Appendix 5 NAPLAN data and model.

The National Assessment Program—Literacy and Numeracy (NAPLAN) and the relationship of mean learning status of cohorts with Year level and age

Preliminary comment.

Through out this appendix (and in Appendices 6 and 7) the *CurveExpert* (Hyams, 2001) software is used to fit a model developed from the Gompertz expression (Gompertz, 1825). The curve fitting is a pragmatic process to idealise the trajectories. Alternative curves can serve this purpose. A quadratic curve can provide an approximately equivalent solution (as can some other models applied in biological research included in the *CurveExpert* software). The Gompertz expression is selected for two reasons. The less important is that it pays homage to Courtis who originally proposed it as a family of curves to model growth in learning over time (see footnote 7 in chapter 2). The more important reason is pragmatic. The smooth curve derived for learning improvement with age fits the data points well. However, in addition, the inflection points for the various curves that can be fitted to the curve of the means or to the trajectories of +/- 1.96 SD, offer further mathematical modelling of slow and faster moving cohorts (relative to the mean cohort) in ways that provide hypotheses for further research into learning growth rates with age. In the process models for the trajectories of SD at particular ages are also predicted. However, for the basic modelling of the means of learning status with age a quadratic function works adequately.

NAPLAN

The first Australian National Assessment Program—Literacy and Numeracy (NAPLAN) tests were conducted in May 2008 for all Years 3, 5, 7 and 9 students in government and non-government schools (National Assessment Program Literacy and Numeracy (NAPLAN) Full Report, 2008). This publication has been timely and helpful in the refinement of the understanding of the general trend in test performance with increasing Year levels and age for this thesis. Prior to the NAPLAN report, information from some US states and from some US test norming programs, had been the best sources for explaining (and justifying for extrapolation and interpolation) some of the ‘dynamics’ of learning in English language and mathematics over an extended period of schooling. While the NAPLAN data are cross-sectional, based on Hilton & Patrick (1970) the trends can be considered as approximately similar to the longitudinal situation and indicative of the likely trends that applied in the South Australian data of 1997 and 1998.

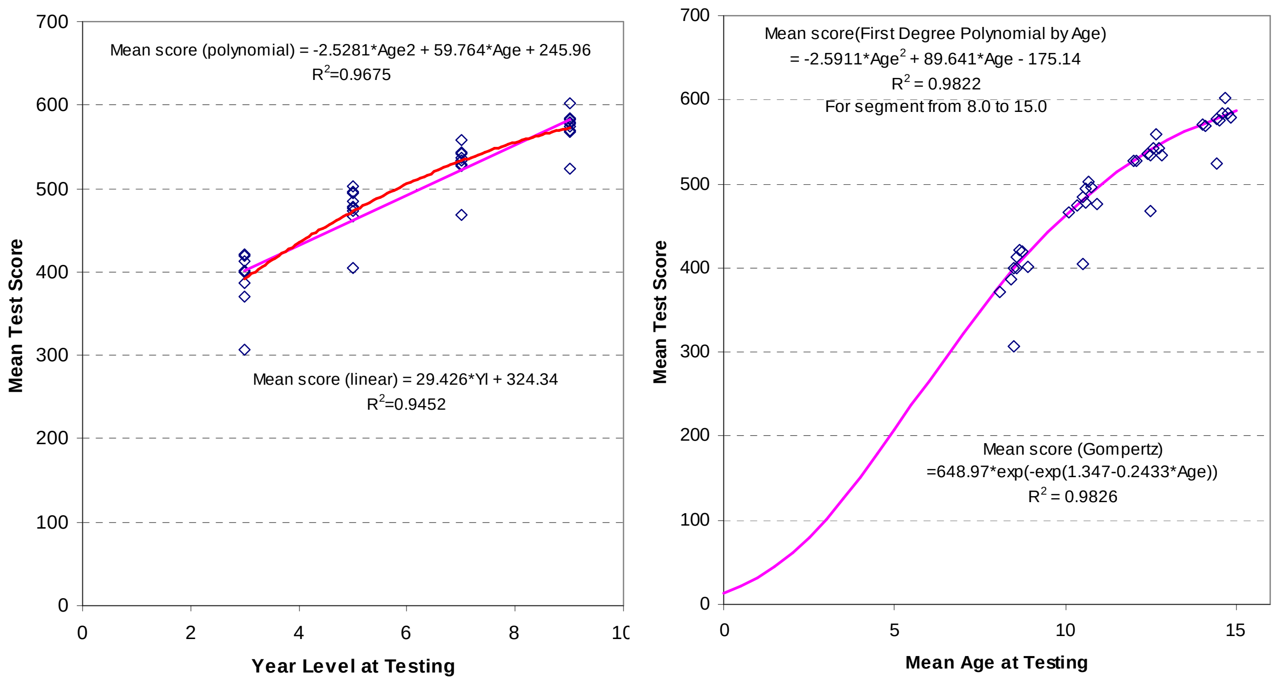
The 2008 collection was the first occasion that students in Australia had been assessed at these Year levels with a set of common and linked tests and an underlying vertical scale. From 1999 to 2007 the National Report on Schooling published data, commencing with reading in 1999 at Years 3 and 5 only. By 2007 the National Report included reading, writing and numeracy at Years 3, 5 and 7. However students sat different tests in each State and Territory and while these were referenced to national benchmarks there were no reported vertical scales. Results were reported as percentages meeting national benchmark criteria (National Report on Schooling in Australia, 2007).

