Applying the Rasch Growth Model (RGM) for the evaluation of achievement trajectories

Leigh Patterson[1]

[1]Australian Curriculum, Assessment and Reporting Authority

April 07, 2022

Abstract

Considerable interest lies in the growth in educational achievement that occurs over the course of a child's schooling. This paper demonstrates a simple but effective approach for the comparison of growth rates, drawing on a method first proposed some 80 years ago and applying it to data from the Australian National Assessment Program. The methodology involves the derivation of a 'meta-metre' – a quantitative mode of variation in growth – which permits comparison between groups defined by time-invariant characteristics. Emphasis is placed upon the novel characteristics of the method and the valuable information it can provide. Unlike complex modeling procedures, the approach provides a parsimonious, easily-interpretable model of growth.

APPLYING THE RASCH GROWTH MODEL (RGM) FOR THE EVALUATION OF ACHIEVEMENT TRAJECTORIES

by

Leigh Cameron Patterson

(Australian Curriculum, Assessment and Reporting Authority)

April 2022

Acknowledgements

The idea for the application of the Rasch Growth Model to National Assessment Program – Literacy and Numeracy (NAPLAN) data took root following a presentation to the Australian Curriculum, Assessment and Reporting Authority (ACARA) by David Andrich in July 2020. This presentation discussed the measurement of growth by the Danish Mathematician Georg Rasch. Rasch had applied his approach to growth of physiological variables, such as weight, which have a natural origin. Andrich showed how Rasch’s approach could be adapted to measured variables with an arbitrary origin, such as those derived using the measurement model carrying his name.

Correspondence between the author and Andrich continued throughout the period of the development of the manuscript, with particular assistance provided in the derivation of the meta-metre. I extent my thanks and acknowledgement to David for his support and encouragement.

Abstract

Considerable interest lies in the growth in educational achievement that occurs over the course of a child's schooling. This paper demonstrates a simple but effective approach for the comparison of growth rates, drawing on a method first proposed some 80 years ago and applying it to data from the Australian National Assessment Program. The methodology involves the derivation of a 'meta-metre' – a quantitative mode of variation in growth – which permits comparison between groups defined by time-invariant characteristics. Emphasis is placed upon the novel characteristics of the method and the valuable information it can provide. Unlike complex modeling procedures, the approach provides a parsimonious, easily-interpretable model of growth, appropriate for a specific set of applications.

Introduction

Recent Australian government, public policy and academic recommendations have highlighted the need to measure student achievement over the course of schooling, rather than concentrating on a specific timepoint (e.g., Gonski et al., 2018; Goss, Emslie, & Sonnemann, 2018; Masters, 2020; McGaw, Louden, & Wyatt-Smith, 2020).

Within the Australian educational landscape, the focus of school-age reporting of achievement has traditionally centred on performance at a single timepoint as measured by the National Assessment Program – Literacy and Numeracy (NAPLAN; Gonski et al., 2018). While such approaches provide meaningful information on educational achievement relative to the population, they fail to address variation in the level of growth for students with differing levels of academic aptitude. They also mask the extent to which student proficiency increases relative to previous performance, particularly for students entering school with low levels of literacy and numeracy. Interest thus lies in a shift away from static measures to a focus on achievement over the course of a child's schooling. Such views are highlighted in several recent reports (e.g., Goss et al., 2018; Masters, 2020; McGaw, Louden, & Wyatt-Smith, 2020) and emphasise a need to focus on variation in levels of achievement in early years of schooling and the subsequent progress that takes place over time – features respectively referred to as a student's 'initial status' and 'growth rate'.

One area of interest raised by Masters (2020) is variability in growth between students from different backgrounds, with the possibility of increasing equity and inclusivity through targeted interventions. To implement such recommendations, methods appropriate for analysing growth between groups are required. Growth modelling approaches that accommodate varying starting points and trajectories, such as the RGM, are a logical consideration.

To effectively measure the rate of growth across schooling, appropriate methods are required. Growth-oriented methodological approaches accommodate varying starting points and trajectories but are often complex in their specification and interpretation (Curran, Obeidat, & Losardo, 2010). An alternative methodology, first proposed some 80 years ago, involves the derivation of “salient features of growth” (Rao, 1958, p.1). This paper demonstrates how the application of this approach to a set of longitudinal student assessment data can provide easily interpretable statistics that lend themselves to effective presentation of results. The approach provides valuable information regarding the overall trajectories of groups of students and demonstrates utility in the evaluation of achievement over time.

The Rasch Growth Model (RGM) proposes a mechanism for evaluating differences in growth across a variety of phenomena (Olsen, 2003). One application for which this approach may be appropriate is in the modelling of achievement trajectories for school-age children, an area for which Rasch’s probabilistic models are most commonly associated (e.g., Rasch, 1961). Since children’s literacy and numeracy abilities vary when they enter schooling and as a result of learning through schooling, it is feasible to measure and evaluate differences in these two features. To do so, one typically seeks to manifest a construct of interest through a set of assessment items (i.e., to elicit a behaviour which demonstrates the psychological phenomena, such as a performance of reading ability demonstrated through a reading test) and subsequently models the variation in change over time by applying a set of methodological approaches commonly referred to as growth modelling (Williamson, 2016). In doing so, the trajectories of achievement for individuals and groups may be compared.

Higher levels of academic achievement and education are associated with well-documented economic and social benefits, including individual and aggregate improvements in health outcomes (Grossman, 2006), civic engagement and social cohesion (Dee, 2004), and distribution of income and earnings (Hanushek & Wößmann, 2007). Research points to educational advantages associated with the tailoring of pedagogical and policy-led interventions based on varying growth trajectories (Williamson, Fitzgerald & Stenner, 2014). As a result, monitoring achievement over the course of schooling plays an important role in informing both and educational and fiscal policy.

The following paragraphs provide a summary of the interdisciplinary features of the RGM, its relationship to educational assessment and measurement, and its embedding within Rasch's more widely recognised work. This approach is then described and applied, with the statistical and descriptive features of the model demonstrated using examples from NAPLAN.

The Rasch Growth Model

The RGM was first proposed by Georg Rasch in 1940 and subsequently articulated in a series of lectures presented at the 1951 meeting of the International Statistical Institute (Olsen, 2003). In both instances, Rasch's focus was on the physiological growth of animals and an attempt to derive a "simple elementary growth law… [with] time expressed in the physiologically adequate unit of time" (Olsen, 2003, p.65). Rasch later applied the same methodology in the analysis of economic data, at which time he emphasised the key statistical features associated with the model (Rasch, 1972). In each application of the RGM, irrespective of the domain to which it was applied, the core feature lay in the capacity to derive the primary features of growth for the purpose of efficient comparisons between groups (Rao, 1958).

The RGM approach attempts to derive an estimated rate of growth for each individual, proportional to the increase over time for the population (Rao, 1958). Underlying this is an assumption that observations represent functional changes alongside a continuous variable, time. Methods of this type provide a parsimonious summary of individual differences (McArdle & Nesselroade, 2003). The RGM aims to identify the principal sources of variation in growth, which are conveniently expressed as a set of derived variables (Jolliffe & Cadima, 2016). Such an approach is consistent with time-ordered analyses that incorporate and acknowledge the role of both individual and group-level differences (Duncan & Duncan, 1995). This has relevance within the context of evaluating differences in developmental trajectories, a point highlighted by Meredith and Tisak (1990) in their call for wider recognition of such procedures by both biological and behavioural researchers.

Educational Assessment and Measurement

Assessment and measurement act not only as a mechanism for obtaining results, but for further understanding the nature of development. Effective assessment development commences with an understanding of how an attribute will influence performance (Borsboom, Mellenbergh, & van Heerden, 2004). By focusing upon a variable of interest, quantification is considered appropriate in some contexts to the extent that it provides opportunities for generalisations to be made regarding underlying processes. For example, approaches that permit the inclusion of large or census samples, more defined criteria for differentiating performance (i.e., the use of multiple items and explicitly-defined rubrics), and a minimisation of construct-irrelevant variance, alongside a defined assessment framework, aid the determination of a clear construct definition. If one adheres to the notion of learning as a cumulative process, it can be inferred through the use of statistical techniques that differing degrees of development can be represented as varying points along a continuum. Different locations on such a continuum can then be used to represent the quantitative, comparable abilities of those undergoing assessment, varying with respect to the measured variable.

In an assessment context, tasks are required for test takers to demonstrate their ability in a specific area of inquiry. Assessments must permit the accurate measurement of variation in ability if they are to be considered valid (Borsboom et al., 2004). To achieve this, items are devised and selected to the degree that they represent the underlying construct (Andrich, 1988). Variable construction is instantiated through the creation of test items that elicit evidence of the psychological phenomena (i.e., a demonstrated manifestation of the construct), which further assists in the development of the operational definition (Wright and Stone, 1979). The aim of the test development process is to devise and select a set of items suitable for elucidating variation in the degree of ability required from the measurement procedure, and to represent those differences as locations on a continuum. In effective test construction procedures, items are selected in a way that permits differentiation between the various levels of ability – that is, a test is composed of items appropriately targeted to the population of interest. When a person interacts with an item, the theoretical construct is made explicit through the manifestation of an observed response. Inferences can then be made regarding the level of ability exhibited in the set of responses provided (Bond & Fox, 2015). Ensuring an appropriate level of differentiation among test takers – based on the functional definition of the variable and the capacity of the items to manifest it – is an essential component of measurement in the social sciences (Andrich, 1988).

The Rasch Measurement Model

The Rasch Measurement Model (RMM) serves a comparative function in which data derived from a testing instrument (e.g., an assessment) can be compared to expectations set under fundamental principles of measurement (Andrich, 2004). These fundamental principles relate to the specific structure of relations between attributes and those relations being wholly attributable to the measurement act itself (i.e., not derived from other measures). Such properties permit a comparison in the degree of differences between two measures (i.e., additivity), but not in the ratio of them (i.e., multiplicativity). In this way, such relations may be considered similar to what Stevens (1946) described as ‘interval scale measurement’. Such features are common to the physical sciences (e.g., temperature as measured in Fahrenheit or Centigrade) and readily permit analysis using linear statistics.

The RMM imposes a priori restrictions on both the model and parameters used to account for the observed structure of data (Andrich & Marais, 2019). In this way, a data set – the numerical summarisation of the qualitative aspects of a testing instrument – can be said to meet the requirements for measurement when it conforms to the structure specified by the RMM, thus permitting quantitative conclusions (Duncan, 1984). It may be argued that the methodology applies an approach consistent with that espoused by Kuhn (1977), in which the merit of the procedure lies not in appraising the appropriateness of a set of models to fit the data, but instead assessing whether observed data suitably represent features expected under fundamental measurement through evaluation of conformity with a pre-specified model.

In the dichotomous RMM, the probability that a person will respond correctly to an item is dictated by the interaction between the person taking the test and the item used to measure the underlying construct. The person’s estimated level of ability determines the likelihood they will respond correctly to the item, given the item’s level of difficulty. This is achieved by calculating the difference between ability and difficulty under certain algebraic constraints. Such an estimate of ability can be conceptualised as the individual's location along a trait continuum, varying according to their capacity.

A test conforming to the RMM can be used to ascertain an estimate of ability on a construct that is independent of the specific set of items used to make that assessment. This fundamental feature of the RMM – known as specific objectivity – states that the comparison of two people should be independent of the items used to assess them, and similarly that the comparison of two items should be independent of the individuals used for the comparison (Rasch, 1977). The algebraic separation of person and item parameters underlies this notion, ensuring that each parameter can be eliminated in its counterpart's estimation. Interestingly, this same feature – consistent with requirements for objective measurement – is present within the RGM (Olsen, 2003).

Measuring Growth in Educational Achievement

The RMM can be implemented to evaluate whether quantities under investigation conform to fundamental properties of measurement when growth trajectories are estimated (Williamson, 2017). By applying the RGM to valid and reliable measures of educational achievement that meet the requirements of the RMM, an attempt can be made to derive a set of summative parameters that characterise the growth trajectories of both individuals and groups. This is undertaken under the assumption that performances observed from psychometrically-sound measures of educational achievement represent functional changes associated with skill acquisition through structured learning (i.e., time spent at school). Such an assertion is consistent with developmental theories that posit an asymptotic decrease in the rate of educational achievement over time (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996). However, like all approaches that are guided by substantive theory, proposed models require subjection to empirical evaluation through data analysis (Williamson, 2016).

Aim, Objective and Research Question

This paper addresses the following question: Can Rasch's Growth Model be used to measure differences in achievement trajectories by representing growth as a function of time? Meaningful application of the RGM will be demonstrated through the use of examples comparing the initial status and growth rates of groups defined by time-invariant characteristics, emphasising the valuable information about achievement trajectories that can be derived using this model.

Methodology

Data Sources

NAPLAN provides annual, point-in-time information regarding student achievement across a variety of domains at the level of the student, school, states and territories, and Australia as a whole. NAPLAN assessments are completed by students in grades three, five, seven and nine. The information collected as part of the program facilitates the monitoring and reporting of performance of specific groups and across states and territories – herein referred to as 'jurisdictions'. While NAPLAN assesses performance across both primary and secondary education, the focus currently lies on reporting levels of achievement based on observations at a single timepoint, with limited supplementary reporting at two timepoints – between Grades three and five and between Grades seven and nine (Australian Curriculum, Assessment and Reporting Authority, 2013). By adopting a longitudinal approach to data analysis that incorporates individual-level data measured across all timepoints, a more cogent understanding of educational achievement trajectories can be developed.

Measures

The measures of interest – student achievement in reading – took the form grade-specific weighted-likelihood estimates (WLEs; Warm, 1989) of reading ability measured in logits, derived using the RMM (Australian Curriculum, Assessment and Reporting Authority, 2020). This fulfilled the requirement stipulated by Williamson (2017) that fundamental properties of measurement be attributable to base quantities when growth trajectories are estimated. These estimates were equated onto the NAPLAN reporting scale, which became the unit of analysis. The NAPLAN reporting scale spans across all tested grades (i.e., Grades 3, 5, 7 and 9) and is standardised using a mean of 500 and standard deviation of 100, with scores ranging from approximately 0 to 1000 (Australian Curriculum, Assessment and Reporting Authority, 2020). This process places reported results across assessment years (e.g., 2013, 2015, 2017, 2019) on the same scale, thus permitting comparability.

Data Preparation and Processing

Prior to model implementation, data appropriate for longitudinal analysis was sourced and prepared. Separate data sets containing matched ‘gain’ data[1] from the 2013, 2015, 2017 and 2019 NAPLAN assessments were provided to the author for the purpose of data linkage and analysis. As current reporting and analysis techniques do not necessitate the linkage of student data across NAPLAN assessments, a considerable degree of variability exists in the consistency and use of common student identifiers. As a result, a small selection of the large number of variables pertaining to each student were used in the matching process. This was undertaken with the view to maximise the likelihood of a correct correspondence in assessment information included across timepoints.

Data was first partitioned such that only cases from the appropriate grade and year were retained (i.e., only Grade 3 data was retained from the 2013 dataset, only Grade 5 data was retained from the 2015 dataset, etc.). The following variables were subsequently used in the matching process:

- Student Identifier[2];
- Date of Birth;
- School Identifier associated with the 2017 (Grade 7) and 2019 (Grade 9) dataset[3];
- Previous NAPLAN Reading score.

---

[1] ‘Gain’ data refers to a two-timepoint data set containing matched student information. These data sets contain information for students in all grades (i.e., three, five, seven and nine).

[2] There are no common controls in place for the establishment and allocation student identifiers used in NAPLAN assessments, which are managed by the jurisdiction in which the student sits the assessment. As a result, there are known issues in the sole use of student identifiers for the purposes of data linkage.

[3] Except in one jurisdiction, where Grade 7 students attend primary rather than high school, thus commonly attending a different school in Grade 9.

