How Has School Productivity Changed in Australia?*

Andrew Leigh
Research School of Social Sciences
Australian National University
andrew.leigh@anu.edu.au
http://econrsss.anu.edu.au/~aleigh/

Chris Ryan
Research School of Social Sciences
Australian National University
chris.ryan@anu.edu.au
http://econrsss.anu.edu.au/Staff/ryan/contact_cr.htm

Abstract

Using two series of data that ask overlapping questions to successive cohorts, we estimate how the literacy and numeracy skills of young Australian teenagers (aged 13-14) have changed over time. We find a small but statistically significant fall in numeracy over the period 1964-2003, and in both literacy and numeracy over the period 1975-1998. The decline is in the order of one-tenth to one-fifth of a standard deviation. Adjusting this decline for changes in student demographics does not affect this conclusion; if anything, the decline appears to be more acute. Next, we estimate long-run changes in real per-child school expenditure. This estimate varies somewhat according to the treatment of private spending, and the chosen price index, but our preferred estimate suggests that real per-child school expenditure increased by 10 percent over the period 1975-1998, and by 258 percent over the period 1964-2003. This increase in spending funded a substantial reduction in student-teacher ratios. Measuring productivity in terms of literacy and numeracy points per dollar, our results imply that the productivity of Australian schools may have fallen over the past 3-4 decades. Although we cannot account for all the phenomena that might have affected test score results, we identify a number of plausible factors that might have led to a drop in school productivity.

Keywords: education production function, literacy, numeracy
JEL Codes: H52, I21, I22

* This paper was originally prepared as a report for the Australian Department of Education, Science and Training (DEST). The views expressed are those of the authors and do not necessarily represent the views of the DEST. The authors thank Dr Sheldon Rothman of the Australian Council for Educational Research for providing one set of the school achievement scales used in this paper. He bears no responsibility for our use of those scales. In revising this report for publication, Elena Varganova provided outstanding research assistance.

I. Introduction

All too frequently, education policy debates focus on inputs rather than outputs. To a large extent, this is a function of the available data. While input measures such as class sizes, teacher salaries and per-student funding are readily comparable over time, output measures such as test scores are often not designed to be compared across years (either because they are re-standardised each time they are administered, or because the test instrument itself is changed from year to year).

In this paper, we present data on one output of the educational production process – literacy and numeracy scores of Australian children aged in their early teenage years. Combining data from two nationally representative sets of tests, we compare numeracy scores from 1964 to 2003, and literacy scores from 1975 to 1998. This is a measure of schooling quality, which we then compare with per-child school expenditure.

Although literacy and numeracy scores are by no means the sole output of the education process, equipping children with good reading, writing and mathematical skills is nonetheless an important function of schools. At an individual level, studies have shown that Australian students with better literacy and numeracy in their early teenage years are more likely to be employed when aged in their twenties (Marks and Fleming 1998a), and that conditional on being employed, tend to earn higher wages (Marks and Fleming 1998b). At a national level, countries where the labour force is more numerate tend to grow faster than countries with lower levels of numeracy (Hanushek and Kimko 2000; Jamison, Jamison and Hanushek 2006).

Only a small number of studies have looked at changes in test scores and spending over several decades. For the United States, Hanushek (1997) showed that US test scores were essentially flat over the period 1970-94, while real per-child expenditure grew by 2.5 to 3 percent per year. Hoxby (2001) demonstrated that this finding held true for the period 1970-98, and that adjusting for student demographics made little difference to the overall pattern. Similarly, Gundlach, Wossmann and Gmelin (2001) find that test scores in 11 OECD countries were basically flat over the period 1970-94, while real per-child funding rose dramatically (though their findings are limited because their approach to linking school performance over time is to benchmark countries against the United States[1]). So far as we are aware, all other studies to have looked at school productivity since the 1960s/70s have relied upon the US National Assessment of Educational Progress (NAEP).

Our analysis of school productivity follows in the extensive literature on public sector productivity (see eg. Rosen 1993; Atkinson 2005; Boyle 2006; Douglas 2006; Weale 2007). As Atkinson (2005) noted, "There is a strong case for devoting significant resources at this time to improving the measurement of public sector output on account of its increased saliency in policymaking and public debate". However, as Rosen (1993) has pointed out, one hesitation that arises when assessing public sector productivity is choosing the appropriate measure. Since no metric perfectly captures all aspects of the effectiveness of a public sector organisation, it is important to recognise that any chosen output measure is merely a proxy variable. Notwithstanding their limitations, studies that have analysed educational

[1] Gundlach, Wossmann and Gmelin (2001) claim that "The problem with measuring the quality of schooling output over time is that consistent time-series data on the cognitive achievement of pupils are available only for the United States, where the National Assessment of Educational Progress (NAEP) began to monitor the performance of US pupils aged 9, 13 and 17 years in mathematics and science in the early 1970s." As our paper demonstrates, this is incorrect. Indeed, the Australian tests allow comparison over an even longer interval than the US NAEP, albeit with fewer observations.

productivity have typically opted to use test scores as their primary output measure (see eg. Atkinson 2005). For example, the UK Department for Education and Skills (2005) noted that it had considered other output measures, such as school inspections – but concluded that test scores were preferable on the basis of data availability, transparency and their potential to be linked to labour market outcomes. In Australia, the Productivity Commission (SCRGSP 2007) lists its preferred outcome measures as test scores for reading, writing, numeracy, science and civics; plus ICT literacy performance, VET in schools, completion, destination, and "other areas to be identified".[2]

To preview our results, we find no evidence that the test scores of Australian pupils have risen over the past four decades, and some evidence that scores have fallen. This finding is consistent with earlier Australian studies using some of the same datasets (Afrassa and Keeves 1999; Rothman 2002). We explain how this finding can be reconciled with evidence on the 'Flynn Effect' (growth in measured IQ scores over the twentieth century in developed countries), and with Australia's ranking on international tests. We also find that adjusting for the observed demographic characteristics of students does not affect this conclusion; if anything, the decline appears to be more acute. We also review other possible explanations, such as changes in early school leaving behaviour, and conclude that they are unlikely to have significantly affected trends in test scores.

