

NAPLAN Results – Fact or Fiction?

NAPLAN 2009 Results – Fact or Fiction?

Margaret Wu

University of Melbourne

Email address: m.wu@unimelb.edu.au

Phone: +61 3 8344 0963

Fax: +61 3 9348 2753

Postal address:

Assessment Research Centre,

234 Queensberry Street,

Melbourne Graduate School of Education,

University of Melbourne,

Victoria,

Australia, 3010

Keywords: trends, equating error

This report discusses the NAPLAN 2009 summary results released on September 11, 2009 (http://www.naplan.edu.au/verve/resources/2009_NAPLAN_Summary_Report.pdf). In comparing the NAPLAN 2009 results with those from long established U.S.'s National Assessment of Educational Progress (NAEP), NAPLAN 2009 results appear anomalous for Grammar & Punctuation, Reading and Numeracy.

In this report, the magnitude of change in achievement from one year to another is measured in *effect size*. Effect size is typically defined as the amount of change in standard deviation unit. For example, one might find that between 2008 and 2009, the mean score has increased by 0.1 of a standard deviation of numeracy scores for Year 3 students. In this case, the effect size is said to be 0.1. The advantage of using effect size is that effect sizes can be compared across different assessments. In addition to effect size, the unit of “typical months of growth” is also used as a reference point. For example, one might say that the gain in average achievement scores in a subject area is around two months of growth.

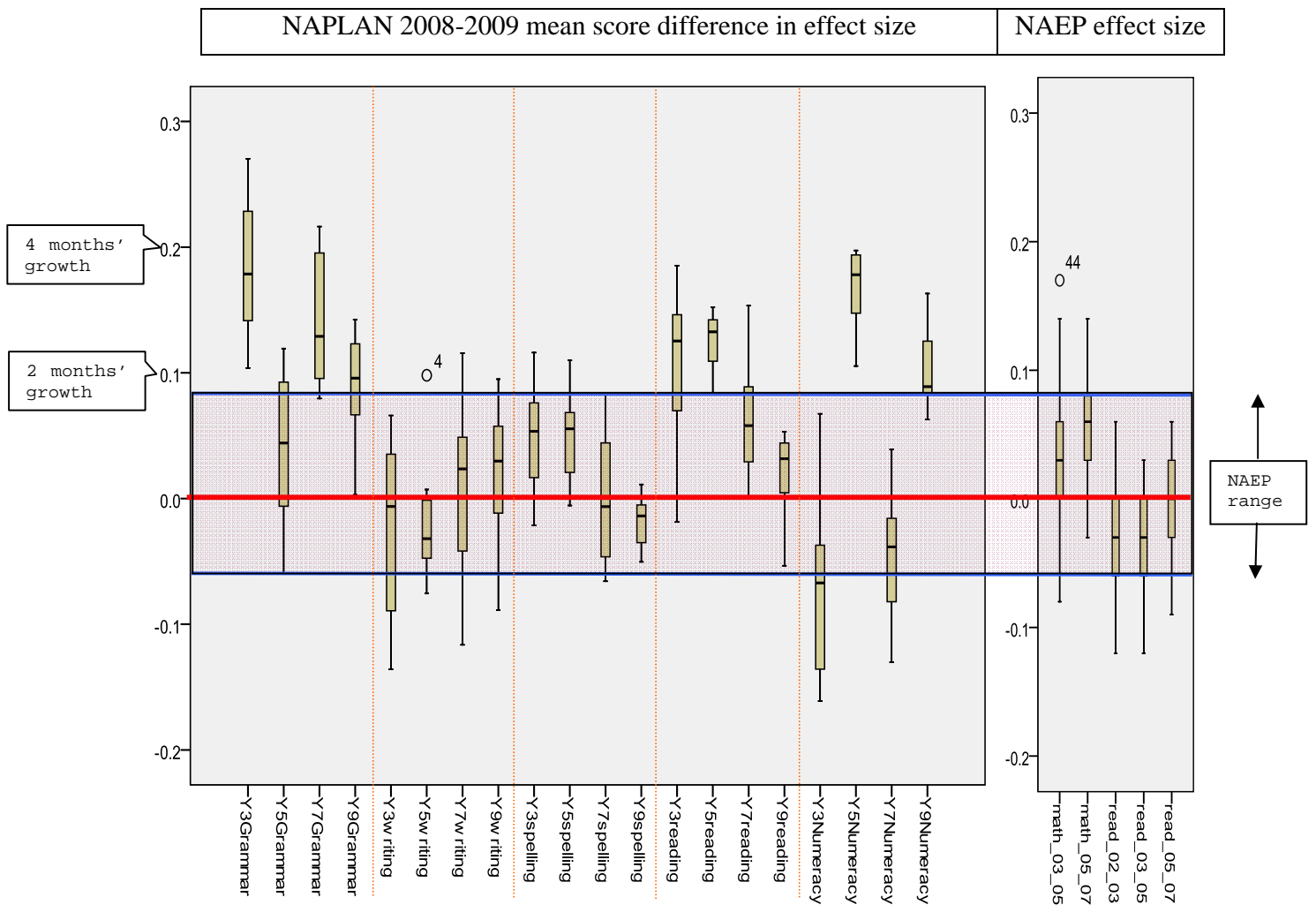


Figure 1 Changes in NAPLAN 2008-2009 Mean Scores Expressed in Effect Size, compared with NAEP

Figure 1 shows boxplots of changes in NAPLAN 2008-2009 mean scores expressed in effect size, compared to the effect size found in NAEP. For NAPLAN, each boxplot is for one year level and one subject area, and the data consist of the change scores in effect size of eight jurisdictions. For NAEP, the data for each boxplot come from around 50 states. The NAEP effect size is measured for two-year intervals (except for one boxplot) rather than for adjacent years as in NAPLAN.

On the right-side of Figure 1, the NAEP (U.S.A.) trend results show that typically, the inter-quartile range of effect sizes is between -0.1 and 0.1. Using this as a reference range, the effect sizes for NAPLAN Grammar & Punctuation, Reading and Numeracy appear to be anomalous.

For example, for Y3 Grammar & Punctuation, all jurisdiction means have increased by a considerable amount. The effect size for Australia is .18. The effect sizes for the 8 jurisdictions range between .10 and .27. In substantive terms, an effect size of 0.20 is around four months of growth. While this effect size is not at all likely for any jurisdiction, the fact that ALL 8 jurisdictions had such a large growth suggests that there are systematic errors in the data analysis, most likely in the equating process, that account for these aberrant results.

For Y7 Grammar & Punctuation, Y3 and Y5 Reading, Y5 and Y9 Numeracy, all effect sizes are too high. For Y3 Numeracy, the effect sizes are too far below zero. With seven out of 20 subject areas showing aberrant results, it is difficult to have confidence in the overall NAPLAN 2009 results.