As the RGM requires complete data across all timepoints of interest, cases with missing data were removed. There was considerable variability in the number students with complete data across all four timepoints in each jurisdiction. To address these imbalances and potential privacy-related issues, a senate-weighted, random sampling procedure was then undertaken in which 1000 students from each jurisdiction with complete data were selected. Due to issues associated with data access and its linkage, data from one jurisdiction was excluded, resulting in a total of 7000 students being included in the analysis sample.

Three time-invariant variables were selected for further analysis – jurisdiction, sex and Aboriginal and Torres Strait Islander identification (ATSI). Jurisdictional data was re-categorised using an anonymised, numerical identifier. Sex data was categorically coded ‘1’ and ‘2’ as per the original data collection definitions (Australian Curriculum, Assessment and Reporting Authority, 2012), with the decision made to not elucidate the associated groups. The same data definitions (Australian Curriculum, Assessment and Reporting Authority, 2012) differentiated students who identified as:

- Aboriginal but not Torres Strait Islander origin;
- Torres Strait Islander origin but not Aboriginal origin; or,
- Both Aboriginal and Torres Strait Islander origin.

Due to the small number of students identifying in the affirmative across these categories, the decision was made to collapse students across these groups into a single, binary-coded variable, ATSI. Students with missing data on either the ATSI or sex variables were excluded from further analysis. Comparisons were subsequently made of descriptive statistics for each of the variables relative to the 2019 ‘gain’ data set, with approximate equivalence found between the two datasets, as reported in

Table *1*. Computation of the level of negative ‘gain’ – in which student progress decreases over the measures of interest – was also undertaken. During this process, distribution checks for concordance with assumptions of analysis of variance (ANOVA) – including normality of distributions as presented in the results section – were undertaken.

Table 1

Percentage of Students Belonging to Each of the Categorically Coded Time-invariant Variables of Interest

Data Set	% Sex 2 (% Sex 1)	% ATSI (% Not ATSI)
2019 ‘Gain’ data (*N* = 859 055[4])	49.8 (50.2)	5.9 (94.1)
Analysis Sample (*N* = 7 000)	51.5 (49.5)	6.6 (93.4)

Analytic Methods

The RGM proposes the existence of a quantitative mode of variation in growth, referred to as the ‘meta-metre’, that is common to all individuals, providing relevant information for the comparison of average growth curves (Rao, 1958). In deriving this ‘age transforming function’ as Rasch first described it (Olsen, 2003), we assert that the rate of growth of an individual is directly proportional to the meta-metre, thus allowing comparisons characterised by linear relationships (Rao, 1958). Such relationships can be expressed in

[4] As previously noted, ‘gain’ data contains students across all grades (i.e., three, five, seven and nine).

terms of a single growth parameter, retaining dynamic consistency, whereby individual and group level differences can be expressed by the same mathematical function (Keats, 1980). In this way, while there exists a trajectory for the population, each individual retains an estimate that varies relative to the group as a whole (McNeish & Matta, 2018). These variables are defined by the data, not a priori, allowing for the modelling of individual growth trajectories (Jolliffe & Cadima, 2016).

In the context of educational achievement, the RGM asserts that:

$$Y_{nt} = a_n + b_n \tau(t) \qquad (1)$$

where:

Y_{nt} is domain-specific achievement for student *n* at timepoint *t*;

a_n is the overall achievement over the time period of interest;

b_n is the rate of growth in achievement for student *n*; and,

$\tau(t)$ is the time transforming function, referred to by Rasch as the meta-metre (Olsen, 2003).

Equation (1) can be conceptualised as a single structural equation expressing growth rate, with coefficients a_n and b_n specifying the growth rate parameters (Stone, 2020). Unlike behavioural models which utilise multiple structural equations – typically requiring specialised software and knowledge for their specification and interpretation (Curran et al., 2010) – the RGM models the group-level effect on the dependent variable, given the meta-metre (Stone, 2020). Through this approach, the RGM asserts that the derived meta-metre should be equivalent for all individuals, allowing one to generalise growth in academic achievement and thus meaningfully represent deviations from it (Stone, 2020).

It is possible to estimate a meta-metre for the comparison of rates of growth (Andrich, Marais, & Sappl, 2022), by taking the mean measurements at each timepoint *t*:

$$(\textstyle\sum_{n=1}^{N} Y_{nt})/N = (\textstyle\sum_{n=1}^{N} a_{nt})/N + [(\textstyle\sum_{n=1}^{N} b_{nt})/N]\tau(t) \quad (2)$$

which can be expressed as:

$$\bar{Y}_{.t} = \bar{a}_{.} + \bar{b}_{.}\tau(t) \quad (3)$$

for the overall sample, and:

$$\bar{Y}_{.at} = \bar{a}_{.a} + \bar{b}_{.a}\tau(t) \quad (3a)$$

$$\bar{Y}_{.bt} = \bar{a}_{.b} + \bar{b}_{.b}\tau(t) \quad (3b)$$

for any specific subgroups (e.g., *a* and *b*) of interest, allowing free estimation of the meta-metre $\tau(t)$:

$$\bar{Y}_{.t} = \bar{a}_{.} + \bar{b}_{.}\tau(t)$$

$$\bar{Y}_{.t} - \bar{a}_{.} = \bar{b}_{.}\tau(t)$$

$$\hat{\tau}(t) = \frac{\bar{Y}_{.t} - \bar{a}_{.}}{\bar{b}_{.}} \quad (4)$$

by substituting $\hat{\tau}(t)$ from (4) into (1), as:

$$Y_{nt} = \hat{a}_n + \hat{b}_n\hat{\tau}(t) = \hat{a}_n + \hat{b}_n\frac{\bar{Y}_{.t} - \bar{a}_{.}}{\bar{b}_{.}} = \hat{a}_n + (\hat{b}_n/\bar{b}_{.})(\bar{Y}_{.t}) - (\hat{b}_n/\bar{b}_{.})(\bar{a}_{.})$$

$$= \hat{a}_n + (\hat{b}_n/\bar{b}_{.})(\bar{a}_{.}) - (\hat{b}_n/\bar{b}_{.})(\bar{Y}_{.t}) \quad (5)$$

Importantly, this can be expressed as:

$$Y_{nt} = \hat{A}_n + \hat{B}_n(\bar{Y}_{.t}) \quad (6)$$

where:

$$\hat{A}_n = \hat{a}_n - (\hat{b}_n/\hat{\bar{b}}_.)(\hat{\bar{a}}_{..}) \quad (7)$$

is the estimated normalised, relative achievement over time (i.e., initial status); and,

$$\hat{B}_n = (\hat{b}_n/\hat{\bar{b}}_.) \quad (8)$$

is the estimated normalised, relative rate of growth. This being due to:

$$\bar{A}_. = [\textstyle\sum_{n=1}^{N} A_n]/N = ([\sum_{n=1}^{N} a_n]/N) - 1_.(\bar{a}_.) = \bar{a}_. - \bar{a}_. = 0 \quad (9)$$

$$\hat{\bar{B}}_. = [\textstyle\sum_{n=1}^{N} \widehat{\bar{B}_n}]/N = ([\sum_{n=1}^{N} b_n]/N)/\bar{b}_. = \bar{b}_./\bar{b}_. = 1 \quad (10)$$

It is noteworthy that it is feasible to estimate $\hat{A}_n$ and $\hat{B}_n$ either by regressing Y_{nt} on $(\hat{Y}_{.t})$, in which $\hat{A}_n$ and $\hat{B}_n$ are dependent, or by estimating $\hat{B}_n$ independently of $\hat{A}_n$ and then estimating $\hat{A}_n$ (Andrich, Marais, & Sappl, 2022), as expressed by:

$$y_{nt} = Y_{nt} - Y_{n(t-1)} = A_n + B_n\bar{Y}_{.t} - [A_n + B_n\bar{Y}_{.(t-1)}], t = 1,2,3 \ldots T \quad (11)$$

And the estimate $\hat{y}_{nt}$ given by:

$$\hat{y}_{nt} = \hat{A}_n + \hat{B}_n\bar{Y}_{.t} - \left[\hat{A}_n - \hat{B}_n\bar{Y}_{.(t-1)}\right], t = 1,2,3 \ldots T = \hat{B}_n\bar{Y}_{.t} - \hat{B}_n\bar{Y}_{.(t-1)} = \hat{B}_n(\bar{Y}_{.t} - \bar{Y}_{.(t-1)})$$

$$= \hat{B}_n\bar{Y}_{.t} \quad (12)$$

where:

$$\hat{B}_n = \textstyle\sum_{t=1}^{T} y_{nt}\bar{y}_{.t} / \sum_{t=1}^{T} \bar{y}^2{}_{.t} \quad (13)$$

$$\hat{A}_n = \bar{Y}_{n.} - \hat{B}_n\bar{Y}_{..} \quad (14)$$

These parameter estimates $\hat{B}_n$ and $\hat{A}_n$ permit the comparison of individuals and groups of interest. The independent separation of these parameters, which incorporate the repeated measurements, subsequently allows the use of univariate procedures for significance testing. Independent samples *t*-test, one-way ANOVA and Tukey’s honestly significant difference post hoc test were used to reveal the significance and degree of differences between parameter estimates for the identified, time-invariant groups of interest. These procedures provided tests of the null hypothesis that rates of growth and their initial status were the same (Greenland et al., 2016).

The linear relationship between growth and time, conditional on the meta-metre, further allowed the representation of growth for separate groups to be presented as straight lines on a plot. These straight lines are easily interpretable as continuous changes, serving useful purposes in the identification of trends (Peebles & Ali, 2015). This approach avoids the issue of non-developmental behaviour (i.e., negative growth) characterised in quadratic representations (Williamson, 2016). As per the recommendations outlined in Nese, Lai and Anderson (2013), sets of growth plots, which permit more readily detected differences in growth, were subsequently used to present variation between multiple groups.

Differences in achievement may also be represented as a function of time, providing another effective and easily interpretable way to visualise growth trajectories (Singer & Willett, 2003). A feature consistent with the RGM is the capacity to transform time such that – by a common transformation – individual growth curves are linearised, with slight variations attributable to error (Rao, 1958). This can be achieved by taking each timepoint to its natural logarithm (i.e., the logarithm of each timepoint taken to the base of the constant *e*). To paraphrase Rasch, this process – utilising the meta-metre – allows one to measure time in a particular way, allowing a uniform description of growth curves for all individuals considered (Olsen, 2003).

Results

Estimates of initial achievement $\hat{A}_n$ and rate of growth in achievement $\hat{B}_n$ were calculated for all individuals within the sample, as well as for the overall sample and groups of interest. As outlined in the Analytic Methods section, the estimated normalised relative rate of growth for the overall sample had a mean of 1.0 (SD=0.45) and the estimated normalised achievement over time (i.e., initial status) had a mean of 0.0 (SD=251.41), consistent with expectations. Overall grade means evidenced an asymptotic decrease in the rate of growth over time (i.e., a decrease in the level of growth occurring between grades over time), prototypical of educational achievement trajectories, as shown in *Table 2*.

Table 2

Grade Means in NAPLAN Scale Scores

Grade 3	Grade 5	Grade 7	Grade 9
420.27	501.71	548.29	582.33

Differences in Jurisdictional Growth

Comparison of Jurisdictional Growth Parameters

Grade means for each jurisdiction evidenced a similar asymptotic decrease in the rate of growth over time as that evidenced overall, as shown in Table 3. The normalised relative growth rate and initial status estimates for jurisdictions are presented in

Table 4. A representation of the distribution of growth rates and initial status across jurisdictions is presented in Figure 1 and Figure 2 respectively.

Table 3

Grade Means for each Jurisdiction

Jurisdiction	Grade 3 Mean	Grade 5 Mean	Grade 7 Mean	Grade 9 Mean
1	419.30	501.06	549.37	594.89
2	441.70	517.22	562.39	596.01
3	419.76	498.11	543.97	572.01
4	417.20	493.41	547.16	579.77
5	418.03	505.50	548.43	579.89
6	377.81	467.18	516.44	551.93

7	448.10	529.48	570.25	601.79

Table 4

Comparison of Growth Rate and Initial Status for Jurisdictions

Jurisdiction	Growth Rate (SD)	Initial Status (SD)
1	1.05 (0.44)	-22.64 (251.41)
2	0.94 (0.44)	45.11 (239.40)
3	0.95 (0.46)	20.52 (259.34)
4	0.99 (0.43)	3.45 (242.03)
5	1.02 (0.42)	-12.22 (233.11)

6	1.08 (0.51)	-77.04 (271.88)
7	0.96 (0.43)	42.83 (239.39)

