Appendix 7 Mathematics Assessment for Learning and Teaching (MaLT) in England

Williams, Wo, & Lewis (2007) and Ryan & Williams (2007) report data from a national sample designed to provide age related performance references for the MaLT test. Year level cohorts of between 1000 and 1400 students were recruited from 111 schools.

Data are summarised on a time dimension calibrated in months. The test was developed using the Rasch model. Vertical equating was achieved partly by common persons for adjacent Year levels (about 1/3 of the cohorts sat adjacent level tests). Common item equating was also applied in the test development phase where about half the items for the next Year level for pre-test cohorts were included in the lower level (Williams et al., 2007, p. 132).

A scatter plot of all students across all age categories, in months is presented as a 'Quintic' model in Williams et al. (2007, Figure 1, p. 134) with test scores reported as logits. From this figure it was possible to estimate some broad values to develop a model as illustrated in Figure A7.1. The data in this model are estimated from the published scatter plots rather the r taken from tables.

The data are presented in Williams et al. (2007) as 5 trajectories representing the paths of the 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentile students. To develop a model similar in form to those in previous appendices, readings of data points from the 50^{th}, 90^{th} and 10^{th} percentile trajectories were made. The assumption was made that the median was very close to the mean and the median then used as a proxy for the mean. The readings for the medians at a set of ages (60, 72, 84, 96, 108, 132, 168 months) were taken. Readings were taken for the same ages on the 90^{th} and 10^{th} percentile trajectories.

A model was estimated through curve fitting using the Gompertz equation for the three trajectories. To achieve this the logit scores were transformed to a scale with 0 logits = 200 and one logit transformed as 20 scale units. Once the model was fitted the results were reconverted to the original logit scale. As in previous appendices, the impact of the choice of the zero position of the scale score influences the shape of the 'modelled' trajectory below age 5 but with little impact above. Transforming the original scale with zero at 200 provided an inflection point at about 5 years of age. Transforming the original scale with zero at 400 produced an inflection at about 3 years of age.

The trajectories of the upper and lower boundaries for 95% of the data were estimated by first developing the model for the 90^{th} and 10^{th} percentiles, establishing the spread between the two lines at age points, re-scaling the area under the normal curve from 80% to estimate a SD (by dividing by 2.56 (2*1.28)). The upper and lower boundaries are then estimated from 1.96 * SD.

Figure A7.1 displays the resulting model for the mean, the actual data points (as estimated) and the upper and lower boundaries for 95% of the data. Also plotted are the annual rate of change, based on the model and the estimate of the SD. Consistent with the NAPLAN model the model SD reduces slightly as the age increases and peaks about a year of age past the inflection point. As previously the model estimates can be used to estimate the effect sizes for year-to-year growth. These are shown in Figure A7.2. For reference the US effect sizes from Hill et al. (2007) for 6 Mathematics tests are included (taken from Chapter 5, Table 5.1).

Figure A7.1 Model of Mathematics Development - Mathematics Assessment for Learning and Teaching, (MaLT)