For the vertical scale, skills and understandings assessed in each domain (reading, writing, spelling, grammar and punctuation, and numeracy) from Year 3 through to Year 9 are mapped onto achievement scales with scores that range from 0 to 1000. By locating all students on a single national scale for each domain, the scales provide a clearer understanding of the general trends of learning with time/age/Year level. The ‘reporting scales are constructed so that any given scaled score represents the same level of achievement over time. For example, a score of 700 in Reading will have the same meaning in 2010 as it has in 2008' (NAPLAN Full Report, 2008, p. 3).

Modelling the learning trajectory

The 2008 data for each Year level are reported in Figure A5.1 in the left panel. Each point represents one state or territory system with government and non-government schools combined. In the right panel, the same data are graphed by average age at testing rather than by Year level. (The data could also be presented by elapsed years of schooling rather than age.) The four low outlier points in each panel are the points for the Northern Territory. These points are omitted for model estimation using 'curve-fitting' software (Hyams, 2001, Curve Expert). The vertical axis is plotted using the original 1000 point scale for the test collection scale. As will be indicated later in this appendix, the assumption about where the test scale origin at time 0 (birth) occurs has an influence on the trajectory of the curve, particularly in the lower ages and also on the model determined final asymptote. Varying the notional test scale score at 0 time also has an influence on the location of the inflection point in sigmoid models. For convenience, the original data are left unchanged, that is the original scale is preserved. The fitted model in this cases passes through Test score =13 at age=0. The Gompertz model assumes that the scale has a value on the vertical axis close to zero at time zero.

Figure A5.1 Plot of NAPLAN System Averages for Reading-2008



In the left panel the effect of centreing points on Year level on the time dimension eliminates the spread of the individual points on the time axis. The Coefficients of Determination (R^2) are shown as one indicator of model fit. For linear fit (Northern Territory omitted) R^2 is 0.9452. The R^2 for a first-degree polynomial is 0.9675.

In the right panel the data points are spread along the time axis at the average age at testing. Two curves are fitted, which at the scale of the graph, appear to be visually identical. The first-degree polynomial fitted from age 8 to 15 (R^2 = 0.9822 –NT omitted), follows approximately the same trajectory as the Gompertz curve (R^2=0.9826 – NT omitted) for the same segment. The Gompertz curve is plotted from age 0 to age 15.

A variety of solutions for a mathematical description of a smoothed curve to model the data points are possible. As described in Chapter 2 and in Chapter 5, the Gompertz curve provides a good fit for the points both statistically and conceptually and brings with it some 'testable' spin-offs related to rates of learning at points on the curve. The relationship between rate of leaning and the dispersion of the data (SD) near the inflection point of the rate of learning for a cohort/groups at specific ages can be explored. As a result the general first order Gompertz model is used to model the trajectory of the mean learning status of various cohorts in the subsequent sections of this appendix.

The Gompertz curve for the model in Figure A5.1 has the form:

Mean Test score = 648.97*exp(-exp(1.347-0.2433*Age)), where 648.97 is the asymptote for the curve.

What is clear from Figure A5.1 is that the general trend for the cross-sectional mean cohort scores by age and Year level for reading is non-linear. The left panel indicates a better fit for a quadratic function than a linear function. In both panels the trajectory can be modelled by a Gompertz model, a quadratic model and a number of other unreported possibilities (Morgan Mercer Flodin (MMF) as one example). The right panel shows a relationship of the mean test scores within systems to the age at testing. From this example there is supporting evidence for allowing the use of a non-linear model for interpolating and extrapolating cohort data discussed in more detail in Chapter 6. A simplification of the model using national norms rather than using each of the states as data points produces an equivalent model. However the model based on state data points highlights an apparent anomaly in state comparisons when data are centred on Year levels. These state comparisons are discussed briefly at the end of the appendix..