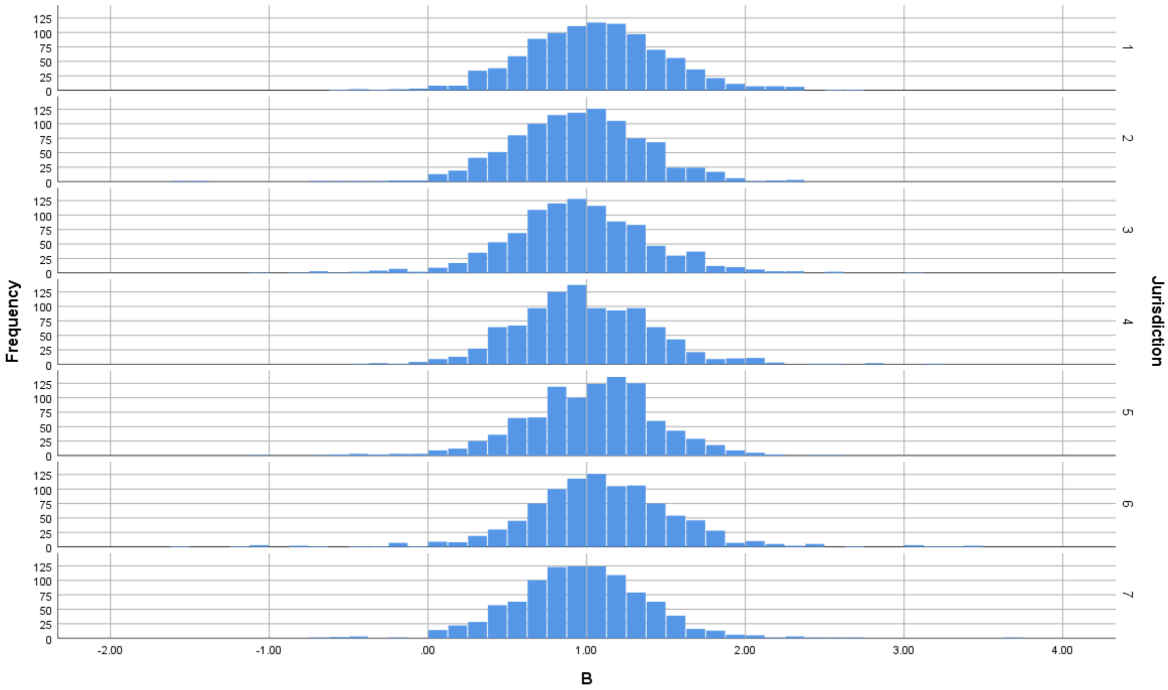

Figure 1 Distribution of growth rates across jurisdictions

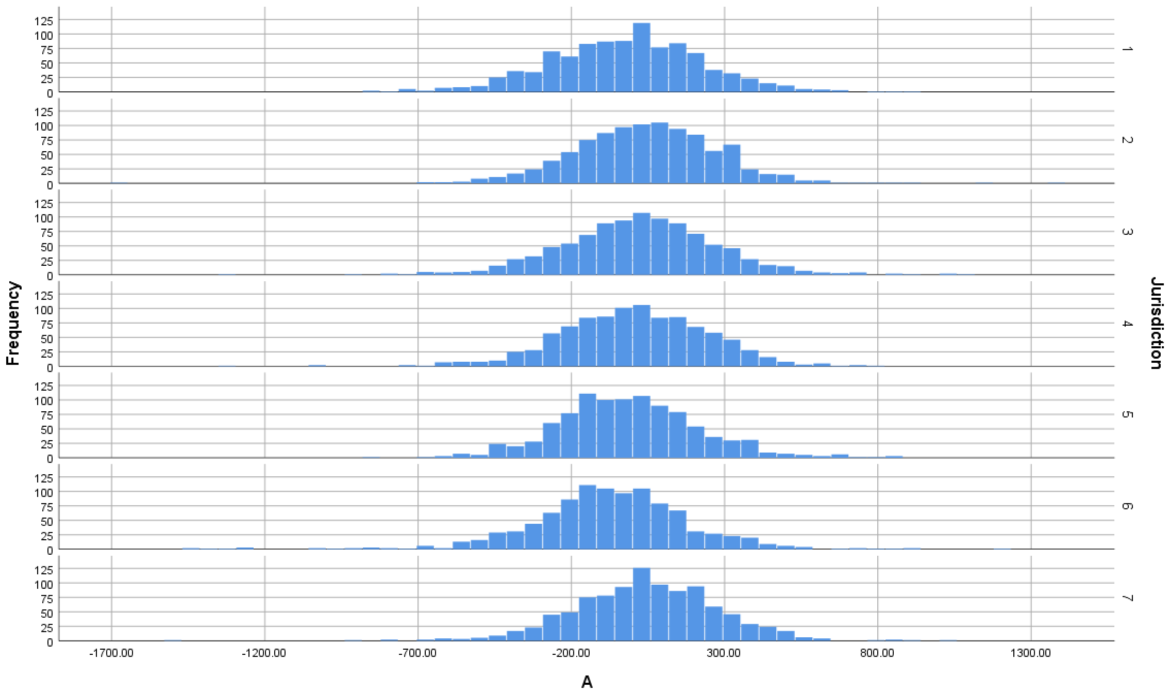

Figure 2 Distribution of initial status across jurisdictions

Noting the linearisation of growth trajectories and independent separation of parameters, one-way ANOVA was used to ascertain whether observed differences in jurisdictional growth rates and initial status were statistically significant. There was a statistically significant difference observed between jurisdictional growth rates ($F(6,6993) = 14.16, p < .05$). Tukey's honestly significant difference (HSD) post hoc test was applied to reveal the jurisdictions to which the statistically significant differences in growth rate applied and their degree. As shown in Table 5, 11 of the 21 paired comparisons between jurisdictions showed significant differences in growth rate, the largest being 0.14 units.

Table 5

Paired-comparisons in jurisdictional growth rate as determined by Tukey's HSD

	1	2	3	4	5	6	7
1	-						
2	0.11*	-					
3	0.10*	-0.01	-				
4	0.06*	-0.04	-0.04	-			
5	0.03	-0.08*	-0.07*	-0.04	-		
6	-0.03	-0.14*	-0.13*	-0.10*	-0.06	-	
7	0.09*	-0.02	-0.01	0.02	0.06*	0.12*	-

*Mean difference is significant at the 0.05 level

There was also a statistically significant difference observed between the initial status' of jurisdictions ($F(6,6993) = 29.43$, $p < .05$). Tukey's HSD post hoc test was applied to reveal the jurisdictions to which the statistically significant differences in initial status applied and their degree. As shown in Table 6, 13 of the 21 paired-comparisons between jurisdictions showed significant differences in initial status, the largest being 122.15 units.

Table 6

Paired-comparisons in initial status between jurisdictions as determined by Tukey's HSD

	1	2	3	4	5	6	7
1	-						
2	-67.75*	-					
3	-43.15*	24.59	-				
4	-26.09	41.66*	17.07	-			
5	-10.42	57.32*	32.73	15.67	-		
6	54.40*	122.15*	97.56*	80.49*	64.82*	-	
7	-65.47*	2.28	-22.31	-39.38*	-55.05*	-119.87*	-

**Mean difference is significant at the 0.05 level*

Visual Representation of Differences between Jurisdictions

It is conceivable to represent differences in means simply by regressing the grade means for each jurisdiction on the overall sample means, as displayed in Figure 3. These correspond to the values obtained from *Table 3* and *Table 2*.

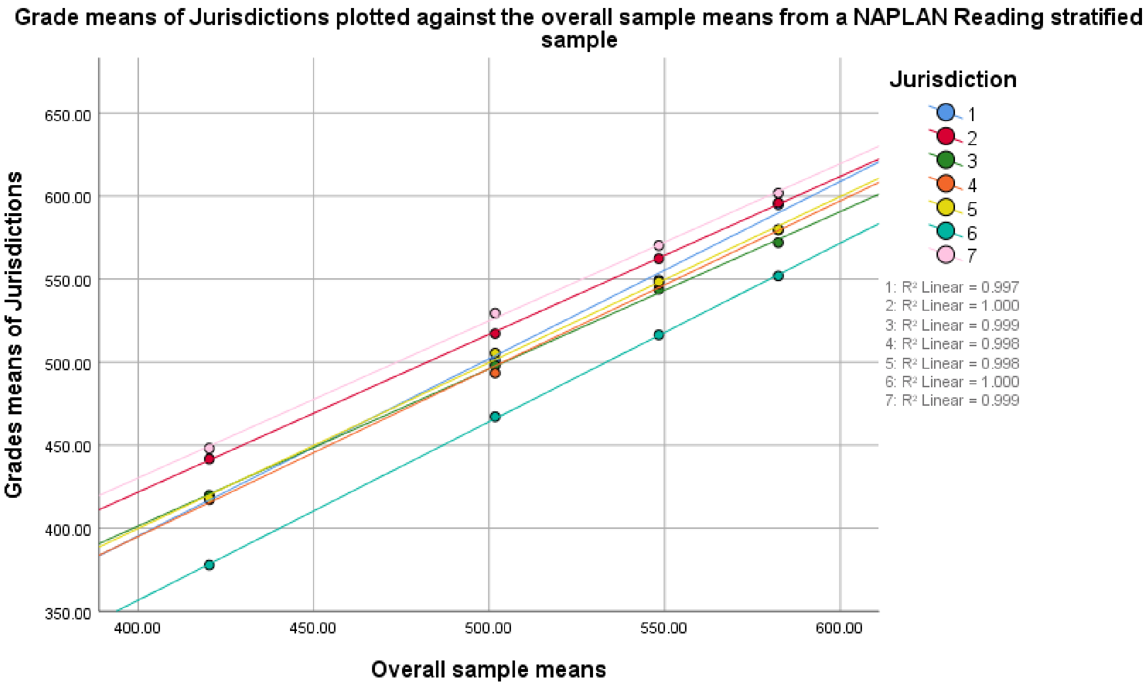

Figure 3 Grade means of each Jurisdiction plotted against the overall sample means

Alternatively, as described in the Analytic Methods section, the linear relationship between growth and time, conditional on the meta-metre, permits a representation of growth for separate groups as straight lines. Figure 4, Figure 5, Figure 6 and Figure 7 show the comparison of grade means for select pairs of jurisdictions, plotted against the transformed means. The line representing growth for each jurisdiction can be characterised by two parameters, identified within the equation on each of the plots, which are consistent with the growth rate and initial status estimates displayed in

Table 4 with their corresponding x-axes obtained as the transformation of the sample means obtained from *Table 3*. By presenting the comparison of a pair of jurisdictions in each plot, the individual trajectories of growth for each jurisdiction may be compared.

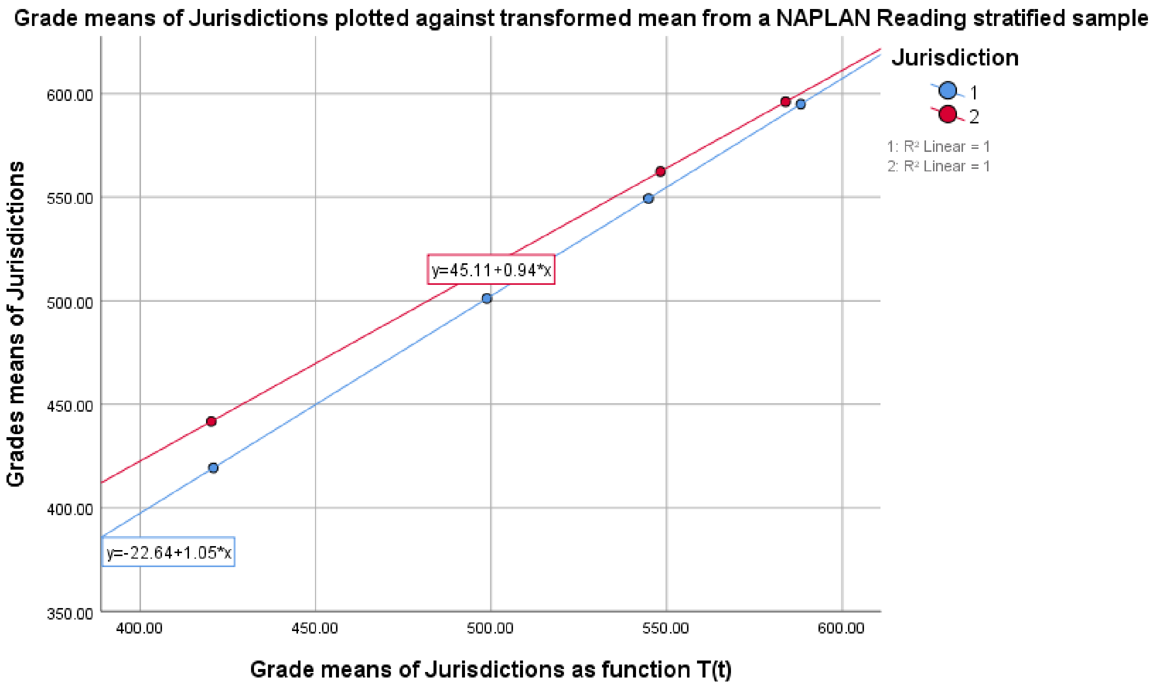

Figure 4 Grade means of Jurisdictions 1 and 2 plotted against the transformed means

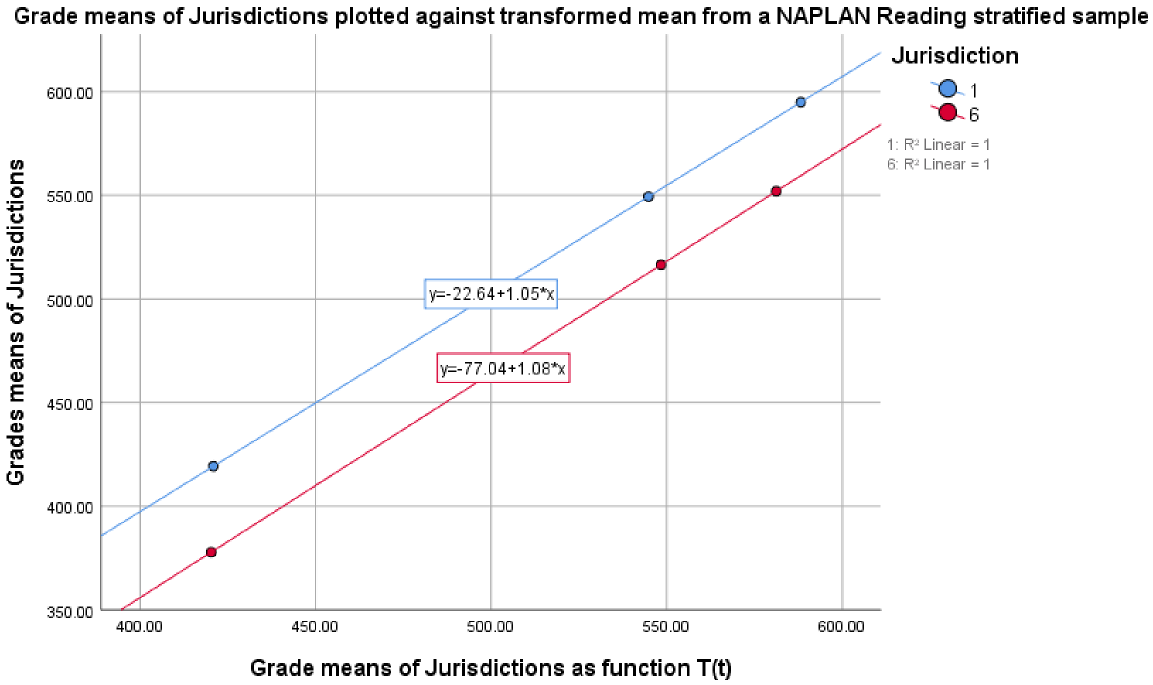

Figure 5 Grade means of Jurisdictions 1 and 6 plotted against the transformed means

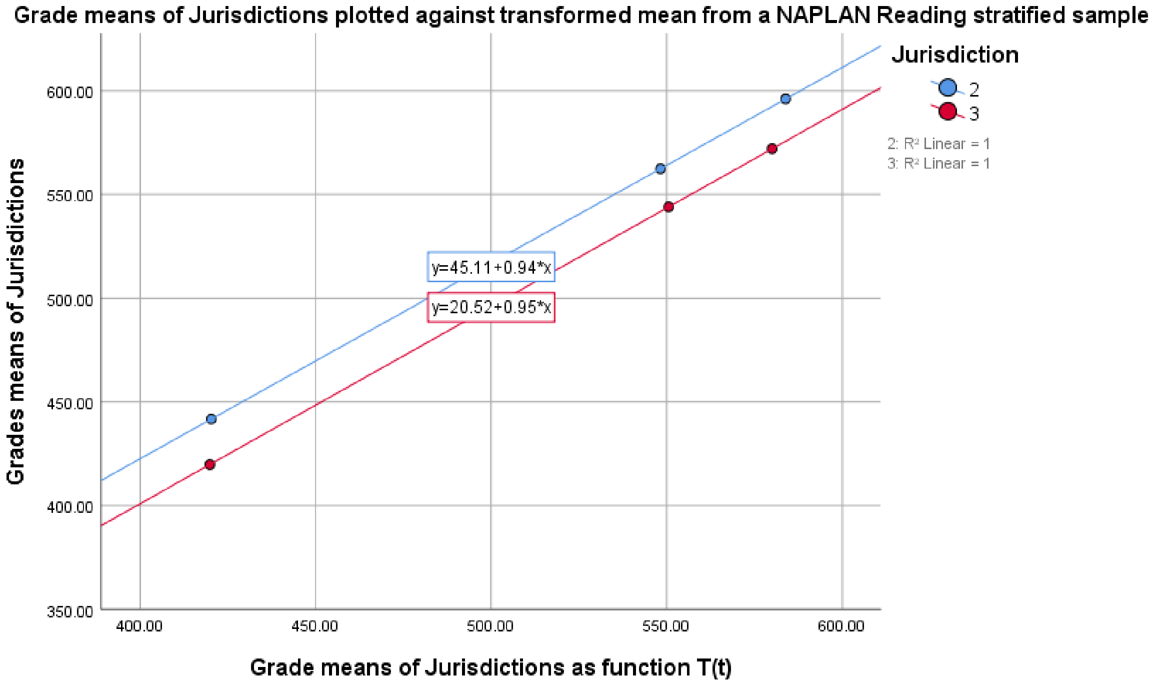

Figure 6 Grade means of Jurisdictions 2 and 3 plotted against the transformed means

