses to negotiate external environments. It encompasses a teachable set of skills and strategies that can be explicitly taught or learned though observations or modelling (Boekaerts & Corno, 2005; Zimmerman, 2002; Zimmerman & Schunk, 1989).

Self-regulation in learning involves the deliberate selection and use of strategies to set direction and understand and plan processes that mediate between person, context and achievement. That is, the learner develops strategies for negotiating distractions and input from others in the classroom, the classroom environment (including school culture) and plan for successful learning.

Teaching SRL involves giving students more control over their learning while encouraging a general awareness of motivation that underlies engagement and provides reasons for individuals' selection of learning strategies. In reading comprehension, instruction and modelling relate directly to monitoring and control of reading strategies; awareness, self-evaluation and calibration of a student's own skills and abilities; the ability to analyse a task and determine expectations and requirements and then match and adapt resources; selecting the process best suited for each situation; and identification of a mediation of motivation are all associated with deep comprehension (Minguela et al., 2015). Awareness and consideration of personal and contextual variables such as student ability and experiences with reading, exposure to texts, student motivation and valuing of reading are among the individual factors whose effects can be influenced by strategic implementation of skills related to self-regulation. Examples include the student's ability to self-reflect and self-motivate, developing an interest in reading and setting realistic goals for developing vocabulary and comprehension of genre and format.

High capacity students can engage with complex texts and activities that often demand higher and more sophisticated levels of reading comprehension and SRL. They can also engage with less complex texts using more complex and higher levels of comprehension. Skilful self-regulated readers systematically adapt their reading strategies to independently meet the demands of changing personal and contextual situations and tasks.

The teacher can support students in their development of SRL strategies to improve reading comprehension by

- discussing specific SRL strategies (such as goal setting, reflection, self-management of on/off task behaviour) and how these relate to reading strategies;
- modelling SRL strategies within reading contexts;
- providing access to reflective dialogue (including strategies for self-talk) and read-aloud strategies;
- providing and modelling corrective feedback to encourage self-correction;
- helping students develop and/or adapt strategies for making connections between abstract concepts within and across texts;
- helping learners automatically link new experiences with prior learning;
- requiring learners to reflect, evaluate and record successful reading strategies and processes for mitigating unsuccessful strategies.

In the current project, the following responses demonstrate how teachers taught SRL skills to their students.

- "I teach students to identify when they are on and off task . . . how to appropriately ask for help and assistance . . . how to speak to their peers that are being distracting"
- "Setting clear performance standards, discussing good study habits, giving students some choice . . . goal-setting, giving self-reflection opportunities"
- "Explicitly discussing and modelling behaviours conducive to self-regulated learning (SRL). Supporting students to think and record SRL behaviours. Supporting students to identify how these behaviours affect their learning"

- “Monitoring thinking, reviewing and critiquing own work, self-management of behaviour, making positive choices for learning (e.g. having a clear workspace, moving away from distractions)”
- “Goal setting by students and providing clear pathways on where to next with achievable short term goals for students”
- “Goal setting so students know what to learn and how to learn and how to plan their approach to learning”
- “Assessment rubrics and success criteria for goal setting . . . and self-evaluation”
- “Using reflection tools, personalised learning (plans), independent (negotiated) investigations, trans-disciplinary skill foci”

Summary

- High growth in reading comprehension for high capacity students was found in classes where teachers reported, including targeted teaching strategies, strategies and opportunity for SRL, and strategies and opportunity for engaging students in higher order thinking.
- Targeting questions towards higher order skills can support the development of higher order thinking, which can in turn increase students’ ability to comprehend and interpret high order texts.
- Planning and auditing lessons for teaching that is appropriately targeted for high capacity students is important for their growth and progression in reading comprehension.
- The inter-relatedness of reading comprehension and students’ SRL behaviour suggests that skills in one can be used to support the development of skills in the other.
- Students demonstrated higher growth when provided with tasks and texts aimed at their level of ZPD, targeting their learning needs.

Teacher Reflection

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

1. Review the evidence-based teaching practices discussed and choose one strategy that you could implement in your teaching of reading comprehension. Create a planning document for trialling this strategy (when, where, how, what resources do you need etc.). Reflect on the success of the strategy after having implemented it for a chosen period.
2. How could you or do you modify the content of teaching and learning activities for high capacity students when teaching reading comprehension? Discuss with at least two colleagues how they modify work for high capacity students.
3. Why is it important to modify questioning techniques according to student level of ability? Design an activity or series of questions that cover three or more reading comprehension levels in your class – use the examples provided in Table 5 and Table 6 to help you.
4. Select an upcoming area of study in reading comprehension and determine the related self-regulated learning skills that could be included in that lesson or series of lessons.
5. How did reading these ideas change or impact your perspective on teaching strategies for high capacity students in reading comprehension?
6. How and when might you share your learning with other colleagues?