[2] Our analysis does not use school completion rates as a metric of school quality, on the basis that students' decisions to obtain more schooling do not necessarily reflect higher school productivity. In the Australian context, youth labour market conditions have been shown to be extremely important. Ryan and Watson (2004) noted that the parameters of the regression equation they estimated, in conjunction with the change in variable values, would have attributed the entire increase in Australian Year 12 retention rates from the 1970s to the 1990s to the decline in full-time job opportunities for teenagers.

Turning to education expenditure, we observe a significant increase in per-child spending from the mid-1960s to the early-2000s. If we measure school productivity in terms of the average test score divided by the average real per-child expenditure, this implies that the productivity of Australian schools has fallen over the past four decades. Using our preferred measure of school input prices, we estimate that school productivity has declined by 12-13 percent between 1975 and 1998, and by 73 percent between 1964 and 2003.

The remainder of this paper is structured as follows. Section II outlines the test scores that we use, and presents the basic trends in literacy and numeracy. Section III examines the extent to which the observed patterns might be due to changes in demographics, or other factors beyond the school gate. Section IV looks at how per-child spending has changed. The final section concludes.

II. Trends in Literacy and Numeracy

Longitudinal Surveys of Australian Youth

The first set of tests that we use to analyse changes in literacy and numeracy are data from four Longitudinal Surveys of Australian Youth (LSAY) cohorts. These cohorts are the Youth in Transition 1961 and 1975 birth cohorts (YIT 61 and YIT 75) and the Longitudinal Surveys of Australian Youth 1995 and 1998 Grade 9 cohorts (LSAY 95 and LSAY 98). Those in the first cohort sat the literacy and numeracy tests in 1975, while those in the last cohort sat the tests in 1998. Although subsequent cohorts of the LSAY have been surveyed, the test instrument is not comparable to that used in earlier cohorts and changes to the survey design mean that the subjects in the later cohorts are older.

Students in the first two LSAY cohorts sat standardised tests at age 14, while students in the last two cohorts sat standardised tests in grade 9. There are four ways that one might compare achievement across time using these data: by using the full sample in all years, by restricting the sample to 14 year olds, by restricting the sample to year 9 students, or by restricting the sample to those who were aged 14 and in year 9.

We adopt the last approach. That is, we opt to compare cohorts by focusing only on the common group: those who were aged 14 and in grade 9. This is designed to limit the impact of changes in the age-grade composition of the surveys over time, since both students' ages and grades may affect their performance on achievement tests.

This is a different approach from the preferred specification used by Rothman (2002), who argues that using those aged 14 and in grade 9 suffers from the problem that underperforming students were more commonly required to repeat grades in the 1970s than in the 1990s. Rothman cites Australian Bureau of Statistics data that shows that the share of 14 year olds who were enrolled in grade 8 or below fell from 24 percent in 1975 to 16 percent in 1995-2000, and argues the average quality of the pool of students aged 14 and in year 9 might have been lower in the later cohorts. Therefore, average school achievement for students aged 14 and in year 9 in the later cohorts would be expected to be lower than in earlier cohorts.

Our view is that any such impact from lower grade repetition on our results from the use of the LSAY data is likely to be very small over this period. First, there were offsetting trends on age-grade distributions across Australian jurisdictions. Grade repetition rates certainly fell over the period in jurisdictions where they were high initially. In other jurisdictions, however, there were trends towards later school commencement, especially among those who would have been youngest in their grade cohorts, and once more this phenomenon might be expected to have affected students who were of below average ability. Hence in some jurisdictions, the proportion of students 'old' for their grade, given the school commencement rules operating in their jurisdiction, actually increased in the later LSAY cohorts, presumably increasing the average ability of the students observed aged 14 and in year 9 in the data, while in others it did not change.

Second, empirically, the shares of the cohorts aged 14 and in year 9 did not change as much in the LSAY data as it did in the ABS data. The (weighted) proportion of the sample aged 14 and in year 9 in the YIT 61 cohort was 60.3 percent. In the LSAY 98 cohort this proportion was 63.7 percent. In unweighted data, the change was just 1.5 percentage points. The reason the change in the proportions is smaller in the LSAY data is one of coverage – students' integer ages in LSAY were recorded as of the 1st of October in the survey year, and in the YIT surveys only those aged 14 as of that date were included in the surveys. In the ABS data, ages are classified as of the 1st of July. Because of the complex pattern in which students born in different times of the year are distributed across school grades in Australia, which also differs substantially by jurisdiction, this different coverage means that the LSAY data has been much less affected by the trend identified by Rothman (2002) (and compositional effects driven by differential population growth between states) than the ABS data.