A general model for the NAPLAN 2008 data based on national means

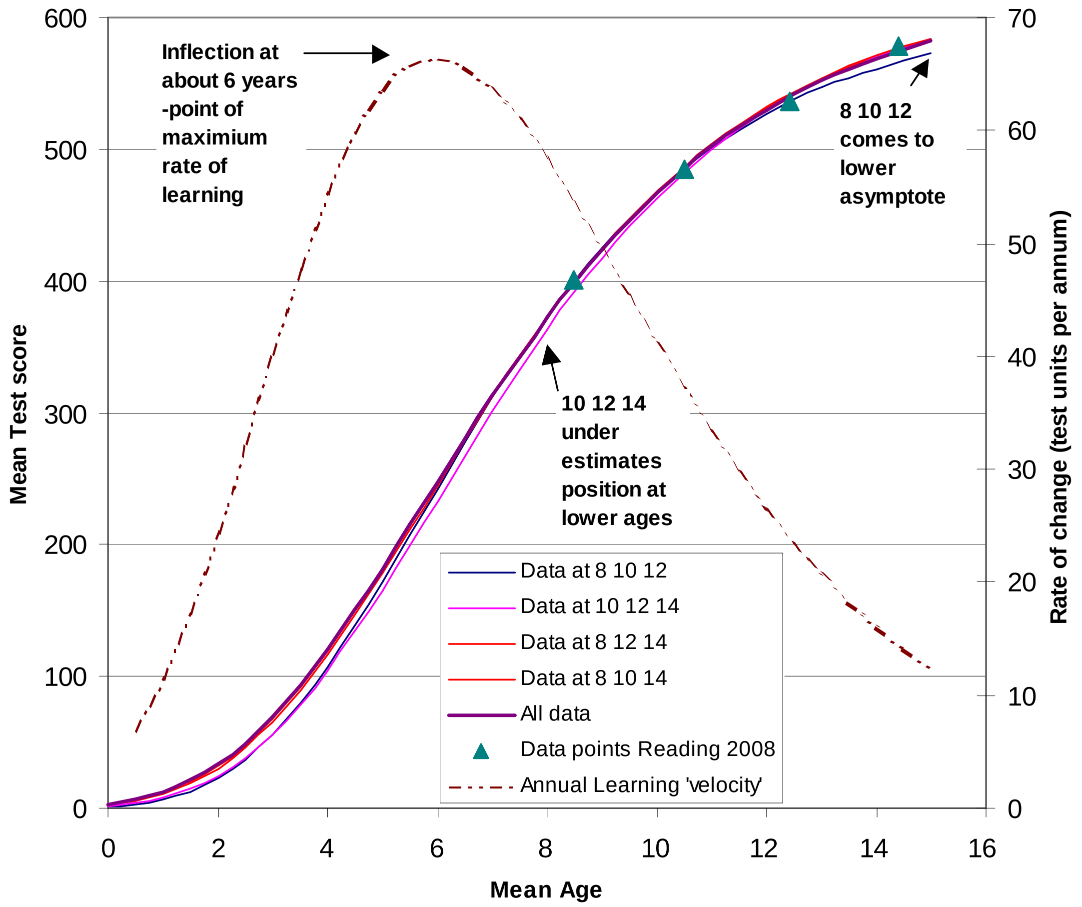
Based on the approximate fit of the Gompertz relation to the individual state and territory data, a model based on the national means can be developed as a general indicator of the trajectory for group means with age in reading and numeracy. On this basis the model is based on almost all students in Australia in Years 3, 5, 7, 9 in 2008. The relationships identified in this model development provide some general insights to be applied in the model for SA data for 1997 and 1998.

The impact of missing data points on curve parameters

The four means for each age cohort enable the impact on the parameters estimated for the curve when data points are missing to be explored along with the impact of these missing points on the trajectory fitted. This exploration will influence approaches adapted in Chapter 6 for extrapolation, where only two real points (Year 3 and Year 5) are available in the calendar years of the data reported, although proxies for Year 7 and Year 9 can be considered, based on more recent data.

Figure A5.2 is developed from the national reading data. The graph provides a comparison of the application of the Gompertz model when varying data points from among the set of available data points are used to estimate the curve. The models are fitted to the national mean for reading for each Year level cohort, plotted at its average age. Age cohorts are identified as 8, 10, 12 and 14 although the actual points are plotted at 8.5, 10.5, 12.4 and 14.4.

Figure A5.2 Comparison of models developed from three data points compared to four data points- for Reading, 2008



A thicker line is drawn for the 'All data' case where all four data points are used to fit the curve. For this initial exploration the Test score at age 0 is taken as 0, and it included as an additional point. The impact of alternative assumptions about the 'best' assumptions for the Test score at age 0 is explored in a subsequent section. The purpose of this section is to understand the effect of missing data by systematically leaving out known data.

The 'All data' curve is regarded as the best estimate of the trajectory for reading growth since it draws on all available data (that is all four points at 8.5, 10.5. 12.4 and 14.4). Ignoring data points that are not at the highest and lowest age has little impact of the trajectory fitted for the data. The 'All data' curve and the curves that use three points that include the lowest and highest age (8 10 14, 8 12 14) produce almost identical trajectories in the range from age 6 to 15.

The two models where either the highest or lowest data points are not included produce slightly different trajectories as marked on the chart. Logically the case that deletes the highest point (8 10 12) tends to a lower asymptote. The case that deletes the lowest point (10 12 14) tracks generally lower than the 'All data' curve below age 10. However in the range of ages from 8 to 14, the models with the widest time separation (even if missing one data point) produce very similar trajectories.