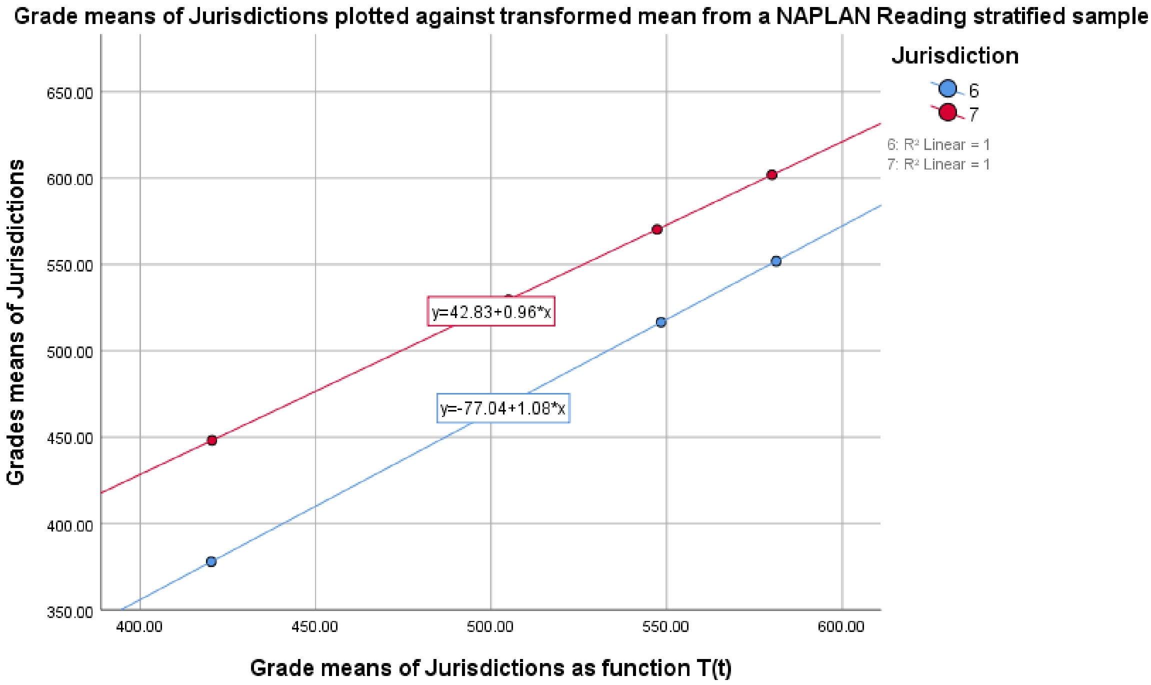

Figure 7 Grade means of Jurisdictions 6 and 7 plotted against the transformed means

Growth trajectories can similarly be represented as a function of time. Attempting to characterise the grade means of jurisdictions from Table 3 as function of linear time (i.e., across the four timepoints) resulted in poor model fit, evidenced by large residuals, and shown in Figure 8. An alternative approach is to fit a quadratic model representing a curvilinear functional form, thereby minimising residuals and increasing the variance explained by the model due to the monotonically decreasing rate of growth. This representation is shown in Figure 9. However, by taking the natural logarithm of each timepoint, this same information can be characterised linearly, with equivalent predictive power (i.e., variance explained, as R^2), as shown in Figure 10.

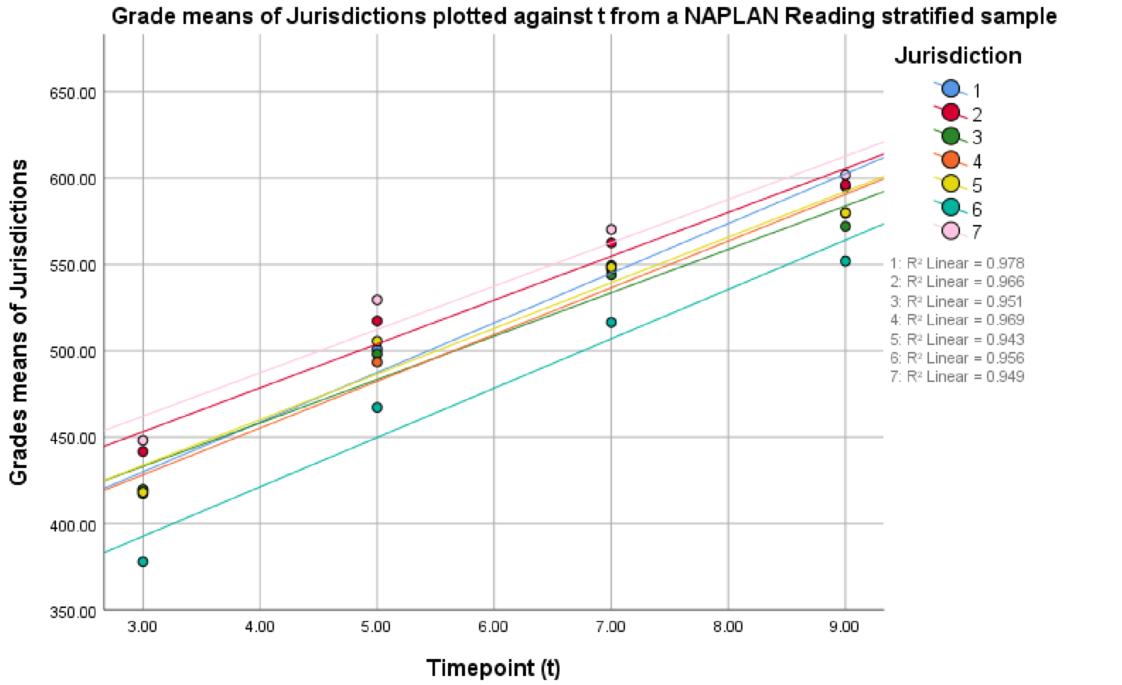

Figure 8 Grade means of each Jurisdiction plotted against time, assuming linear growth

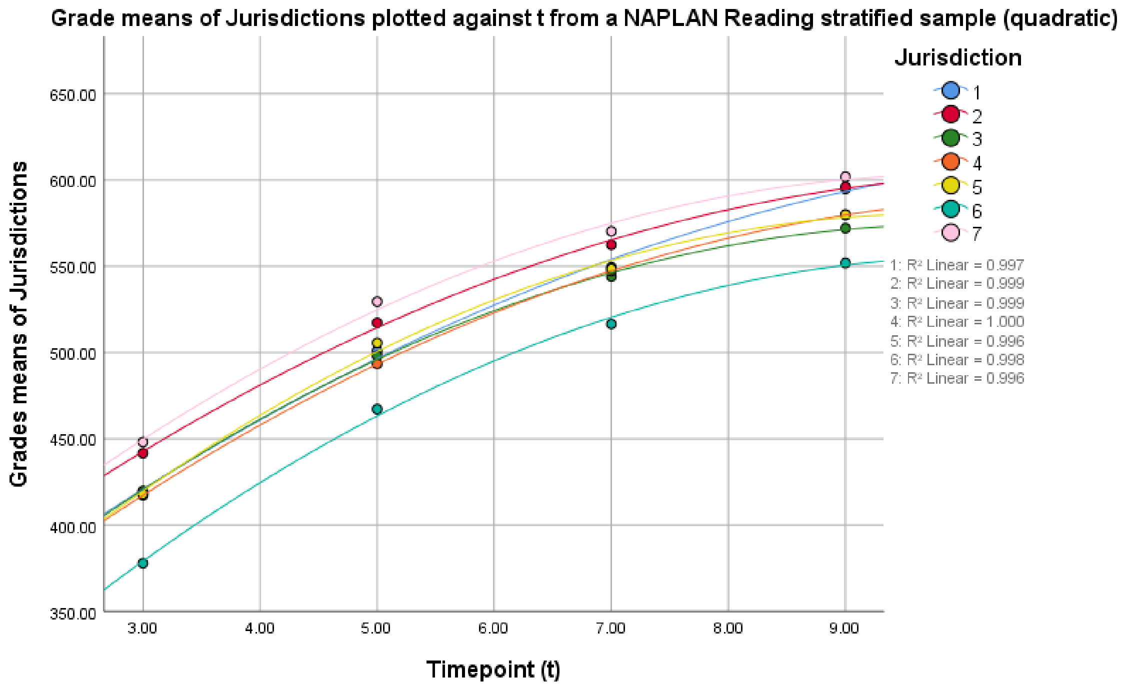

Figure 9 Grade means of each Jurisdiction plotted against time, assuming curvilinear (quadratic) growth

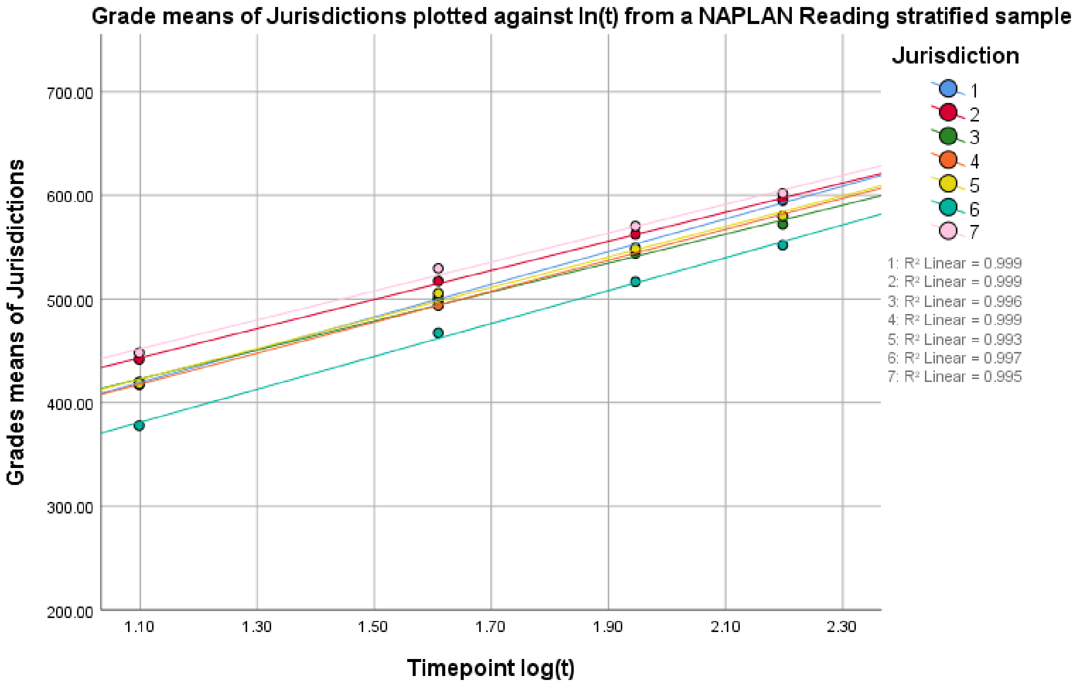

Figure 10 Grade means of each Jurisdiction plotted against time taken to its natural logarithm

A subsequent extension of these properties is the ability to extrapolate beyond measured timepoints. By taking the natural logarithm of each timepoint, the expected achievement can be inferred on the basis of the observed values. Through extrapolation to log-zero (i.e., grade 1) the varying degrees of achievement in initial status at a time approximating entry to school can be shown, as presented in Figure 11.

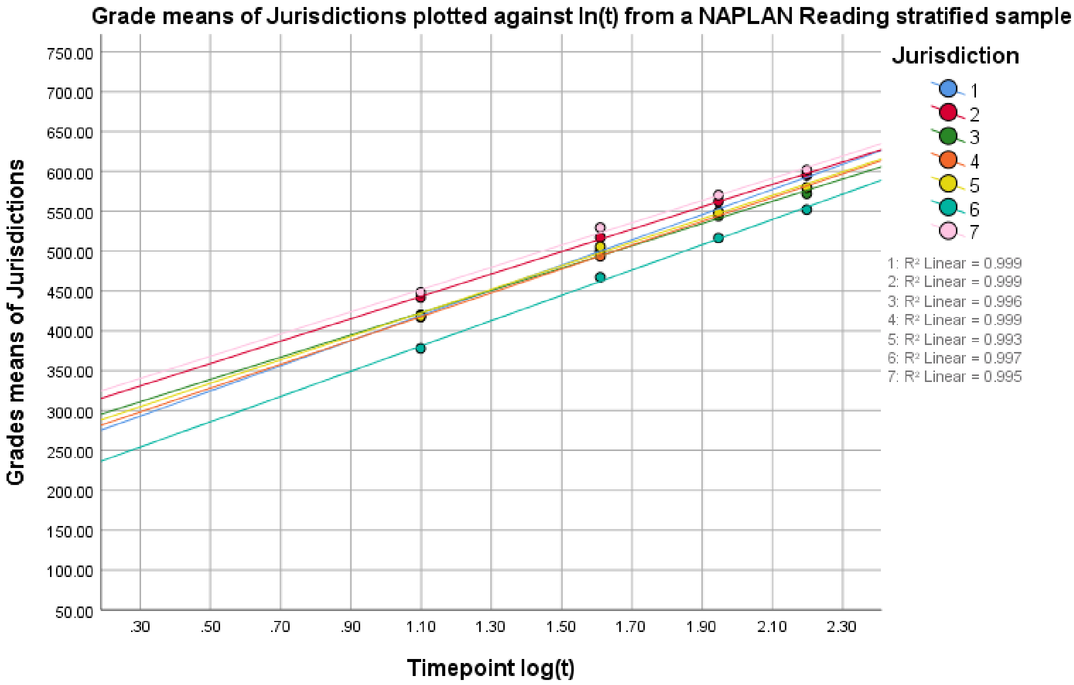

Figure 11 Extrapolated grade means of each Jurisdiction plotted against time taken to its natural logarithm

Differences in Growth between Sexes

Comparison of Growth Parameters for each Sex

As for each of the jurisdictions, grade means for each sex evidenced an asymptotic decrease in the rate of growth over time (i.e., a decrease in the level of growth in achievement progressing from earlier grades to later ones). These are shown in *Table 7*.

Normalised relative growth rate and initial status estimates for sex are presented in

Table 8. A representation of the distribution of growth rates and initial status across sexes is presented in Figure 12 and Figure 13 respectively.