Finally, it is possible to estimate how much such phenomena might have affected our results. First, we can identify students aged 14 and in year 9 who should come from similar parts of the ability distribution in the different cohorts. These are individuals born in the same months of the year from the same state, but in different cohorts, where the proportion of the sample aged 14 and in year 9 were the same. It is also possible to identify cohorts of students where the distribution changed substantially in the way identified by Rothman (2002), that is where the proportion aged 14 and in year 9 increased. We find the trends we present below for the entire group aged 14 and in year 9 are also borne out for these other groups, but that the decline in literacy and numeracy performance is most pronounced in the group where the proportion aged 14 and in year 9 increased most, the effect identified by Rothman (2002).[3]

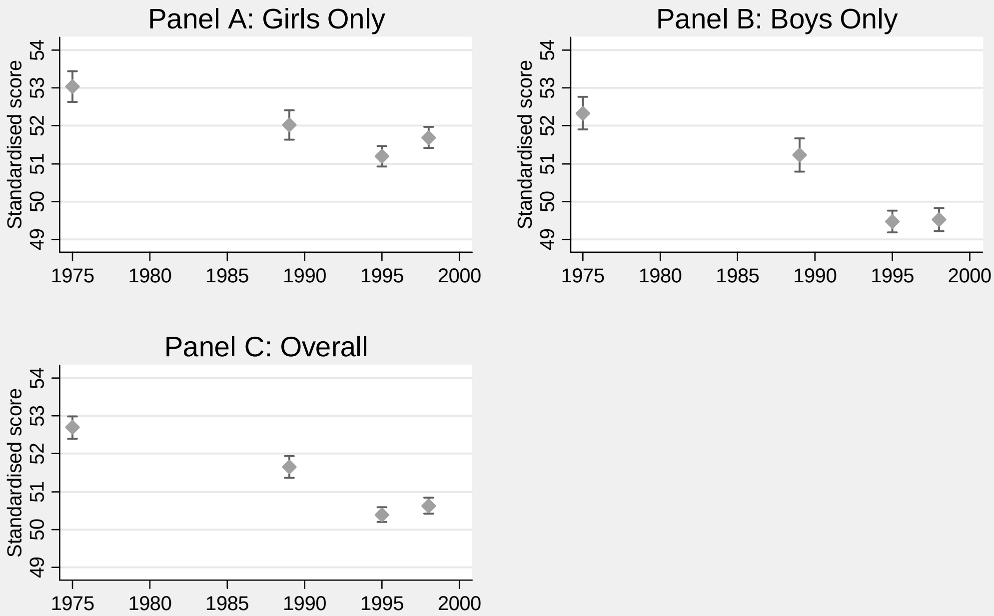
The LSAY literacy and numeracy scales used here are taken from Rothman (2002). They were developed using the common items asked of the cohorts using a Rasch item response model. The scales were constructed to have a mean of 50 and standard deviation of 10 across the cohorts.

Figures 1 and 2 contain trends in test scores from the LSAY, with dots representing the mean, and error bars representing the 95 percent confidence interval for the mean. Across all tests, the mean of the standardised test scores is set at 50 and the standard deviation at 10. The results show that over the period 1975-98, there was a statistically significant decline in the literacy and numeracy test scores of both boys and girls.[4]

[3] Moreover, the trends are identical where we also exclude individuals who appear ‘old’ or ‘young’ for their grade given the typical age-grade level of other in their state born in the same months of the year.

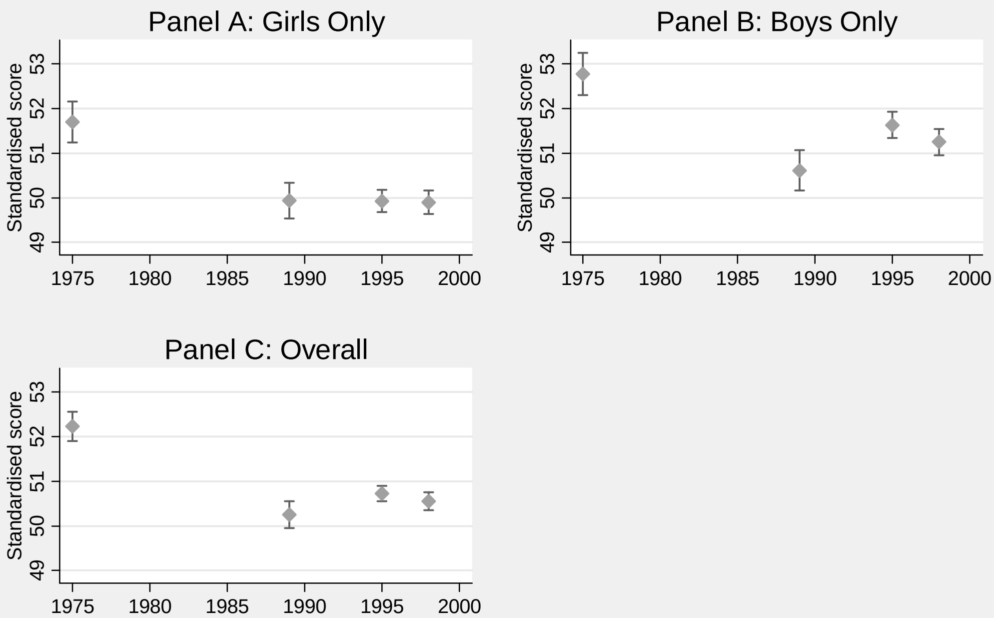
[4] This finding is sensitive to our choice of students aged 14 in year 9. Results presented in Rothman (2002) show that when the comparison uses the full sample, or just 14 year olds, literacy and numeracy scores are flat from 1975-98.