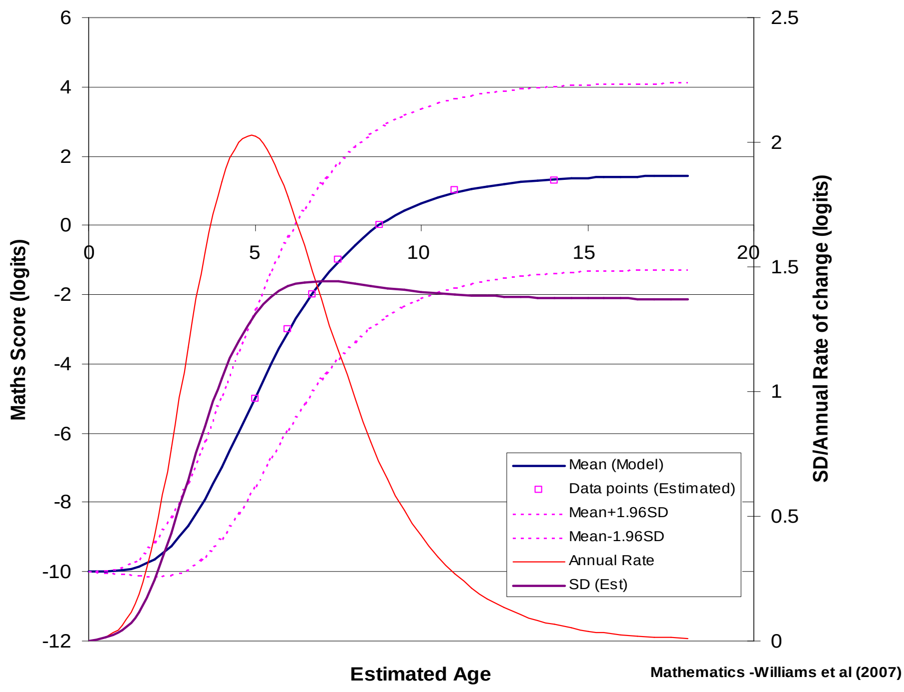
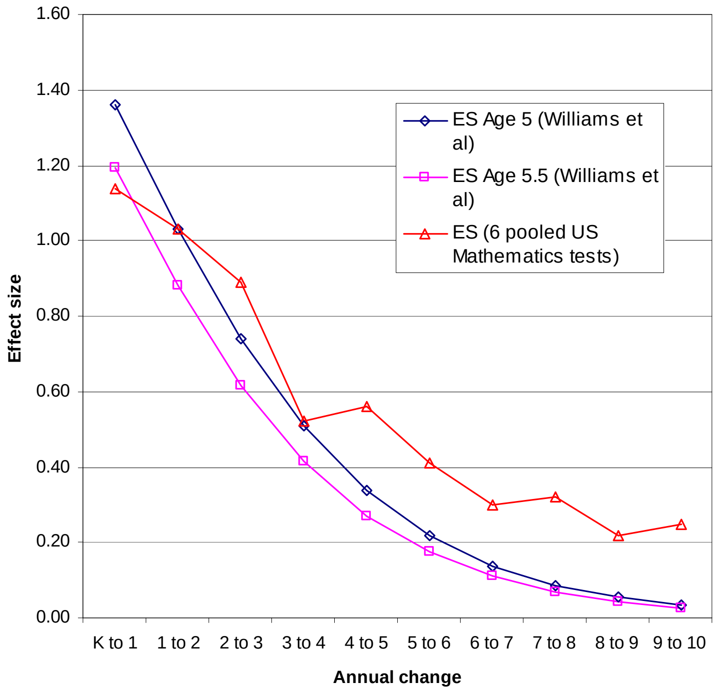


Figure A7.2 Effect sizes for Mathematics Assessment for Learning and Teaching compared with pooled US tests.



The general pattern of reducing learning growth in mathematics is exhibited, through reducing effect sizes, in both the Williams et al. and US data. Somewhat surprisingly, for the UK model, the growth in logits per annum from Figure A7.1 and the effect size in SDs from Figure A7.2 are both close to zero by the transition from Year 8 to 9, much lower than the US comparison. (The Year 10 to 11 effect in both figures is an extrapolation- Williams et al. only tested to Year 9). It was this plateau effect that was the focus of the Williams et al. (2007) article since it implies almost no mathematics development from Years 7 to 9.

The validity of the vertical scale is considered in the article but even with the qualifications to the phenomenon that might be related to the inadequacy of the scaling, Williams et al. conclude that even though the

> …the plateau … must be interpreted with the limitations of the vertical equation methodology in mind …. closer examination of the three year model with superimposed one-year models confirms that the plateau is not an artifice of the full 10 year vertical equating model. It seems realistic to conclude that progress is indeed very slow (about 0.2 logits per year) over this period. …… One speculates that the repeated exposure to the same curriculum in secondary school has a negative effect on these common learning outcomes. (Williams et al., 2007, p. 139)

The Williams et al.(2007) data are also helpful in illustrating the age effect within a Year level cohort. This phenomenon appears to apply quite generally and is discussed in Chapter 5.