Table 7

Grade Means for each Sex

Sex	Grade 3 Mean	Grade 5 Mean	Grade 7 Mean	Grade 9 Mean
1	412.48	496.59	543.33	576.00
2	427.90	506.72	553.14	588.52

Table 8

Comparison of Growth Rate and Initial Status for Sex

Sex	Growth Rate (SD)	Initial Status (SD)
1	1.02 (0.47)	-15.25 (259.32)
2	0.98 (0.43)	14.94 (242.52)

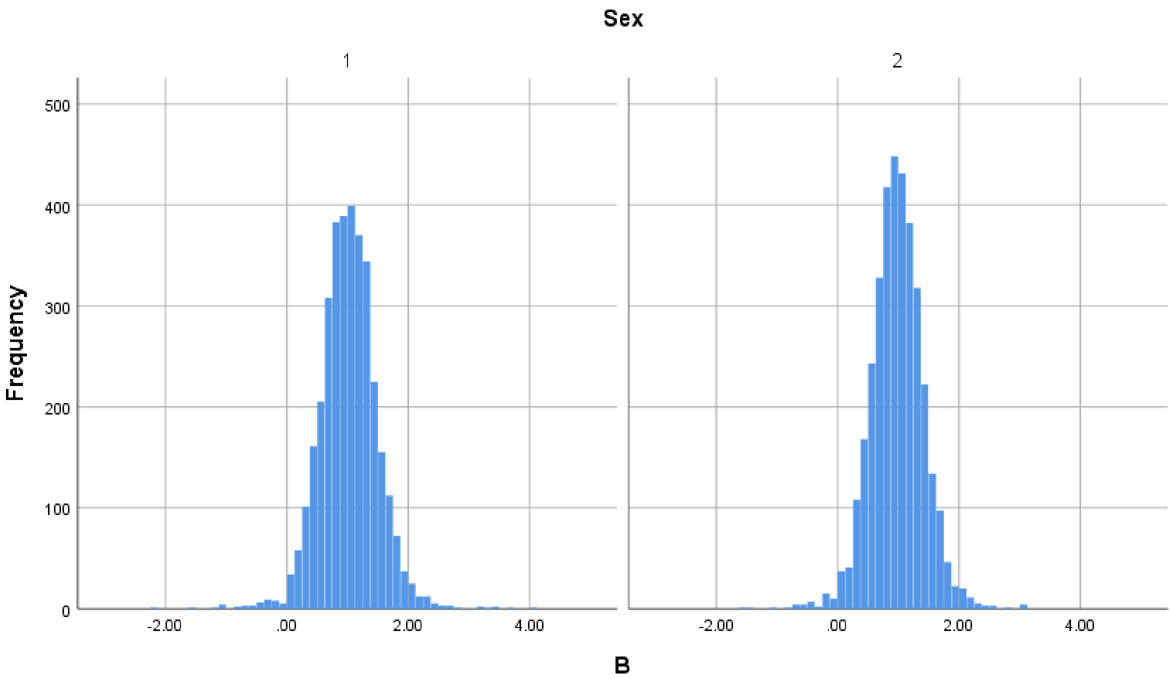

Figure 12 Distribution of growth rates between sexes

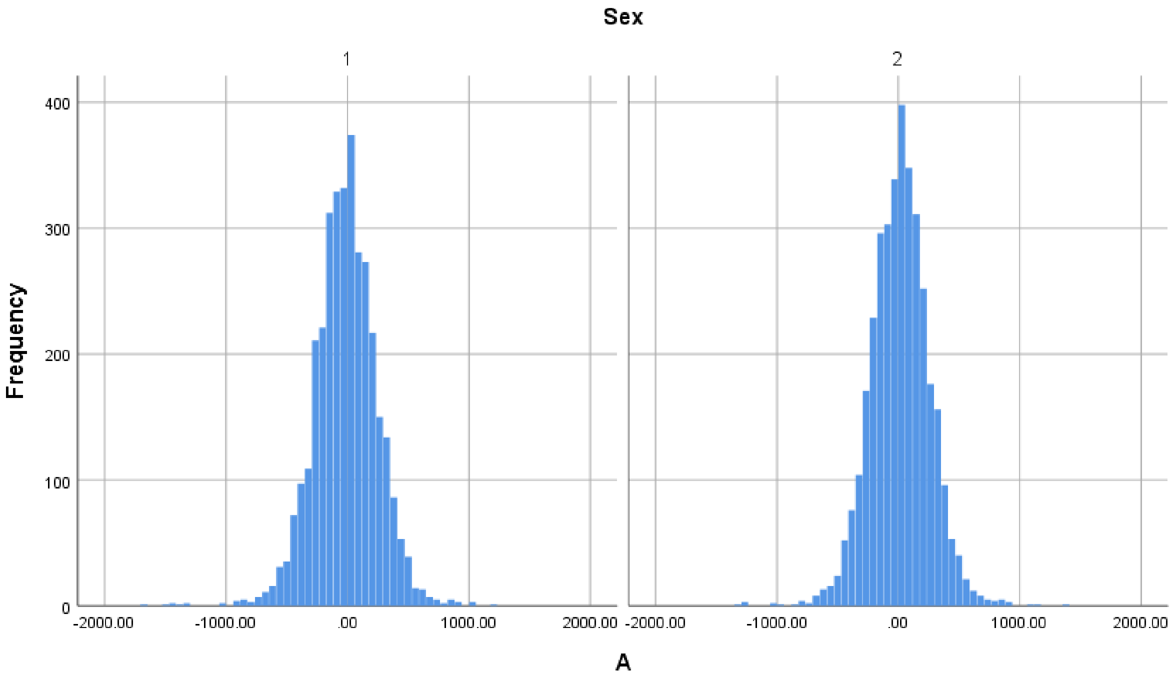

Figure 13 Distribution of initial status between sexes

Independent samples *t*-tests were used to ascertain whether observed differences between sexes as shown in

Table 8 were statistically significant. Levene's Test for Equality of Variances indicated heterogeneity of variance between growth rates of the two sex groups (F = 12.70, p < 0.05). Degrees of freedom were subsequently adjusted to accommodate this, with a very small statistically significant difference between the growth rates of the two groups identified (t(6917.23) = 3.31, p <0.05, d=0.08).

Levene's Test for Equality of Variances also indicated heterogeneity of variance between the sexes' initial status (F = 8.63, p < 0.05). Degrees of freedom were subsequently adjusted, with a very small statistically significant difference between initial status of the two sex groups, as shown in the right-hand column of

Table 8, identified (t(6944.93) = -5.028, p <0.05, d=0.12).

Visual Representation of Differences Between Sexes

As outlined previously, it is possible to represent differences in means simply by regressing the grade means for each sex on the overall sample means, as displayed in Figure 14. These correspond to the values obtained from *Table 7* and *Table 2*.

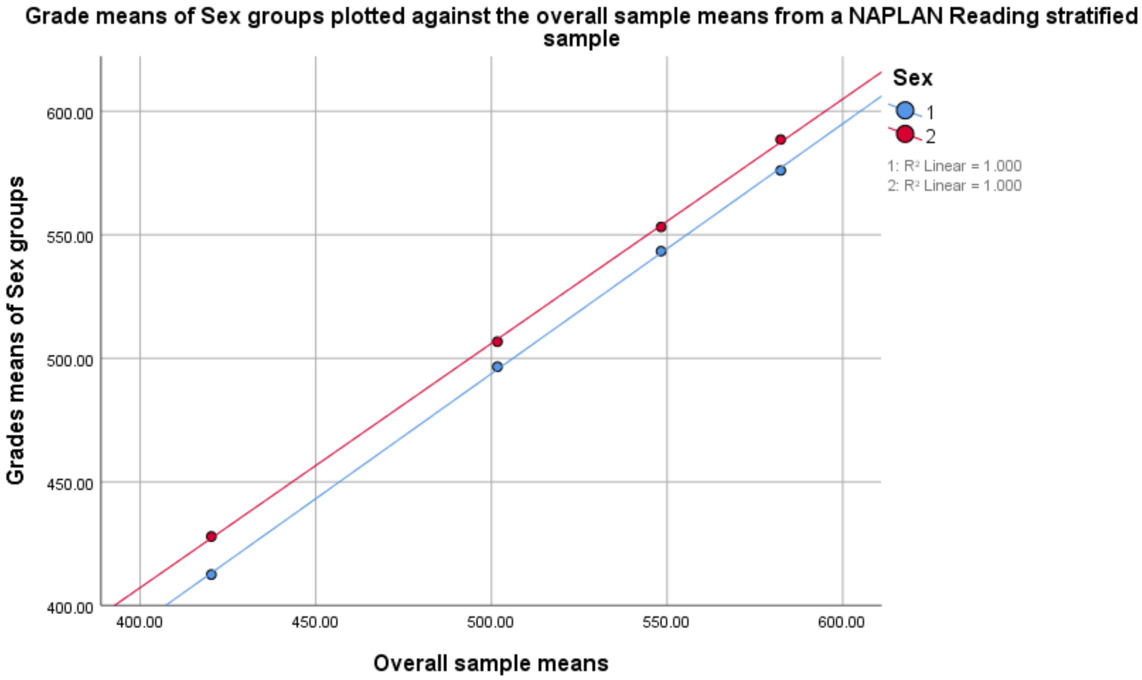

Figure 14 Grade means for each sex plotted against the overall sample means

Similarly, the linear relationship between growth and time, conditional on the meta-metre, that is afford by the RGM permits a representation of growth for separate groups as straight lines. Figure 15 shows the grade means for each sex plotted against the transformed means. The line representing growth for each group can be characterised by two parameters, identified within each of the equations shown on the plots, which are consistent with those growth rate and initial status estimates displayed in

Table 8, with x-axes corresponding to the time-based transformation of grade means obtained from *Table 7*.

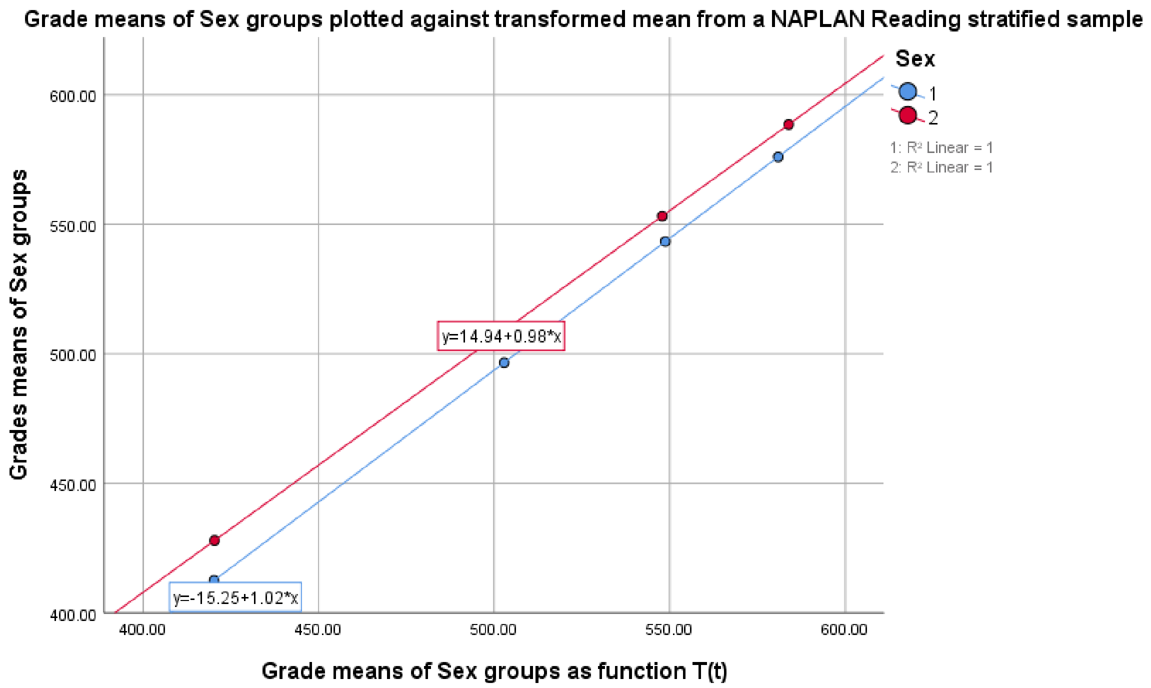

Figure 15 Grade means for each sex plotted against the transformed means

Attempting to characterise the grade means for each sex from Table 7 as function of linear time (i.e., across the four timepoints) once again resulted in poor model fit as evidenced by large residuals and shown in Figure 16. An alternative approach in which a curvilinear (i.e., quadratic) model is fitted minimised residuals and increased the variance explained by the model due to the monotonically decreasing rate of growth, as presented in Figure 17. As applied to the jurisdictional data, taking the natural logarithm of each timepoint allowed the same information presented in Figure 17 to be characterised linearly with an equivalent level of predictive power (i.e., variance explained as R^2), as shown in Figure 18.

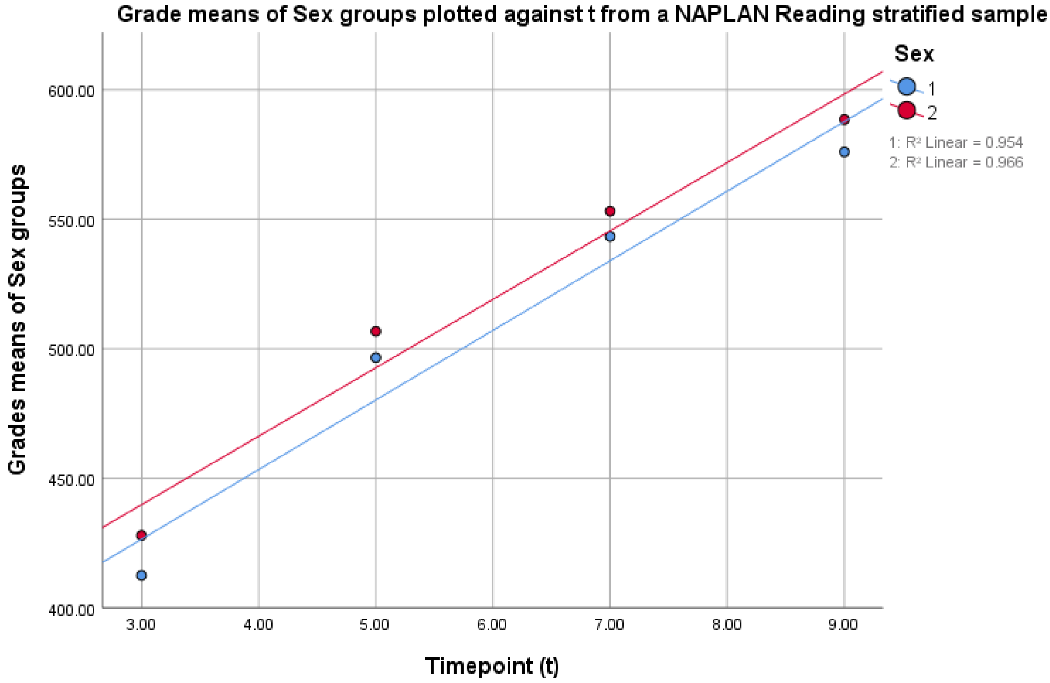

Figure 16 Grade means for each sex plotted against time, assuming linear growth

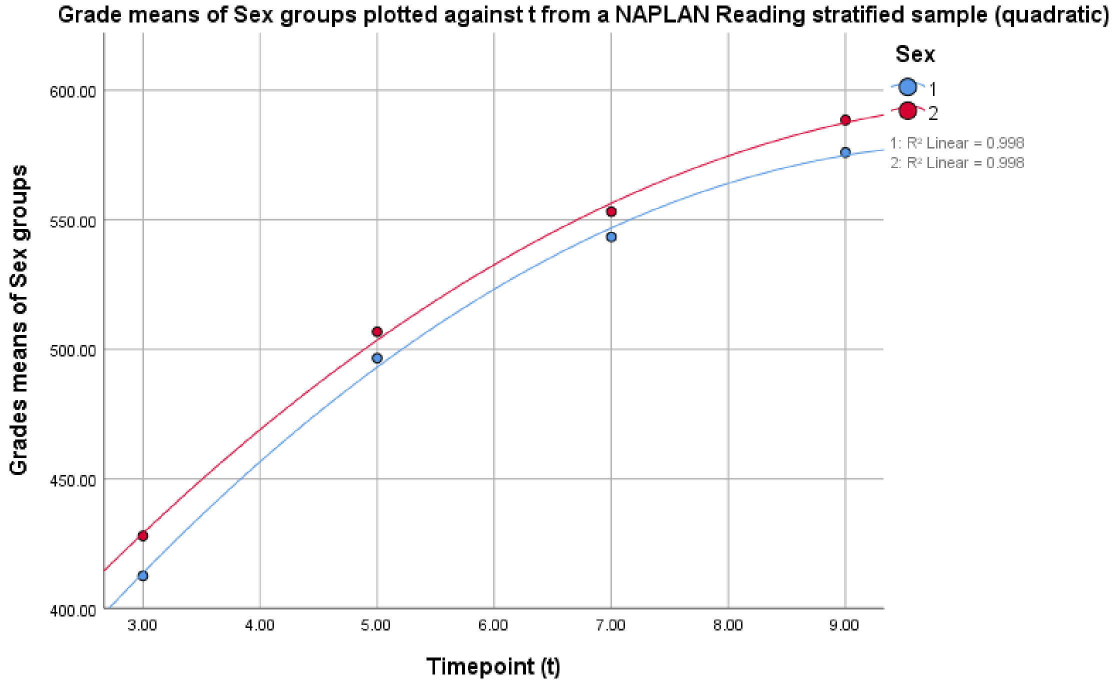

Figure 17 Grade means for each Sex plotted against time assuming curvilinear (quadratic) growth