The chart also illustrates why the Gompertz model is attractive as a model for the general trajectory. The asymmetrically placed inflection point offers a possible explanatory mechanism for what may be happening (at a group level with quite a spread of performance) with reading development with age. Accepting the sigmoid shape as a reasonable general model, and the asymmetric Gompertz curve as an appropriate choice, a model for the 'rate of learning' is provided. The curve of the annual rate of learning at each point on the trajectory for 'All data' is shown and scaled on the right axis. The rate is increasing as the inflection point (at about age 6) is approached from the left. On the curves fitted to the 2008 reading data the rate of learning is increasing rapidly for the ages from 3 to 6, peaks at about age 6 (at about 66 scale points per annum) and then reduces from ages 6 to 15.

There is however a potential impact on the fit of models from the choice of an appropriate value for the 'notional' test score at birth (age=0). The test scale zero is arbitrary and it is assumed that the NAPLAN developers gave no consideration to the choice of an 'absolute zero' for learning. If the Gompertz model applies, the degree of fit is markedly influenced by the choice of Test score value at age=0. This issue is considered in the following section as part of the development of a general model for reading learning trajectories.

Test scale transformation effects on a model for the trajectory of NAPLAN Reading 2008

In developing a likely trajectory for reading, the impact of alternative placements of the zero position for the test scale at age = 0 were investigated under the Gompertz model. Figure A5.3 shows the effects of various placements of the origin on the location of the fitted curve. Each case is a displacement from the original position and then the rescaled data are fitted to a Gompertz model. Once the model has been fitted, the data are rescaled to the original scale. As an example consider the case of OS-100. The data are rescaled by subtracting 100 from each of the original four Test scale means, fitting the new values to a Gompertz model, calculating the model value for each age from 0 to 18 and then adding 100 to these values. With this procedure the 0 position on the original scale was raised 100 units, the curve fitted and then the points were recalculated relative to the previous origin.

Figure A5.3 Effect of changing the location of 0 on the Test scale at age =0

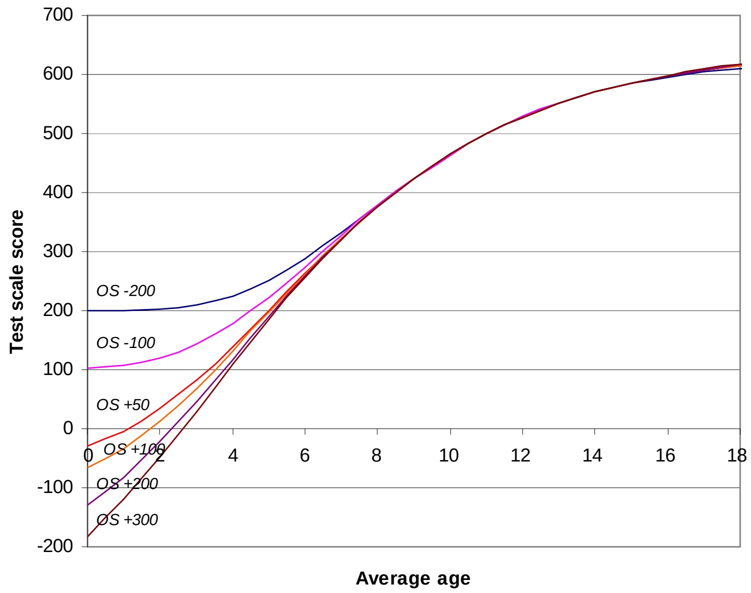


Figure A5.3 shows that the curves of all models are effectively coincident from age 7 to 16. The variation in the Test 0 position influences the shape of the trajectory from age 0 to 7. It does this by repositioning the inflection point. The higher up the scale the 'real 0' is assumed to be, the higher the inflection point. For OS-200 (Original scale minus 200) the inflection point is close to 8 years of age, for OS+300 the inflection point is around 4 years. The curve past the inflection point remains essentially the same, whatever the assumed Test zero. This exploration establishes that while the assumed zero position is important in determining at which age the inflection point occurs, the trajectory after this age is essentially the same for all cases within a range of 250 scale units above or below the original zero position. This means that the curve segments for the ages for which data are obtained by tests are essentially the same. The choice of origin influences the steepness of the curve below age 8 and the model determined point at which the rate of learning peaks. This insight implies that the match of the extrapolation of learning status for younger ages in a Gompertz model is problematic but adds some general considerations for hypothesis testing.

A reconsidered model based on the original test scale

Given the insight above that the selection of the zero position for the test scale in the Gompertz model influences the location of the inflection point, the selection of the appropriate value for the test scale at age=0 should be based on the assumption (or known reality) of the likely age for the highest 'rate' of learning. A reasonable assumption for this period is in the first year of school, the ages of 5 to 6. As it turns out the NAPLAN scale, unchanged, fitted to a Gompertz model, has its highest rate of learning in the period from age 5 to 6. For this reason the further refinement of the model to match data points for Australian students is based on the original test scale. As indicated earlier the purpose for doing this is to provide an understanding of what cohort growth appears to look like in Australian school systems, to help select parameters for extrapolating data as outlined earlier is addressed in Chapter 6.