Figure 1: Literacy (LSAY)



Note: Sample is 14 year olds in grade 9. Capped spikes depict 95% confidence interval for mean.

Figure 2: Numeracy (LSAY)

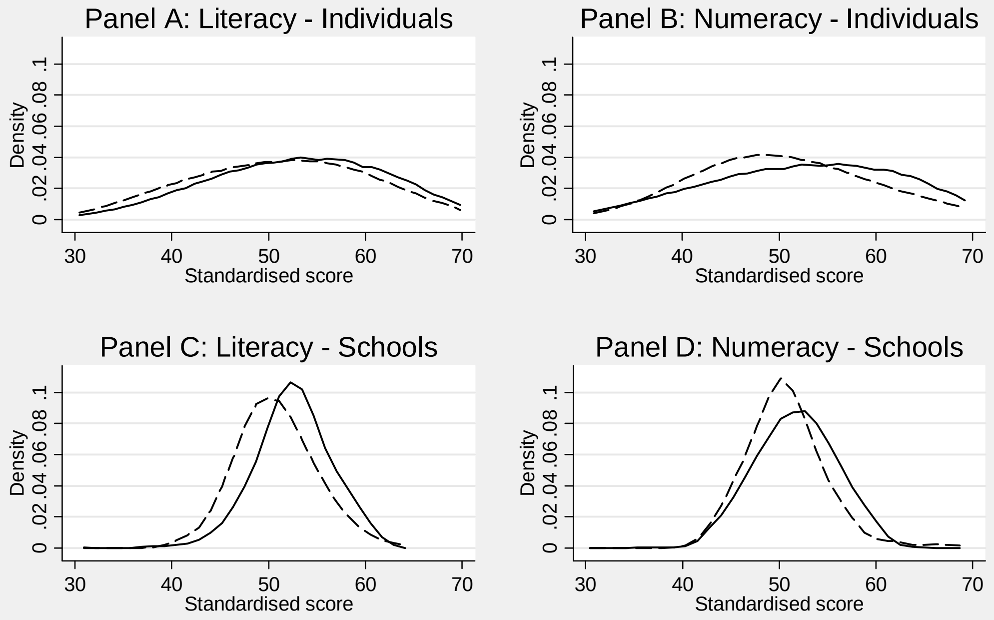


Note: Sample is 14 year olds in grade 9. Capped spikes depict 95% confidence interval for mean.

Was the fall in mean scores driven mostly by changes in the left tail or right tail of the distribution? And how did the distribution of test scores across schools change over time? To answer these questions, Figure 3 presents kernel density plots (effectively smoothed histograms) of the distributions in 1975 and 1998. In these graphs, the horizontal axis shows the standardised score, while the vertical axis shows the share of observations that are at or close to each score. Panels A and B show the distribution of individual literacy and numeracy scores. In the case of individual literacy scores, it appears that the distribution has simply moved towards the left. By contrast, the distribution of individual numeracy scores has also changed shape slightly, with fewer children scoring around 60, and more scoring just below 50.

Panels C and D show the distribution of school average literacy and numeracy scores. In both cases, the graphs overlap on the far right tail, indicating a similar number of high-performing schools in both years. Elsewhere, the graphs appear to have essentially shifted to the left. For literacy, the distribution of school performance has become slightly more dispersed over time; while for numeracy, the distribution of performance across schools has become slightly more compressed over time.

Figure 3: Distributions (LSAY)



Note: Sample is 14 year olds in grade 9. _____ 1975, - - - 1998.

IEA Surveys

The second source of data that we analyse are five mathematics surveys conducted by the International Association for the Evaluation of Educational Achievement (IEA). These are the 1964 First IEA Mathematics Study, the 1978 Second IEA Mathematics Study, the 1994 Third IEA Mathematics Study (also known as the 1995 Third International Mathematics and Science Study or TIMSS, though it was conducted in Australia in 1994), the 1999 TIMSS, and the 2003 TIMSS.

In comparing the cohorts, we need to take into account shifts in the composition of test-takers. There are three aspects to this. First, the geographic coverage of the test steadily increased over time, with the 1964 survey covering five states (NSW, VIC, QLD, WA and TAS), the 1978 survey covering these plus the ACT and SA, and the 1994 and subsequent surveys also covering the NT.[5] Second, the surveys extended their coverage across school types, with the 1964 survey covering only government schools, and later surveys covering both government and non-government schools. And third, the surveys changed their age/grade coverage, with the 1964 survey covering 8th graders, the 1978 survey covering 13 year olds, the 1994 survey covering 7-9th graders, the 1999 survey covering 8-9th graders, and the 2003 survey covering 8th graders.

To take account of these various shifts, we use a Rasch item response model to estimate a standardised score, using common questions across the tests.[6] For ease of comparison, we then express the tests on the same scale as the LSAY results (mean 50, standard deviation 10). Where T is the test score of student i, tested in year y, in grade g, and of age a, we estimate the following regression:

$$T_{iyga} = \alpha + \beta_y I^{test\ year}{}_y + \gamma_g I^{student\ grade}{}_g + \delta_a I^{student\ age}{}_a + \varepsilon_{iyga} \quad (1)$$

The parameter β_y denotes how mean scores have changed, holding constant changes in the age and grade composition of students taking the test.[7] Since the version of the 1994-2003

[5] Although the ACT was part of the NSW schooling system in 1964, ACT schools do not appear to have been sampled in the 1964 survey (Afrassa and Keeves 1999).