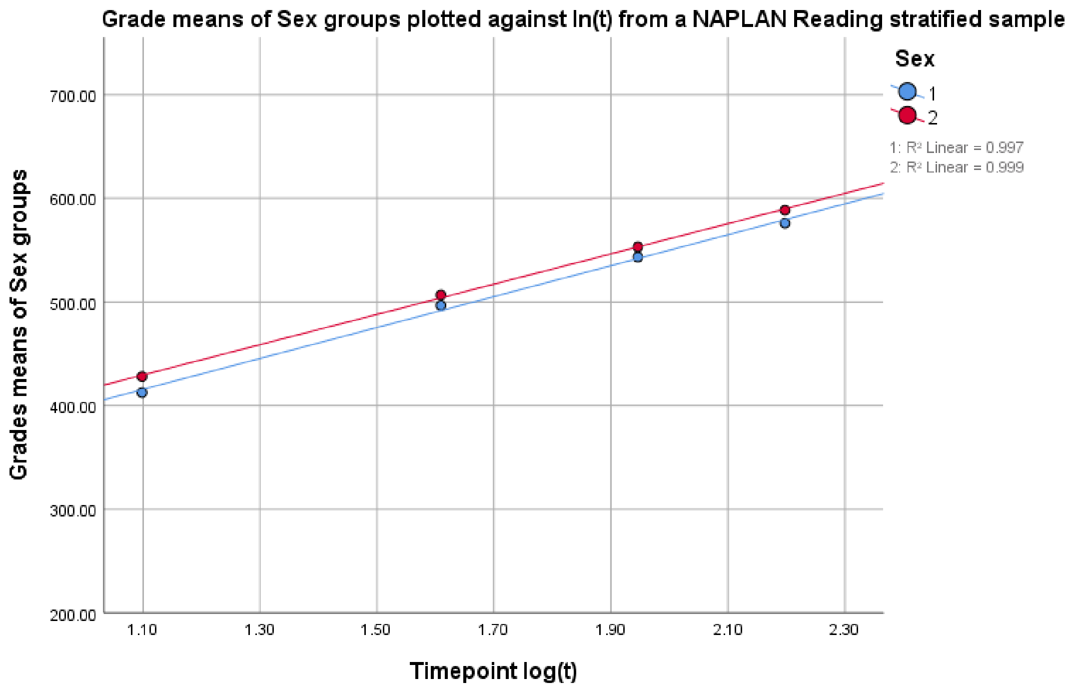

Figure 18 Grade means for each Sex plotted against time taken to its natural logarithm.

Applying the same extension of the RGM to extrapolate beyond measured timepoints for each sex group as was undertaken for jurisdictions, the natural logarithm of each timepoint was taken, thereby permitting an inference regarding expected achievement on the basis of the observed values. A very small difference in the degree of achievement at a time approximating entry to school through extrapolation to log-zero (i.e., Grade 1) is shown in Figure 19.

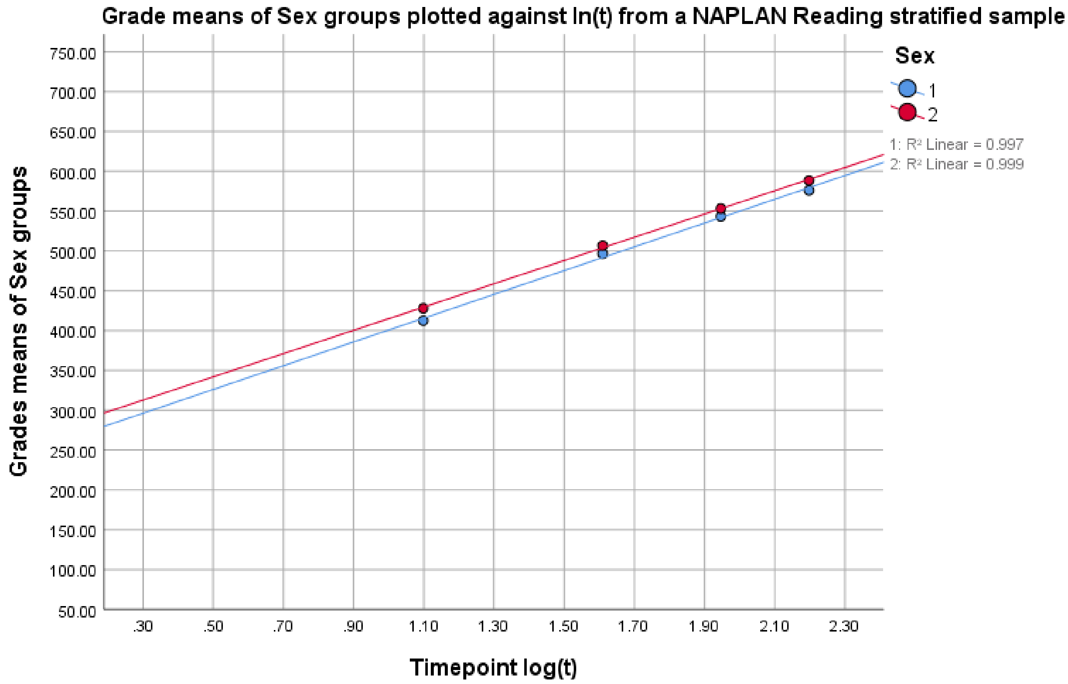

Figure 19 Extrapolated grade means for each Sex plotted against time taken to its natural logarithm

Differences in Growth between ATSI groups

Comparison of Growth Parameters for ATSI groups

Grade means for each ATSI group evidenced the same asymptotic decrease in the rate of growth over time (i.e., a decrease in the level of growth in achievement from earlier grades to later ones) as both jurisdiction and sex analyses, as shown in

Table 9. The normalised relative growth rate and initial status estimates for each ATSI group are presented in

Table 10. A representation of the distribution of these two parameters across the two ATSI groups is presented in Figure 20 and Figure 21 respectively.

Table 9

Grade Means for each ATSI group

ATSI	Grade 3 Mean	Grade 5 Mean	Grade 7 Mean	Grade 9 Mean
0	425.82	507.26	553.61	587.51
1	342.03	423.45	473.27	509.29

Table 10

Comparison of Growth Rate and Initial Status for ATSI groups

ATSI	Growth Rate (SD)	Initial Status (SD)
0	1.00 (0.44)	6.20 (244.92)

1	1.02 (0.60)	-87.31 (317.18)

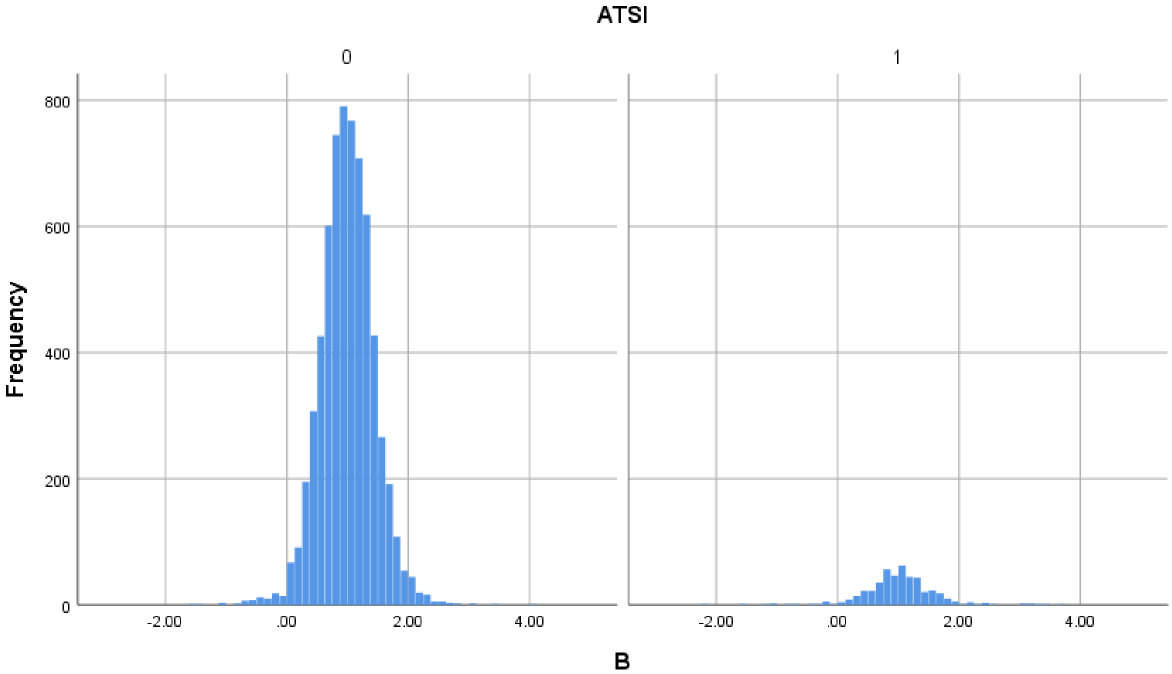

Figure 20 Distribution of growth rates between ATSI groups

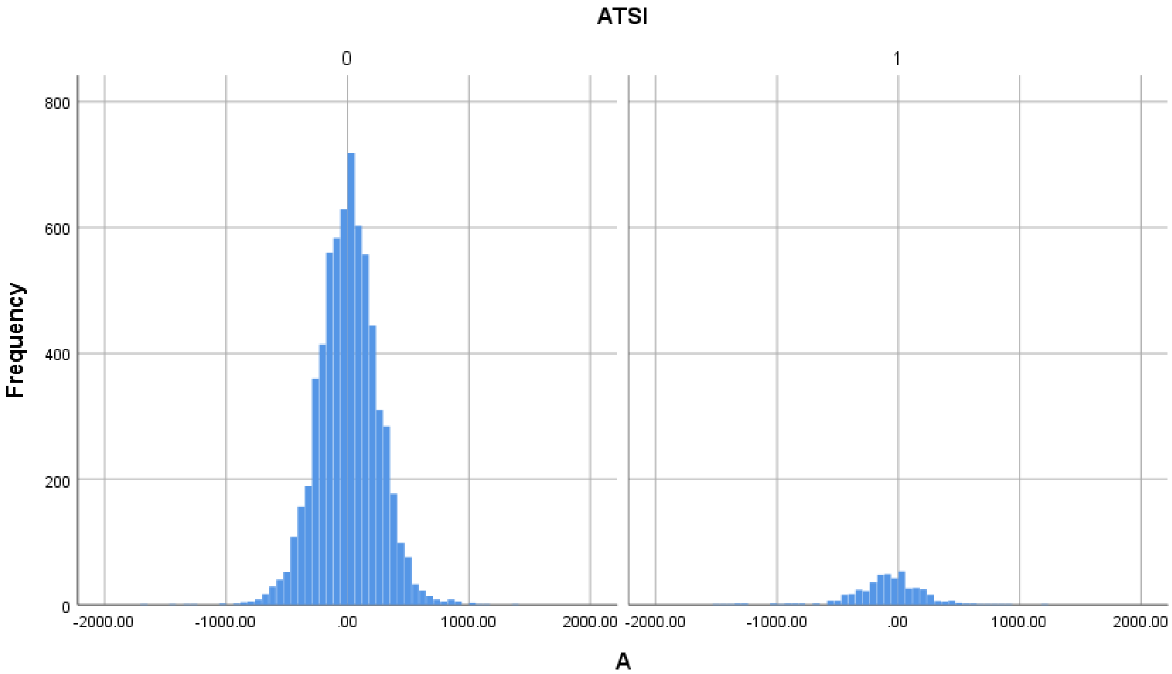

Figure 21 Distribution of initial status between ATSI groups

Independent samples *t*-tests were used to ascertain whether observed differences between ATSI groups as shown in

Table 10 were statistically significant. Levene's Test for Equality of Variances indicated heterogeneity of variance between growth rates of the two ATSI groups (F = 29.71, $p < 0.05$). Degrees of freedom were subsequently adjusted to accommodate this, with no statistically significant difference found between growth rates of the two groups shown in the left-hand column of

Table 8, ($t(497.98) = .818$, $p = .41$).

Levene's Test for Equality of Variances also indicated heterogeneity of variance between ATSI-group initial status (F = 16.81, $p < 0.05$). Degrees of freedom were subsequently adjusted, with a small statistically significant difference between initial status of the two ATSI groups, as shown on the right-hand column of

Table 10, identified ($t(502.96) = -6.22$, $p < 0.05$, $d = 0.33$).

Visual Representation of Differences between ATSI Groups

The same representation of grade means regressed on overall sample means as was presented for jurisdictions and sexes is displayed in Figure 22. The values correspond to those obtained from *Table 2* and

Table 9.

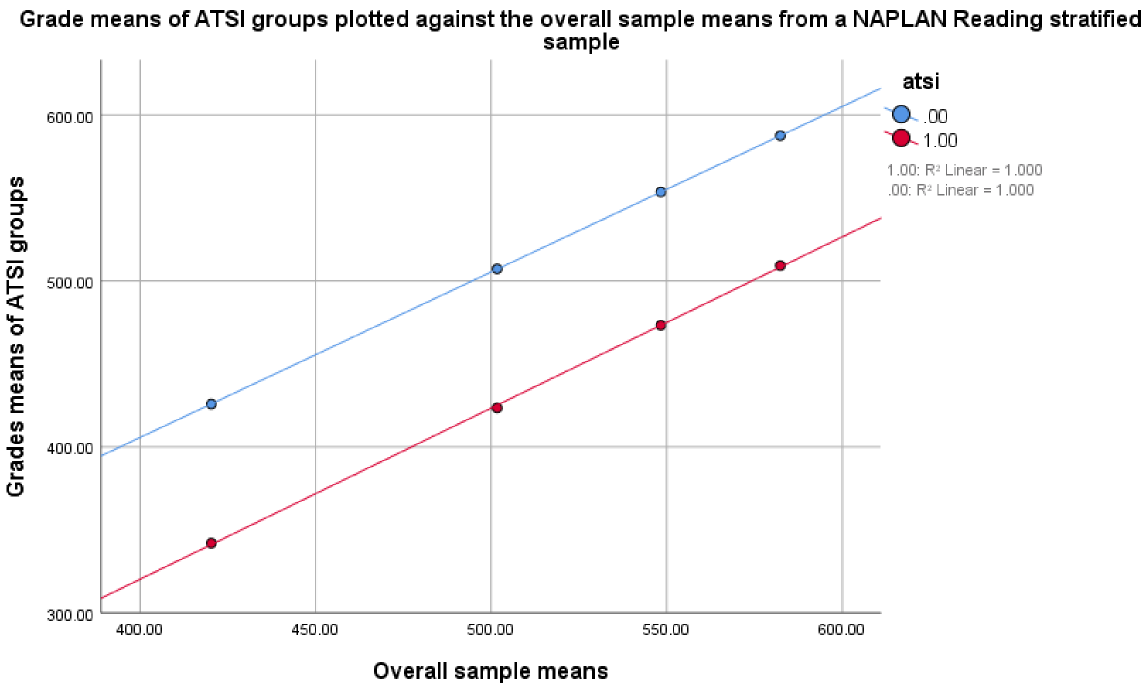

Figure 22 Grade means of ATSI groups plotted against the overall sample means

Figure 23 shows the grade means of each ATSI group plotted against the transformed means (i.e., conditional on the meta-metre), which are represented as straight lines. The line representing growth for each group is be characterised by two parameters, identified within the each equation on the plots, which are consistent with those growth rate and initial status estimates displayed in

Table 10, with their corresponding x-axes obtained from the time-based transformation of means from

Table 9.