While the original scale is used, the value assigned to the test scale at age = 0 remains a matter of choice. The effect of the value of the test score at age = 0 can be considered through varying the value for the test score at this point, in a narrower range than explored above, to see if there is a optimum position for the data point at time 0, i.e. age = 0. Figure A5.4 shows the effect on the parameters of the fitted curves, on the standard error of the fit and the correlation of data points with the model, as the value on the test score axis at age 0 is systematically varied.

A matter to clarify is where the Gompertz model would place the value of the test score at age=0. In seeking a fit to all four data points (8,10,12,14), the modelling software iterates a solution that has parameters a, b and c as 642.1, 1.38, 0.251 in four iterations, with a standard error of 2.66. By adding a range of fifth points (age=0, and test scores varying from 0 to 60) model fit can be optimised by author-managed alternative models to minimise the standard error and maximise the coefficient of correlation (R^2). While the four data points (ages 8,10,12,14) are fitted to the model in four iterations, it is possible to find a value for a fifth point, the test score at age zero, that reduces the required iterations to 3 and minimises the standard error of the fit of the data to the model, and maximises the correlation of coefficient of the fit. This point is found at Test scale =12. Figure A5.4 panels show the variations in the parameters as the test scale value at age zero is varied. With a value of 60, the asymptote (a) increases to over 700, relative to a value of 623 at Test scale =0.

In Figure A5.4 the model with Test scale =12 has an asymptote at 642 the same as the four points only solution. The effect of the placement of the fifth point at Test scale = 12 is to reduce the iterations to solution and to reduce the standard error to below 2. While the 5th point is possibly unnecessary, the optimised Y-axis intercept is close to the value of 13 derived in Figure A5.2. As noted in that case, a model based on all states and the ACT (but not including the NT) has an asymptote of 649 (648.97) compared with 642 above and a model-selected intercept of 13. The model based on the national averages only, has parameters a, b, c and a Y- intercept close to those resolved in Figure A5.2.

Based on the Test scale value of 12 at age 0, and the four test scale mean points, a general model of the national age cohort trajectory for reading can be plotted as illustrated in Figure A5.5.

Figure A5.4 Panels of Parameters and Fit indicators for Options for Test score at Age=0

Gompertz Relation expression: y=a*exp(-exp(b-c*Age))