[6] The Rasch item response model is used to create a standardized score for the 1964, 1978 and 1994 tests. However, not all the common items in the 1964-94 tests also appear in the 1999 and 2003 tests. We therefore take advantage of the fact that for the 1994-2003 tests, the IEA has already created a variable called "National Math Rasch Score". We standardize this variable so that in 1994, it has the same mean and standard deviation as the Rasch score derived from common items in the 1964-94 tests.

[7] On the face of it, this approach differs from that used to analyse the LSAY data. If we estimate the equivalent of equation (1) with LSAY data using all observations with student age, grade, school sector and state controls, the declines between the first and last cohorts for all students are 1.7 and 1.2 points for literacy and numeracy respectively, smaller but close to the estimates in Figures 1 and 2. These declines were statistically significant at the 1 percent level. The alternative methodologies do not produce results that differ qualitatively. Where we also address the potential endogeneity of age and grade in the equation using LSAY data

datasets that we are using does not contain information on state, nor on whether the student attended a non-government or government school, we first estimate the model just for 1964 and 1978, with the sample restricted to the student population covered by the 1964 test. We observe a decline of 2.5 points, which is significant at the 1 percent level. We then estimate the model using data for all tests, and find that the drop from 1964 to 1978 is 2.0 points (still significant at the 1 percent level). Given the similarity between these two estimates, we opt to compare across all cohorts.

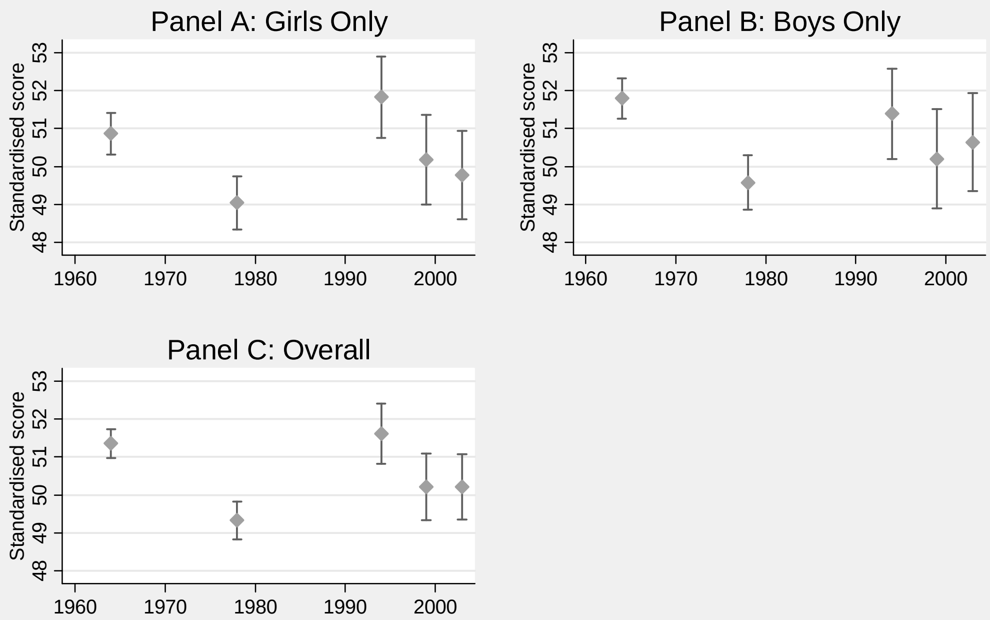
Figure 3 charts the results. Over the 39 years from 1964 to 2003, we observe a statistically significant decline in test scores, with the typical student scoring 1.1 points lower in 2003 than in 1964 (significant at the 5 percent level). When we analyse boys and girls separately, the drop is of similar magnitude for each group, and statistically significant at the 10 percent level. Rescaling the scores to a common scale, the mean scores (for boys and girls combined) were 51.4 in 1964, 49.3 in 1978, 51.6 in 1994, and 50.2 in both 1999 and 2003. In 1999, the most recent year that the IEA surveys covered multiple grades, students' mathematics scores rose by an average of 4 points per grade. These findings therefore imply that the numeracy of the typical young teenage student in 2003 was approximately a quarter of a grade level behind his or her counterpart in 1964.

Although we use a somewhat different methodology, our results from 1964 to 1978 are close to those of Afrassa and Keeves (1999), who find a decline of approximately half a year of

(which is another way of characterising Rothman's argument for using the whole Year 9 cohort) the results do not change in any qualitative way. We use as instruments for age and grade student birth month and its interaction with school commencement rules. These rules, which vary between jurisdictions and changed over time, largely determine students' ages and grades when surveyed in the LSAY cohorts. The declines in performance between the first and last surveys were 1.4 and 0.8 points for literacy and numeracy, respectively, with the former significant at the 1 percent level and the latter at the 5 percent level.

mathematics learning between 1964 and 1978. However, while Afrassa and Keeves find a decline of approximately one year of mathematics learning between 1964 and 1994, we observe no change scores over this period. This is most likely due to the fact that our results adjust for both age and grade effects, while Afrassa and Keeves only adjust for grade effects.