References

Afflerbach, P., Cho, B.-Y., & Kim, J.-Y. (2015). Conceptualizing and assessing higher-order thinking in reading. *Theory into Practice, 54*(3), 203–212.

Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives: Complete Edition*. New York: Longman.

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.

Blair, T. R., Rupley, W. H., & Nichols, W. D. (2007). The effective teacher of reading: Considering the "what" and "how" of instruction. *The Reading Teacher, 60*, 432-438. doi: 10.1598/RT.60.5.3

Bloom, B. (1984). *Taxonomy of educational objectives.* Boston, MT: Allyn & Bacon.

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology, 54*(2), 199-231.

Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom.* Alexandria, VA: ASCD.

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice, 30*(1), 3–12.

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

Coiro, J. (2003). Exploring literacy on the internet: Reading comprehension on the internet: Expanding our understanding of reading comprehension to encompass new literacies. *The Reading Teacher, 56*(5), 458–464.

Dermitzaki, I., Andreou, G., & Paraskeva, V. (2008). High and low reading comprehension achievers' strategic behaviors and their relation to performance in a reading comprehension situation. *Reading Psychology, 29*(6), 471–492.

Duke, N., & Pearson, P. (2008). Effective practices for developing reading comprehension. *Journal of Education*, 107–122.

Dymock, S., & Nicholson, T. (2010). "High 5!" Strategies to enhance comprehension of expository text. *The Reading Teacher, 64*(3), 166-178.

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

Griffith, R., Bauml, M., & Barksdale, B. (2015). In-the-Moment Teaching Decisions in Primary Grade Reading: The Role of Context and Teacher Knowledge. *Journal of Research in Childhood Education, 29*(4), 444-457.

Harding., S., Griffin., P., Graham., L., Arnup., J., Szymakowski, J., English., N., . . . & 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, Australia.

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

Hess, K., Carlock, D., Jones, B., & Walkup, J. (2009, updated 2013). What exactly do "fewer, clearer, and higher standards" really look like in the classroom? Using a cognitive rigor matrix to analyze curriculum, plan lessons, and implement assessments. In Hess's Local Assessment Toolkit: Exploring Cognitive Rigor. Retrieved from https://www.uen.org/literacyresources/downloads/linking_research_with_practice.pdf

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.

Jones, B., & Flannigan, S. L. (2006). Connecting the digital dots: Literacy of the 21st century. *Educause Quarterly, 29*(2), 8–10.

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.

McGee, C., Ward, R., Gibbons, J., & Harlow, A. (2004). *Transition to secondary school: A literature review*: Report to the Ministry of Education, New Zealand, University of Waikato, Hamilton.

Minguela, M., Solé, I., & Pieschl, S. (2015). Flexible self-regulated reading as a cue for deep comprehension: Evidence from online and offline measures. *Reading and Writing, 28*(5), 721–744.

Nibali, N., Harding, S-M., Graham, L., & Griffin, P., (2017). When the apprentice surpasses the master: Teacher content knowledge relative to their students. *European Association for Research on Learning and Instruction (EARLI),* Poster presentation. Tampere, Finland.

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

Peterson, D. S., & Taylor, B. M. (2012). Using higher order questioning to accelerate students' growth in reading. *The Reading Teacher, 65*(5), 295–304.

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, *48*, 85-112.

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.

Pressley, M., & Ghatala, E. S. (1990). *Self-regulated learning: Monitoring learning from text. Educational Psychologist, 25*(1), 19–33.

Resnick, L. B. (1987). The development of mathematical intuition. In M. Perlmutter (Ed.),

Minnesota symposium on child psychology, 19, 159–194. Hillsdale, NJ: Erlbaum.

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

Solheim, O. J. (2011). The impact of reading self-efficacy and task value on reading comprehension scores in different item formats. *Reading Psychology*, 32(1), 1–27.

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).

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/

Tobin, K. (1987). The role of wait time in higher cognitive level learning. *Review of Educational Research, 57*(1), 69–95.

Topsfield, J. (2013). 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

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.

Webb, N. L. (2002, 28 March). Depth of knowledge levels for four content areas. Unpublished manuscript. Wisconsin Center for Education Research, University of Wisconsin-Madison, Madison, WI. Retrieved from http://www.hed.state.nm.us/uploads/files/ABE/Policies/depth_of_knowledge_guide_for_all_subject_areas.pdf

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice, 41*(2), 64–70.

Zimmerman, B. J., & Kitsantas, A. (2007). A writer's discipline: The development of self-regulatory skill. *Writing and Motivation, 19,* 51.

Zimmerman, B. J., & Schunk, D. H. (1989). *Self-regulated learning and academic achievement*. New York, NY: Springer Science & Business Media.