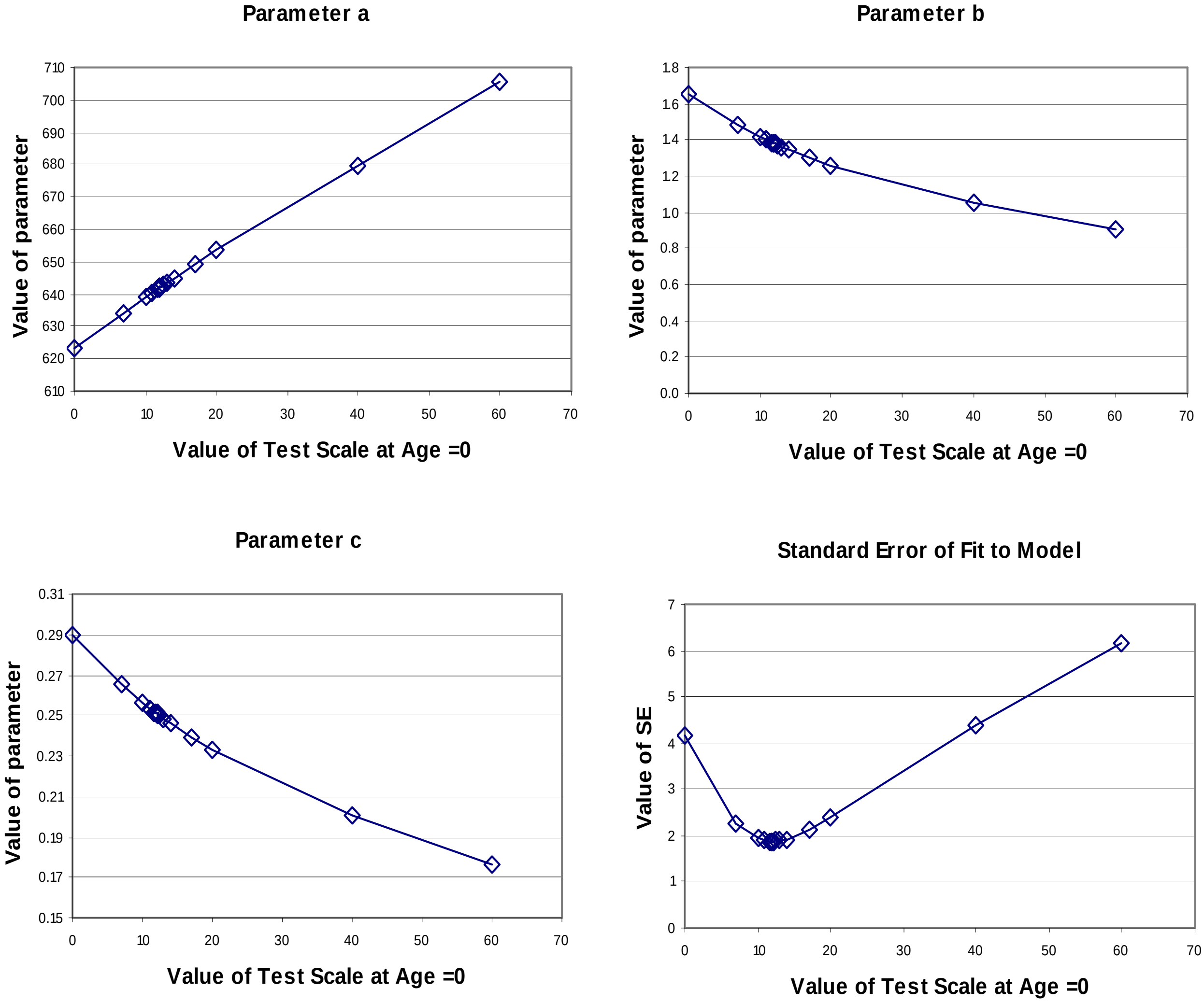
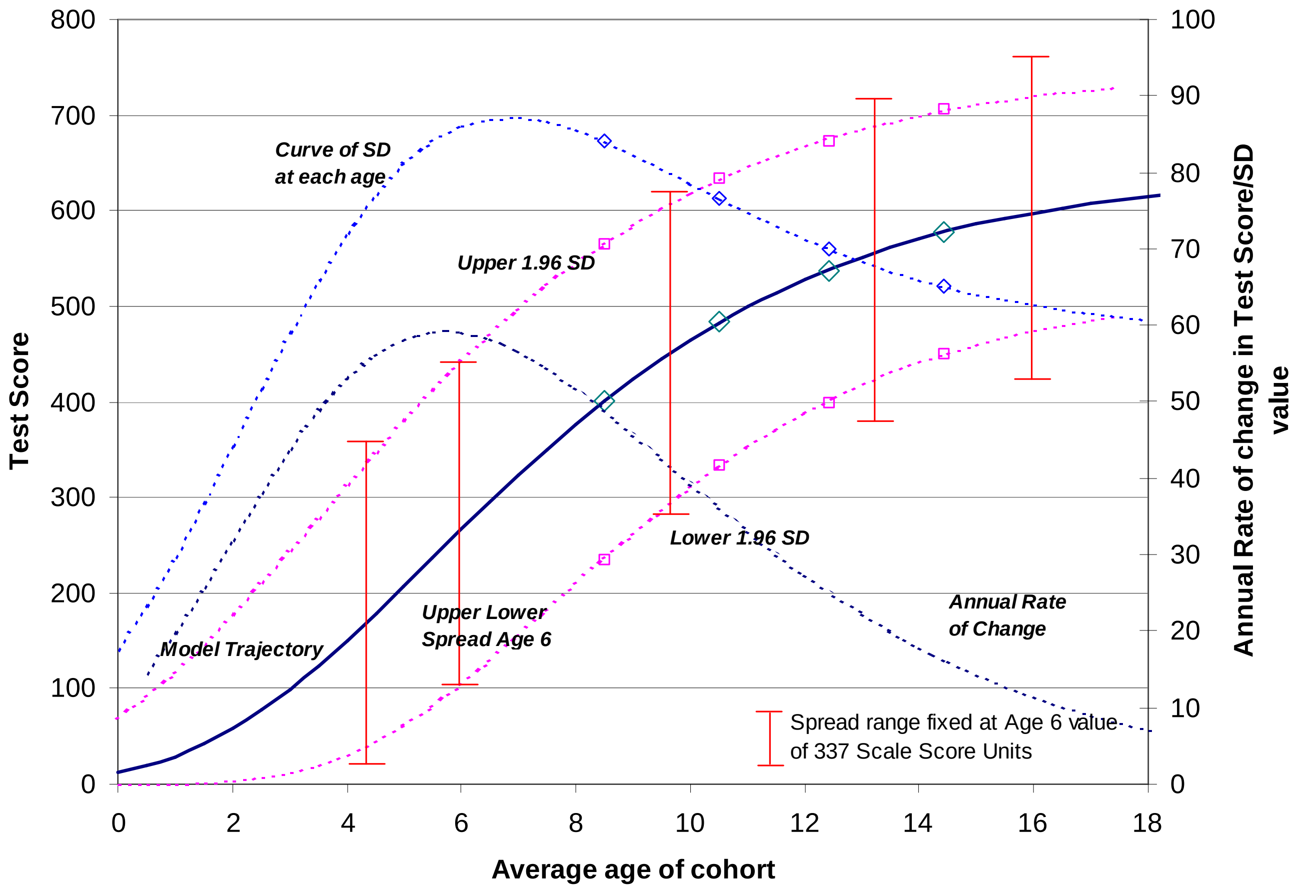