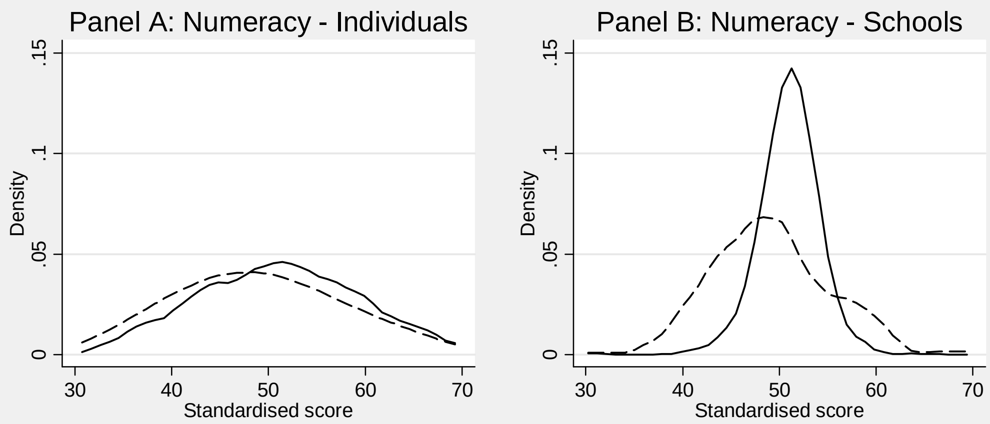
Figure 4: Numeracy (IEA)



Note: Effects are adjusted for student age and grade. Capped spikes depict 95% confidence interval for mean.

To see whether the fall in mean scores was driven by changes in particular points of the distribution (either across individuals or across schools), Figure 5 presents kernel density plots of the distributions in 1964 and 2003. As with the LSAY density functions, the horizontal axis shows the standardised score, while the vertical axis shows the share of observations that are at or close to each score. Panels A shows the distribution of individual and numeracy scores – indicating that the distribution has simply moved towards the left, with little change in dispersion. Panel B shows the distribution of school average numeracy scores. At a school level, the distribution has shifted to the left over time. In addition, there appears to be greater school-level dispersion. However, the apparent change in dispersion is probably due mostly to changes in the sampling frame. While the 2003 IEA test sampled schools across Australia, the 1964 test only included government schools in five states. The omission of non-government schools from the earlier test is the most likely reason why the school-level distribution appears narrower in 1964 than in 2003.

Figure 5: Distributions (IEA)



Note: Effects are adjusted for student age and grade. ____ 1964, ---- 2003.

Other Evidence – Flynn Findings

In addition, it is worth mentioning other sources of data on the cognitive abilities of Australian children which may (at least initially) appear to contradict the results presented here. The first is the well-known ‘Flynn Effect’, under which IQ scores in many developed nations have shown an increase during the twentieth century (Flynn 1987, 2006). If the test scores of young teenagers are stagnant, how can it be the case that IQ scores are rising in the general population?

There are three answers to this. First, in the case of Australia, Flynn's data are drawn from an earlier era. Table 1 juxtaposes Flynn's results with those shown above. We follow Flynn in converting the LSAY and International Mathematics tests into the IQ scale (mean 100 and standard deviation 15). While Flynn (1987) demonstrated that Australian test scores grew by 0.4 to 0.5 points per year in the interwar period and the immediate post-war decades, the data shown above indicate that the last three decades of the twentieth century saw a slight decline in test scores.[8]

Table 1: Long Run Test Score Changes in Australia

Age	Test	Period	Total Change (LSAY scale)	Total Change (IQ scale)	Years	Change per year (IQ scale)
Panel A: Flynn Findings (1936-81)						
10-14	Jenkins	1949-81	10.45	15.67	32	0.490
10-16	Ravens	1950-76	5.84	8.76	26	0.337
10-14	Otis	1936-49	3.67	5.50	13	0.423
Panel B: Trends in School Tests (1964-2003)						
14	LSAY Literacy	1975-98	-2.07	-3.10	23	-0.135
14	LSAY Numeracy	1975-98	-1.67	-2.50	23	-0.109
13	IEA Numeracy	1964-2003	-1.14	-1.71	39	-0.044

Source: Panel A is from Flynn (1987). Panel B is from authors' calculations. LSAY scale has a standard deviation of 10. IQ scale has a standard deviation of 15. All changes are statistically significant at the 5 percent level or better.

The second point to be made about Flynn's findings is that – across a wide range of countries – they appear to be more robust when tests are administered to adults rather than to younger children. Flynn (1987) presents 'strong data' on test score gains for seven countries. In five out of seven of these countries, the data are for adult subjects. As he acknowledges, one factor that might be driving higher test scores among the adult population is higher school

[8] Contrary evidence comes from Nettlebeck and Wilson (2004), who found that the Peabody Picture Vocabulary Test scores of a small sample of Australian school children rose between 1981 and 2001. However, both of their samples are drawn from a single primary school, making it quite possible that the observed changes were due to local demographic shifts, and were not nationally representative.

completion rates, and rising rates of post-secondary education. Even if school-aged test scores were flat, more years of education might be expected to boost adult test scores.

The third factor to bear in mind when reconciling Flynn's findings with those presented here is that the Flynn effect is largest for standard IQ tests (such as the Ravens, Stanford-Binet and Wechsler tests), and virtually absent for curriculum-based tests. Discussing evidence for the United States and other developed countries, Flynn (1999) notes that long-run test score gains are: "small to nil on achievement tests. That is to say, they fall away the closer we come to the content of school-taught subjects".