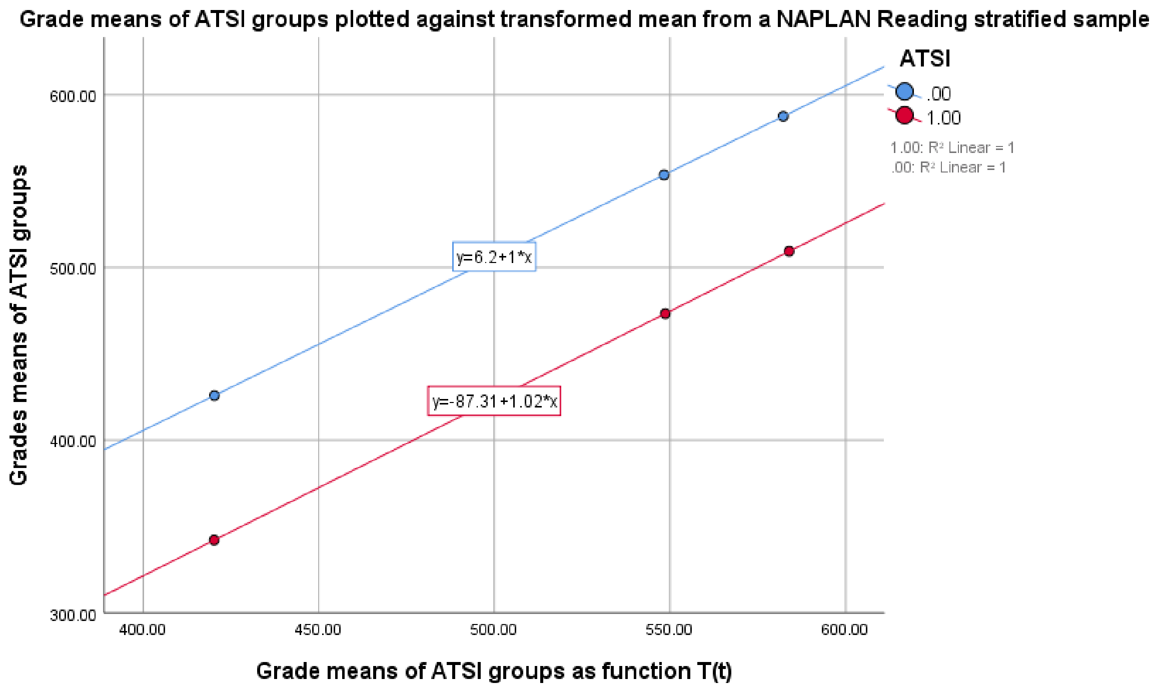

Figure 23 Grade means of ATSI groups plotted against the transformed means

Growth trajectories can similarly be represented as a function of time. Attempting to characterise the ATSI grade means from

Table 9 as function of linear time (i.e., across the four timepoints) once again resulted in poor model fit as evidenced by large residuals and shown in Figure 24. Fitting a quadratic model that assumes a curvilinear functional form minimised residuals and increased the variance explained by the model due to the monotonically decreasing rate of growth, as shown in Figure 25. By taking the natural logarithm of each timepoint, this same information can be characterised linearly, with the same level of predictive power (i.e., variance explained as R^2), as shown in Figure 26.

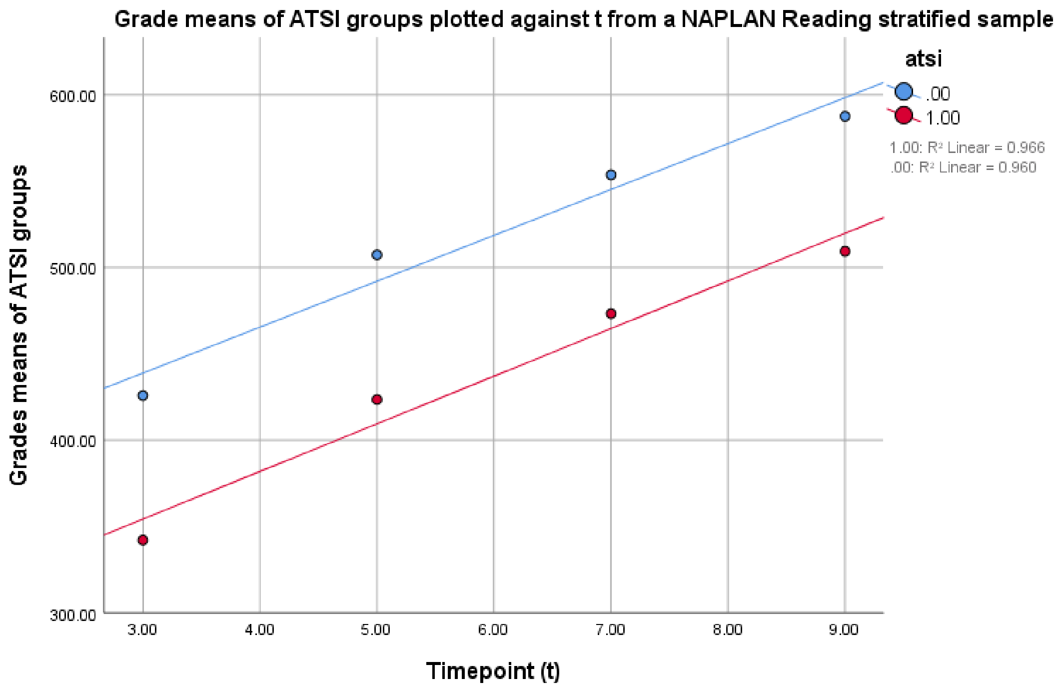

Figure 24 Grade means of ATSI groups plotted against time, assuming linear growth

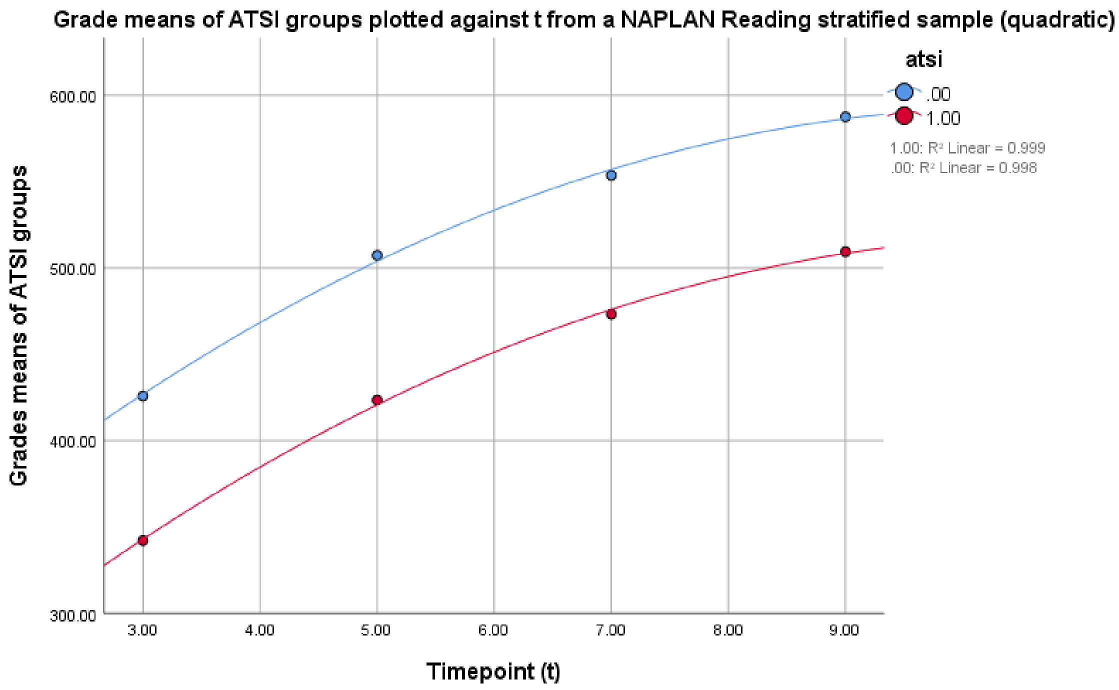

Figure 25 Grade means of ATSI groups plotted against time assuming curvilinear (quadratic) growth

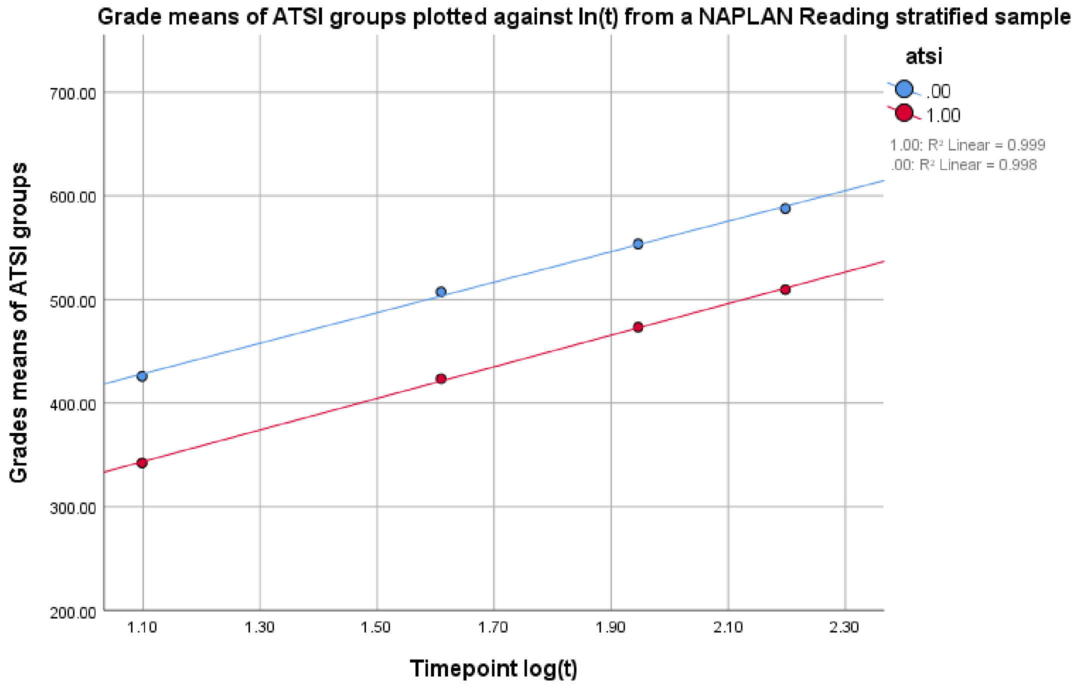

Figure 26 Grade means of ATSI groups plotted against time taken to its natural logarithm.

By applying the same extension to extrapolate the grade means of each ATSI group beyond measured timepoints as was applied to previous groups, inferences regarding expected achievement on the basis of the observed values measured against the natural logarithm of time can be derived. Notable differences in the degrees of achievement at a time approximating entry to school through extrapolation to log-zero (i.e., grade 1) are shown in Figure 27. As noted previously, there was no statistically significant difference in the growth rates of the two groups.

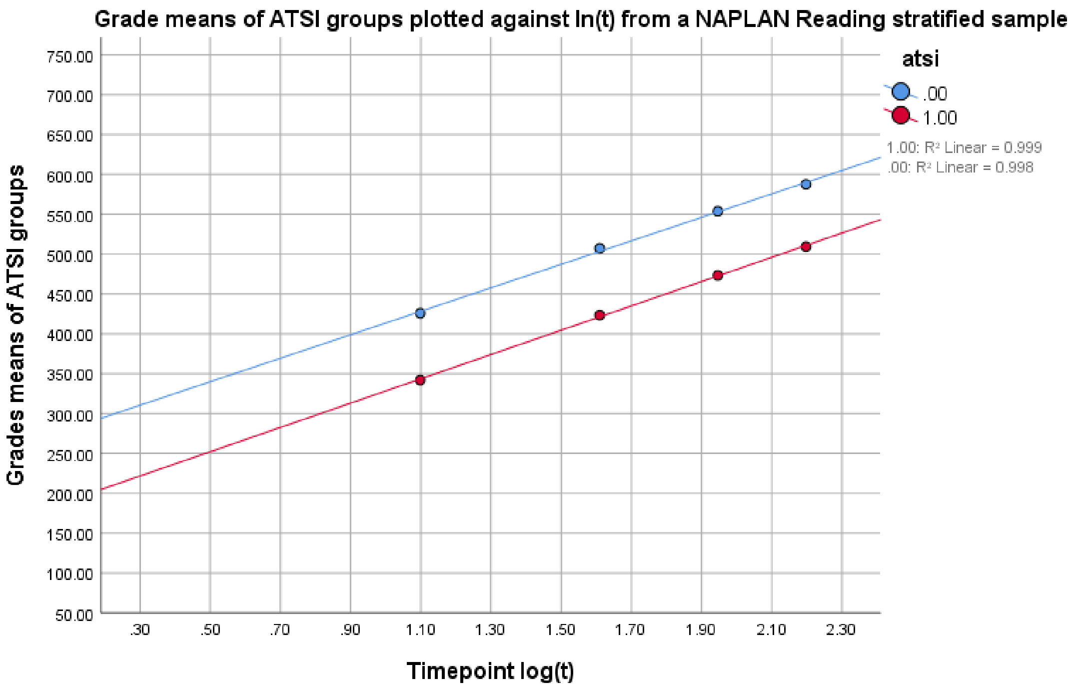

Figure 27 Extrapolated grade means of ATSI groups plotted against time taken to its natural logarithm

Discussion

Characterising Growth in Two Parameters

By applying the RGM to a longitudinal data set containing individual measures of educational achievement in the form of NAPLAN scale scores, and using time-invariant variables pertaining to student jurisdiction, sex and ATSI identification, it was possible to derive a set of parameters by which growth rates and initial status for these groups could be compared.

As summarised in

Table 4, it was possible to characterise the growth of each jurisdiction by two parameter estimates – the growth rate and initial status. Utilising the independent separation of these parameters, univariate tests of significance and multiple comparisons (i.e., post hoc tests) were applied to determine the degree to which differences were statistically significant. These two parameter estimates were subsequently used to visually represent the variation in growth trajectories between each jurisdiction, as shown in Figure 4, Figure 5, Figure 6 and Figure 7. By combining the use of statistical methods and visual representation, it is possible to observe the degree to which differences exist simply by comparing the linear trajectories. For instance, in Figure 4 it is clear that while jurisdiction 2 starts with a higher rate of mean achievement as reflected in the grade 3 mean, the rate at which achievement grows in jurisdiction 1 is considerably greater. Confidence in this assertion is provided via the use of ANOVA. Such findings can be contrasted against those shown in Figure 5, in which the rates of growth between jurisdiction 1 and 6 do not differ (i.e., no statistically significant difference in growth rate was found) as exemplified by the parallel trajectories.

The capacity to derive parameters, plot them and compare them visually was also demonstrated for both sex and ATSI groups. Figure 15 shows that while statistically significant differences were observed in both the growth rate and initial status of the sex groups, as summarised in

Table 8, the degree to which they differed was negligible. This was reflected statistically in the very small value of the Cohen's *d* statistic (Cohen, 1988). The variation in observed mean scores across grades shown in Figure 15 can be contrasted against Figure 23[5], in which clear differences exist in the initial status of ATSI and non-ATSI students, as summarised in

Table 10. The differences shown graphically, and supported by significance testing, indicate considerable, persistent disparities in educational achievement from the outset. This occurs despite a lack of variation between the groups in their rate of growth. Such observations highlight group-level differences requiring further exploration and may be the catalyst for further research and publicly funded initiatives.