Figure A5.5 Model of NAPLAN Reading 2008 with indication of spread of data



The model in Figure A5.5 illustrates the annual rate of change in reading development at each point on the age scale by comparing reading status at successive age points. Estimates of the general patterns of the SDs can be made. The points at 1.96 SDs above and below the four national means encompass 95%of the student scores at each age. Curves can be fitted to the four actual points that delineate these upper and lower boundaries of the 95% of cases (assuming a normal distribution at each age point). The upper curve has an asymptote at 751 and notional test score intercept at age = 0 of 68 test scale units. The lower curve has an asymptote at 521 and a notional test score intercept at age = 0 of 0.08 test scale units. By subtracting the lower boundary values from the upper boundary values an estimate of the SD at each age point can be made by dividing the resulting value by 3.92 (2 x 1.96). On this basis, estimates of the SD can be made for any age points, enabling as a consequence the estimation of the effect sizes for annual growth at those age points (assuming n as the values used in generating the data in the first place)[35].

The resulting SD estimates are plotted with their scale on the right hand axis. The estimated SDs start small, grow to about 87 test scale units at about age 7 and reduce from that point on. Based on the observation of Schulz & Nicewander (1997) that growth spurts (eg puberty for

---

[35] In a model where all students are spread on the age axis at their actual age at testing the general distribution of data points remains essentially the same. The value of n for any point is the cohort n divided by the number of age categories for the cohort. Since values of n above 50 make little difference to the effect size calculation the actual n is almost immaterial.

human height) lead to greater variance at that spurt point, it is confirmed in this model that the SDs are greater around points of rapid growth, that is near the inflection point. The peak SD lags the peak rate of learning development by about a year. The peak rate of learning is around 6 years, the peak SD around 7. In this model logical and mathematical reasons are provided for the 'scale shrinkage' (Yen, 1986; Camilli, Yamamoto & Wang, 1993), the shrinkage of SD within a year level (very small) and more observable, the reduction of SD at higher Year levels.

The estimated annual rate of learning at each age is also plotted on the right hand scale. This curve illustrates an implication of the model. Students near the trajectory of the mean are likely to be learning at their maximum rate about age 6.

On the basis of estimating the SDs from the modelled upper and lower bounds, Figure A5.5 reflects a model of what the data might look like if students were assessed by an appropriate process at all ages from 0 to 15 on the one day and those data plotted by the average age for the students, aggregated to cohorts of average age, say in 0.1 decimal years of age at testing. An alternative view is that if all student scale scores were plotted individually by actual student age at assessment as individual points, 95% of these points would lie within the upper and lower bounds assuming a normal distribution by age on the test scale.

The actual data points reflecting the upper and lower bounds based on actual data points are identified on the curves, along with the actual scores on the model trajectory and the actual SDs.

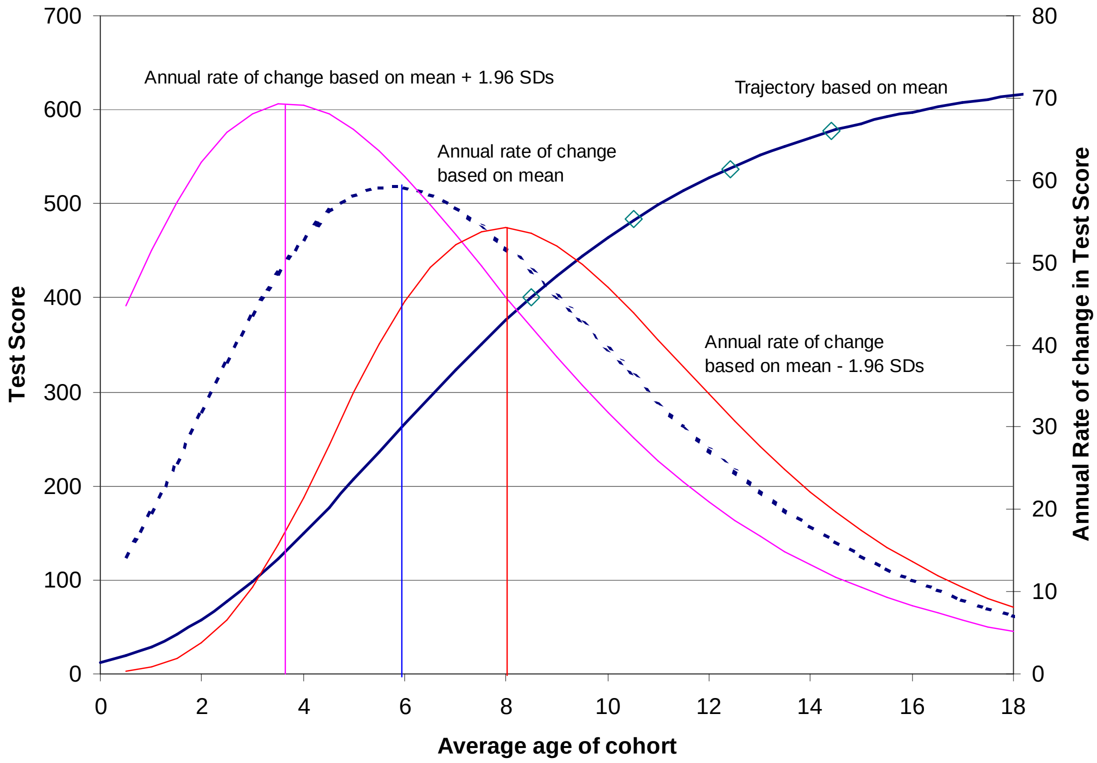
The bars on the chart are all of constant length, based on the estimated SD at age 6, and indicate visually the reducing spread of the scores at higher ages. Based on the explorations of test scale transformations above, speculating about what happens below the inflection point will be inaccurate in the estimate of the actual trajectory from age 0 to 6. A diminishing SD is however plausible because the rate of growth is smaller and the actual quantum of learning that is possible is less. A better basis for estimating the range of 'pre reading skills' would provide a better model. The author has allowed the trajectory to start at age=0 for completeness. It can be assumed that at some future point a better understanding of learning of the appropriate skills in younger children could be incorporated into such a speculative model.