Together, these three factors – different time periods, different ages, and different types of tests – suggest that the absence of a discernable increase in Australian test scores of young teenagers on a curriculum-based test is readily reconcilable with the available evidence on the Flynn effect.

Other Evidence – International Comparisons

The other piece of evidence that may (to some eyes) appear to contradict the long-run trends for Australia is the fact that Australia ranks highly when compared with other developed countries. Pooling data for three international tests: the 1995, 1999 and 2003 Trends in International Maths and Science Study (TIMSS), the 2000 and 2003 Programme for International Student Assessment (PISA), and the 1994-8 International Adult Literacy Survey (IALS), Brown et al. (2005) compare rankings for a group of 18 OECD countries.[9] This comparison has the advantage that all of the countries are of a similarly high income level (on

[9] Although IALS covers the entire adult population, Brown et al. (2005) only analyse respondents aged 16-24.

a purchasing power parity-adjusted basis, Australia's GDP per capita is ranked 9th among these 18 countries).

Pooling these data sources, Brown et al. find that the ranking of the median Australian respondent is, on average, 6.6 out of 18, indicating that the literacy, numeracy and scientific knowledge of young Australians is somewhat above the OECD average. For PISA reading and for TIMSS and PISA mathematics, the typical Australian student is 2/3rds to 3/4s of a standard deviation above the OECD mean. For PISA science and TIMSS science, the typical Australian student is nearly a full standard deviation above the OECD mean, and for the IALS literacy tests, the typical Australian student is at or slightly below the OECD mean.

Just as a rich country can experience a period of no economic growth (yet still remain better off than many other countries), so Australia's ranking on international tests can easily be reconciled with the failure of test scores to rise over recent decades. Countries that rank reasonably well in international comparisons are not necessarily also improving over time. Indeed, the findings of Gundlach, Wossmann and Gmelin (2001) indicate that test scores in OECD countries were essentially flat over the period 1970-94. It is therefore unsurprising that the failure of Australian test scores to rise has not dragged the nation to the bottom of the international league tables.

III. Alternative Explanations

In this section, we canvass several possible explanations for the fact that test scores have not risen over time. Our main focus is on the possibility that demographic shifts led scores to be lower than they would otherwise be. We then briefly address a number of other alternative explanations.

Demographic Shifts

One plausible explanation for the decline in test scores over time is that the demographic characteristics of students have changed in such a way as to affect test scores. For example, it tends to be the case that students from a language background other than English (LBOTE) do worse on literacy and numeracy tests (see eg. Rothman 2002). Since the share of students has risen over time, this would be expected to reduce the average test scores. Conversely, it tends to be the case that students with university-educated parents do better on literacy and numeracy tests (see eg. Cardak and Ryan 2006). Since the share of students with university-educated parents has risen over time, this would be expected to raise average test scores. Thus the overall effect of changing demographics is ambiguous.

To separate the effect of changing demographics, we employ a technique known as Oaxaca decomposition (Oaxaca 1973). Although traditionally used to decompose gender and racial pay differences, Oaxaca decompositions have also been employed to look at test scores (see eg. Cook and Evans 2000). In the present case, such a decomposition allows us to separate the change in test scores into three components: changing attributes of the student body, changing returns to these attributes, and shifts that are not explained by demographics. However, it is important to note that such an exercise is limited by the fact that our datasets do not contain a comprehensive set of demographic characteristics. We return to this issue below.

In the case of comparing LSAY test scores from 1975 and 1998, the Oaxaca decomposition involves estimating the following regressions, where t denotes the test score, Z_j is a vector of student demographics, α is a constant, β_j are estimated parameters, ε is a normally distributed mean-zero error term. Subscript i indexes individuals, j indexes demographics, and 75 and 98 denote the 1975 and 1998 tests respectively.

$$t_{i75} = \alpha_{75} + \beta_{j75} Z_{ij75} + \varepsilon_{i75} \quad (2)$$

$$t_{i98} = \alpha_{98} + \beta_{j98} Z_{ij98} + \varepsilon_{i98} \quad (3)$$

Denoting mean scores in 1975 and 1998 as T_{75} and T_{98} respectively, the change in mean test scores can be written as:

$$\Delta T = T_{98} – T_{75} = (\alpha_{98} + \beta_{j98} Z_{j98}) – (\alpha_{75} + \beta_{j75} Z_{j75}) \quad (4)$$

Equivalently,

$$\Delta T = \beta_{j98}(Z_{j98} - Z_{j75}) + Z_{j75}(\beta_{j98} - \beta_{j75}) + (\alpha_{98} - \alpha_{75}) \quad (5)$$

The first term on the right side of equation (5) is the change in test scores that can be attributed to the change in student demographics; the second term is the change due to changes in the coefficients on those demographics, and the third term is the change that cannot be explained by demographics or coefficients.

To see this more precisely, note that if student demographics are exactly the same in 1975 and 1998, then the first term should be zero. If the returns to student demographics are exactly the same in 1975 and 1998, then the second term should be zero. And if the change from 1975 to 1998 is entirely explained by changes in demographics and the returns to those demographics, then the third term should be zero.

In addition, it is important to note that the effects need not go in the same direction. That is, it is perfectly possible for the returns to student demographics to contribute to a rise in test scores and an unexplained component to contribute to a fall in test scores.