Through the presentation of group means as a function of time, it was demonstrated that attempts to fit simple relationships based on a linear functional form were inappropriate, as evidenced in Figure 8, Figure 16 and Figure 24. A more appropriate fit based on curvilinear relationships was observed in Figure 9, Figure 17 and Figure 25, although the complexities of such relationships, with quadratic terms that can be challenging in their interpretation[6], obviate the benefits sought. It was demonstrated that through the application

[5] It should be noted that variation exists in the scales used to represent y-axes in each of the plots, which should be considered in the interpretation of differences in magnitude.

[6] While the interpretation of quadratic relationships is beyond the scope of this paper, one should consider that to make a determination the vertex of the parabola must first be calculated and compared to

of a semi-log transformation of time, an essentially equivalent degree of fit to that of quadratic approaches could be demonstrated, as shown in Figure 10, Figure 18 and Figure 27.

Extrapolation Beyond Measured Timepoints

A subsequent extension of the linearity property is the capacity to extrapolate beyond measured timepoints. In his paper on academic growth, Williamson (2016) emphasises the need for caution in the process of extrapolation for two reasons: first, because there is no measured data and no capacity to check whether growth from other grades can be described by the same model; and second, the complex nature of quadratics and the reversal in direction of curvature associated with the curvilinear functional form. While the former point requires careful consideration, the latter is resolved through the approaches permissible under the RGM.

Noting the requirement to exercise appropriate levels of prudence in result interpretation, several interesting observations across the groups of interest can be made. As shown in Figure 11, there are varying degrees of expected achievement at the extrapolated timepoint equivalent to Grade 1 (i.e., ln(1), visible at the intersection of each line and the y-axis). For example, by comparing jurisdictions 6 and 7 – similarly presented without extrapolation using the growth rate and initial status coefficients in Figure 7 – a clear difference at the extrapolated timepoint is observed, with students in jurisdiction 7 demonstrating higher achievement earlier. However, one can also note the steeper growth trajectory evidenced by

specific ranges for each group. In each of the examples presented in this paper, the quadratic term is negative, meaning the linear and quadratic terms have differential effects on the curve at different locations.

jurisdiction 6, indicating a higher growth rate. These differences can be contrasted to the linearised curves of jurisdictions 4 and 5, which evidence very similar – and not significantly different – growth rates and initial status. While such values are reflected in the parameter estimates (e.g., in

Table 4), their presentation in a linearised, extrapolated form allows for a further efficiency in their comparison. Furthermore, significant findings may point to evidence of early educational advantage in certain jurisdictions and varying degrees of improvement over the course of schooling, as consistent with previous research (e.g., Goss et al., 2018).

As previously articulated, both in Figure 15 and Table 8, the degree to which sexes differed was small although significant with respect to both summative parameter estimates, which are evident at the extrapolated Grade 1 values shown in Figure 19. Such sex differences have been noted elsewhere (e.g., Thomson, De Bortoli, Underwood, & Schmid, 2019). This is in contrast to the evidence provided in Figure 27, which suggests considerable differences in the overall levels of educational achievement at the commencement of schooling for ATSI and non-ATSI students. These findings are consistent with both qualitative and quantitative research across a variety of disciplines (e.g., Guenther, Bat, & Osborne, 2013; Gray & Beresford, 2008; Thomson, De Bortoli, Underwood & Schmid, 2019) as well as Australian government reporting (e.g., Department of Prime Minister and Cabinet, 2018). One observation of note, however, is the lack of difference in growth rates between these groups. Such findings contradict other research in the area (e.g., Bradley, Draca, Green, & Leeves, 2007) and, while requiring further investigation as to the source (whether divergent due to methodology or substantive reasons), provide a positive message regarding the growth trajectories of a group known to face considerable educational disadvantage and may substantiate the need for further policy reform in this area.

Limitations of the RGM

While the RGM provides an effective method through which group growth can be described and compared, the model itself is not explanatory. As in analytic methods such as principal components analysis, the purpose of the model is largely descriptive (Jolliffe & Cadima, 2016); therefore, no attempt is made to provide clarity on the multitude of possible factors contributing to observed differences. While providing an effective method for reporting and lending itself to use as a preliminary step in research activities, the testing of complex theories is reserved for alternative models. As broad sets of such methodological approaches exist, it is recommended that further research compare estimates derived from such models against those of the RGM, while also considering the properties associated with their outcomes.

A further limitation of the RGM is the requirement for complete longitudinal data. While the model retains dynamic consistency (Keats, 1980), the incorporation of individual differences in the estimates of group-level summative parameters results in a requirement for non-missing data. One outcome of this is the possibility of correlations between rates of attrition and other variables of interest. While the current application pertains to a census-based assessment with very high participation rates (i.e., greater than 95%; Australian Curriculum, Assessment and Reporting Authority, 2019), the possibility remains that high rates of withdrawal or attrition could feasibly give rise to systematic biases within certain sub-groups, thus limiting the interpretability of results.

Similarly, while the present analysis was undertaken predominantly for demonstrative purposes, it should be noted that it was applied to a limited data set. While efforts were made to retain consistency with the original data through a comparison of the distribution of variables of interest, the analysis and results may include sampling biases that obscure true results. For example, as no data was available one jurisdiction, the present derivations are based on a sub-set of the total population. Equally, there were data linkage challenges that resulted in a loss of cases through the matching of students across each NAPLAN cycle – 2013, 2015, 2017 and 2019. To resolve this, the introduction of universal student identifiers for the purposes of data linkage would likely be required, particularly if such approaches were to be implemented at scale. While recommendations put to government have advocated for the implementation of these to serve the needs of longitudinal approaches (e.g., Gonski et al., 2018), such decisions may require further ethical examination prior to implementation (Arnold, 2013).

Concluding Remarks and Future Directions

This research investigated the application of the RGM for evaluating differences between groups. This came in response to calls for reporting of progress in student achievement throughout the course of schooling. The research utilised the summative properties of the RGM and its capacity to represent growth in an easily interpretable manner to evaluate variations in NAPLAN reading achievement between jurisdictions, sexes and groups categorised on the basis of ATSI identification.

By applying the RGM, it was possible to derive a set of parameter estimates through which growth could be compared. The degree of variability in the initial status and growth of jurisdictions was considerable. Very small but significant differences were observed in both the growth rate and starting location of students grouped by sex, while significant differences were observed in the starting locations of ATSI and non-ATSI students but not in their growth rates. Each set of results was presented graphically, visually emphasising areas of variability while also demonstrating qualities associated with the RGM. It was also possible to highlight potential variation in the achievement of students at entry to school through the extrapolation of observed results, although caution was highlighted in the adoption of this process. While descriptive in their presentation, the results point to areas for further investigation.

In utilising the RGM, it is noted that there exists a wide variety of current growth models that incorporate individual variability and permit the modelling of complex developmental theories (Duncan & Duncan, 2004). Such approaches often draw on the traditions of hierarchical linear modelling (HLM; Raudenbush & Bryk, 2002) and structural equation modelling (SEM; Meredith, & Tisak, 1990). While the RGM serves an effective, niche method – providing what Goss et al. (2018) describe as a need for approaches that accommodate non-linear rates of growth for the purpose of comparing relative student progress – additional value could be sought through the comparison of group estimates derived under different methodologies. In the case of the RGM, the linearisation of growth rates via the instantiation of the meta-metre, conforming to properties consistent with objective measurement, provides a benefit that may not be realised in alternative approaches.

Noting the use of the meta-metre in deriving comparative estimates, a focus for further research lies in the impact of violations of this time transforming function, and the conditions under which these are likely. While Rao (1958) describes a process in which the existence of a common transformation can be empirically tested, the degree to which such violations may impact parameter interpretation has not been investigated. Due to the novelty of these methods, further inquiry into such conditions would be an appropriate next step in evaluating the application of the RGM. Future applications would also benefit from exploration into the error terms associated with the RGM parameter estimates. Citing Brody (1993), Stone (2020) highlights that the approach used in the model essentially averages over incidentals (i.e., individual variation and error) with the view to overwhelm individual levels of variation and move to a superordinate level of aggregation. While this may provide appropriate summative properties, the implications of doing so – as well as the consequences of the incorporation of measurement error into specific measurement types (i.e., WLEs; von Davier, Gonzales, & Mislevy, 2009) – requires evaluation.

Despite the limitations associated with the uniqueness of this under-utilised method, the present investigation has provided support for the effectiveness and efficiency of the approach to growth modelling first proposed by Rasch. Through a combination of examples, it has demonstrated that the RGM allows for the efficient comparison of groups in a manner that is both statically rigorous and commonly accessible. Such an approach provides both a novel and effective means by which growth may be reported and an avenue for further investigation into group-level differences via qualitative and quantitative research.

References

Andrich, D. (1988). *Rasch Models for Measurement*. SAGE Publications: Thousand.

Andrich, D. (2004). Controversy and the Rasch model: a characteristic of incompatible paradigms?. *Medical care, 42*(1), I7-I16.

Andrich, D., & Marais, I. (2019). A Course in Rasch Measurement Theory. *Measuring in the Educational, Social and Health Sciences*. Singapore: Springer.

Andrich, D., Marais, I., & Sappl, S. (2022). *Adapting Rasch Measurement to the Study of Growth: A Meta Metre for some Intelligence and Attainment Tests*. (Forthcoming) Springer Nature: Singapore.

Arnold, B. (2013). National student identification (but not quite yet)?. *Privacy Law Bulletin, 9*, 111.

Australian Curriculum, Assessment and Reporting Authority. (2012). *Data Standards Manual: Student Background Characteristics*. Sydney: Author.

Australian Curriculum, Assessment and Reporting Authority. (2013). *My School Technical Report: My School Student Gain*. Sydney: Author.

Australian Curriculum, Assessment and Reporting Authority. (2019). *NAPLAN Achievement in Reading, Writing, Language Conventions and Numeracy: National Report for 2019*. Sydney: Author.

Australian Curriculum, Assessment and Reporting Authority. (2020). *NAPLAN 2019 Technical Report*. Sydney: Author.

Bond, T. & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. New York: Routledge.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). 'The Concept of Validity'. *Psychological Review, 111*(4), 1061-1071.

Bradley, S., Draca, M., Green, C., & Leeves, G. (2007). The magnitude of educational disadvantage of indigenous minority groups in Australia. *Journal of Population Economics, 20*(3), 547-569.

Brody, T. A. (2012). *The philosophy behind physics*. L. DeLaPena & P. Hodgson (Eds.). Berlin, Germany: Springer Verlag.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum Associates.

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of cognition and development, 11*(2), 121-136.

Dee, T. S. (2004). Are there civic returns to education?. *Journal of public economics, 88*(9-10), 1697-1720.

Department of Prime Minister and Cabinet. (2018). *Closing the gap: Prime Minister's report.* Canberra: Author.

Duncan, O.D. (1984). Rasch Measurement: Further Examples and Discussion. In Turner, C. F. & Martin, E. (Vol. 2), *Surveying Subjective Phenomena* (pp. 367-401). New York: Russell Sage Foundation.

Duncan, T. E., & Duncan, S. C. (1995). Modeling the processes of development via latent variable growth curve methodology. *Structural Equation Modeling: A Multidisciplinary Journal*, *2*(3), 187-213.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational psychology, 88*(1), 3-17.

Gonski, D., Arcus, T., Boston, K., Gould, V., Johnson, W., O'Brien, L., Perry, L.A. & Roberts, M. (2018). *Through growth to achievement: Report of the review to achieve educational excellence in Australian schools*. Canberra: Commonwealth of Australia.

Goss, P., Sonnemann, J., and Emslie, O. (2018). *Measuring student progress: A state-by-state report card*. Melbourne: Grattan Institute.

Gray, J., & Beresford, Q. (2008). A 'formidable challenge': Australia's quest for equity in Indigenous education. *Australian Journal of Education, 52*(2), 197-223.

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European journal of epidemiology, 31*(4), 337-350.

Grossman, M. (2006). Education and nonmarket outcomes. *Handbook of the Economics of Education, 1*, 577-633.

Guenther, J., Bat, M., & Osborne, S. (2013). Red dirt thinking on educational disadvantage. *The Australian Journal of Indigenous Education, 42*(2), 100-110.

Hanushek, E. A., & Wößmann, L. (2007). *The role of education quality for economic growth*. The World Bank.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374*(20150202), 1-16.

Keats, J. A. (1980). Comparing Latent Trait with Classical Measurement Models. In Spearitt, D. (1982), *The Practice of Educational and Psychological Measurement. Improvement of Measurement in Education and Psychology*, 17-35. VIC: Australian Council for Educational Research.

Kuhn, T.S. (1977). The function of measurement in modern physical science. In T.S. Kuhn (1977), *The Essential Tension: Selected Studies in Scientific Tradition and Change* (pp. 161-190). Chicago: The University of Chicago Press.

Masters, G. N. (2020). Nurturing Wonder and Igniting Passion, designs for a new school curriculum: NSW Curriculum Review. NSW Education Standards Authority.

McArdle, J. J., & Nesselroade, J. R. (2003). Growth curve analysis in contemporary psychological research. *Handbook of psychology*, 447-480.

McGaw, B., Louden, W., & Wyatt-Smith, C. (2020). *NAPLAN Review Final Report*. NSW (Department of Education), QLD (Department of Education), VIC (Department of Education and Training) & ACT.

McNeish, D., & Matta, T. (2018). Differentiating between mixed-effects and latent-curve approaches to growth modeling. *Behavior Research Methods, 50*(4), 1398-1414.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*(1), 107-122.

Nese, J. F., Lai, C. F., & Anderson, D. (2013). A Primer on Longitudinal Data Analysis in Education, Technical Report No. 1320. *Behavioral Research and Teaching*, 1-21.

Olsen, L. W. (2003). *Essays on Georg Rasch and his contributions to statistics*. Københavns Universitet, Økonomisk Institut.

Peebles, D., & Ali, N. (2015). Expert interpretation of bar and line graphs: The role of graphicacy in reducing the effect of graph format. *Frontiers in psychology, 6*, 1673. doi: 10.3389/fpsyg.2015.01673

Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics, 14*(1), 1-17.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In J. Neyman (Eds.), *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability: Volume IV: Contributions to biology and problems of medicine* (pp. 321-333). Berkeley: University of California Press.

Rasch, G. (1972). Georg Rasch 1972 Retirement Lecture: Objectivity In Social Sciences – A Method Problem. In *Rasch Measurement Transactions* (2010), 24:I, 1252-1272.

Rasch, G. (1977). On specific objectivity. An attempt at formalizing the request for generality and validity of scientific statements. In the Danish Yearbook of Philosophy (Vol. 14), *Proceedings of the Symposium in Scientific Objectivity*, 58-93. Copenhagen: Munksgaard.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: *Applications and data analysis methods – Advanced Quantitative Techniques in the Social Sciences (Vol. 1)*. Thousand Oaks, CA: Sage.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford: Oxford university press.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*(2684), 677-680.

Stone, M. H. (2020). Rasch's Logistic Model Applied to Growth. *Journal of Applied Measurement*, *21*(1), 1–16.

Thomson, S., De Bortoli, L., Underwood, C., & Schmid, M. (2019). *PISA 2018: Reporting Australia's Results (Vol. 1).* VIC: Australian Council for Educational Research.

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

Williamson, G. L. (2016). Novel interpretations of academic growth. *Journal of Applied Educational and Policy Research, 2*(2), 15-35.

Williamson, G. L. (2017). "Measuring" Academic Growth: Where Does Measurement Reside?—A commentary on Measurement: A Very Short Introduction by David J. Hand. *Measurement: Interdisciplinary Research and Perspectives, 15*(1), 19-22.

Williamson, G. L., Fitzgerald, J., & Stenner, A. J. (2014). Student reading growth illuminates the Common Core text-complexity standard: Raising both bars. *The Elementary School Journal, 115*(2), 230-254.

Wright, B. D., & Stone, M. H. (1979). *Best Test Design*. Chicago: MESA Press.