Further speculative implications of the NAPLAN Reading model

The model can be explored further by plotting the annual rates of learning for the mean in comparison to the rates implied in the upper and lower boundary curves.

Figure A5.6 shows these three speculative annual rates of learning. The small number of students near the upper boundary are learning at a high rate and peak at just below 4. The group near the average have a gentler increase in rate and peak at 6. The students near the lower bound start slowly and reach a peak at age 8. Surprisingly the model suggests their rate of learning continues at a higher rate than the mean group or the upper tail. Were this simple model anywhere near a model of reality, some hypotheses about individual learning trajectories could be tested using individual longitudinal data. The issues relating to individual trajectories are addressed briefly later in the Chapter 5 and Appendix 10.

Figure A5.6 Model of NAPLAN Reading 2008 with annual rates of change in learning at mean, 1.96 SDs above and 1.96 SDs below the mean



While the model illustrates in an approximate way what data might look like if say all students were assessed at the one point in time and their data plotted, the model can also estimate what the mean score for a cohort at a particular cohort average age might look like. Using the model in this way enables the effect size for 'normal' year-to-year growth to be estimated.

Impact of data spread on State comparisons

A specific insight from the curve-fitting with average age at testing is the impact of this placement of the test mean on comparisons between states and territories (or school systems generally), where the average ages for systems vary markedly from each other or from the national average age. A similar situation applies when the data are analysed by elapsed years of schooling rather than age. Both differ from the impression gained when 'age' is centred on Year level. Table A5.1 lists the grand means of the displacements from the national test score mean at each year level for each system, averaged over the four Year levels, compared with the displacement from the fitted curve (Gompertz).

Table A5.1 NAPLAN Reading 2008 – Comparison by State and Territories by displacement from National mean

	Mean Age at Year 9	Mean displacement from National mean (averaged over all four Year levels)	Mean displacement from curve (averaged over all four Year levels)	Difference (Curve Displacement minus –Nat. Mean Displacement)
Qld	14.08	-16.48	-0.57	**15.90**
WA	14.00	-10.58	-0.27	**10.30**
NT	14.42	-73.78	-70.85	**2.93**
SA	14.50	-3.15	-2.29	**0.86**
NSW	14.58	8.30	8.33	**0.03**
ACT	14.67	21.25	18.77	**-2.48**
VIC	14.75	11.20	6.24	**-4.96**
Tas	14.83	-2.20	-11.23	**-9.03**

The table is ordered by the impact of the difference between the two methods of establishing displacement. Queensland is displaced from the national means at each Year level by an average of -16.46 test scale units. Apart from the Northern Territory this is the largest negative (below mean) displacement. However when the mean test scores for Queensland are compared with the curve of mean test scores by average age at testing as the X axis, as shown in the right panel of Figure 5.1, the displacement from the curve is less than one test scale unit. Effectively when an adjustment is made for the lower average age in Queensland at each Year level, Queensland sits almost where it would be expected to be, accepting the plotted curve as a reasonable model for scores by average age. Similarly Western Australia is almost where it would be expected (-0.27 units from the curve). NSW and SA maintain the same relationship with the curve model as they do with the national mean (due to the closeness of their average ages to the national average age). The ACT's displacement is slightly reduced as the average age is slightly higher than the national average. Victoria, which on the national mean score comparison is 11.2 units above the national mean, is only 6.24 units above where the model by age would place it. Tasmania, which appears to be close to the national mean, is actually 11.23 units below the curve when age is considered. The Northern Territory being at the national average age is unchanged in the comparison. Whether or not the model suggested is the most appropriate, the principle is established that an age adjustment when making national comparisons produces quite different conclusions.