Table 2 presents results. Among the sample of students for whom we have non-missing demographic characteristics in the LSAY, the decline in test scores from 1975 to 1998 is only around half as large as across the full population (-1 point for literacy, and -0.6 points for numeracy). However, when we decompose the shift, we find that shifts in demographic characteristics should have acted to increase test scores over time. Taking account of shifts in the levels of and returns to student demographics, the decline in test scores is considerably larger: around 3.4 points for literacy and 3.9 points for numeracy.

For the IEA test, our approach is very similar, except that because the age and grade coverage of the test changes over time, we take as the dependent variable the residual from a regression of the student's test score on indicator variables for age and grade. This can be thought of as parsing out differences in test coverage. The Oaxaca decomposition then compares the 1964 and 2003 test cohorts. Among the sample of students for whom we have non-missing demographic characteristics in the IEA, the decline in test scores from 1964 to 2003 is similar to the decline in the full sample. When we decompose the shift, we find that shifts in demographic characteristics should have acted to increase test scores over time. Taking account of shifts in the levels of and returns to student demographics, the decline in test scores is 7.6 points. However, this is partly a function of the very limited demographic characteristics available in the IEA tests. If we had the same demographic variables in the IEA as we have for the LSAY, we expect that the adjusted decline in test scores would be smaller than 7.6 points.

Table 2: Decomposition of Changes in Test Scores

	(1) **LSAY Literacy (1975 to 1998)**	(2) **LSAY Numeracy (1975 to 1998)**	(3) **IEA Numeracy (1964 to 2003)**
Change in average test score (all students in sample)	-2.07	-1.67	-1.14
<u>Decomposition</u>			
Change in average test score (students with non-missing demographics)	-1.00	-0.58	-0.90
<u>Earlier year as base</u>			
Change attributable to levels of student demographics	+1.44	+1.92	+3.64
Change attributable to returns to student demographics	+0.98	+1.36	+3.02
Change not attributable to observed demographics	-3.42	-3.86	-7.56
<u>Later year as base</u>			
Change attributable to levels of student demographics	+1.21	+1.19	+3.81
Change attributable to returns to student demographics	+1.21	+2.09	+2.85
Change not attributable to observed demographics	-3.42	-3.86	-7.56

Note: LSAY: Sample is 14 year olds in grade 9, and demographics are indicators for student gender, whether the school is in a metropolitan area, whether it is a Catholic school, and whether it is an Independent school, whether the student was born in a non-English speaking country, whether the student's mother was born in a non-English speaking country; the share of students in the school with a mother born in a non-English speaking country; the number of siblings that the student has; whether the student's father has a degree, whether the student's mother has a degree; parents' occupational status (on the ANU 3 scale); and indicators for each state and territory. IEA: Test scores are the residual from a regression of the Rasch score on indicators for age and grade. Demographics are student gender, an indicator for whether the student/school is in a metropolitan area, and indicators for whether the student's mother and father have lower secondary education, upper secondary education, a trade qualification, or a university degree.

In interpreting the results of this decomposition, it is important to note that in both datasets, the demographic variables available to us are less extensive than one might like. In particular, we lack precise information on whether the child is in a single-parent family, and our data on neighbourhood characteristics is not as precise as would be ideal.

Have 1960s and 1970s Questions Become Irrelevant?

Another possible alternative explanation for the observed changes is that the questions asked in earlier years are simply irrelevant today. For example, if it were the case that literacy questions in the mid-1970s required a knowledge of words that had fallen out of common usage by the 1990s, or if mathematics questions in the mid-1960s required a level of mental arithmetic that many would judge to be unnecessary in the early-2000s, one might wonder whether it was reasonably possible to compare performance on the two tests.

Although there is not a comprehensive way to address this critique, one straightforward approach is to look at the tests themselves. Appendix 1 presents several examples of questions that were common across IEA tests. In our view, these common questions are reasonably representative of modern-day mathematics tests, although they are all computational rather than conceptual. It is therefore possible that the IEA comparison placed more emphasis on computational skills.

Marks and Ainley (1997) analysed responses to the tests conducted in 1975 and 1995 as part of the LSAY series and found that there were counteracting tendencies in both the literacy and numeracy tests. For the literacy tests, the report found students in the later cohort were more likely to answer correctly questions relating to newspaper articles, but were less likely to answer questions dealing with more difficult textual passages correctly. For the numeracy tests, Marks and Ainley found that students' performance was poorer on computational items in the later cohort, but had improved on conceptual items.

Early School Leaving

One plausible explanation for the failure of test scores to rise over time is the possibility that in our earliest cohorts, a significant number of students dropped out of school before the test was conducted. If a substantial number of children dropped out, and those who left school would have scored below average on the test, then early school leaving in the early years might have biased upwards the earliest test results. However, note that if all students obeyed the compulsory school leaving laws, this should not have occurred. In the 1964 IEA, most students were 13 years old, an age at which Australian children were required to attend school in all states and territories. In the 1975 LSAY, all students were 14 years old, again an age at which Australian children were required to attend school in all parts of the country.

To test this, we checked official statistics on the enrolment of the cohort that were to be 13 year olds in 1964 (using the *Australian Yearbook* for various years). If early dropout was a problem, one would expect to see that the school enrolment of this cohort was larger in prior years. However, we observe little evidence of this. By comparison with the cohort of 13 year olds in 1964, the cohort of 12 year olds in 1963 was 0.14 percent larger, the cohort of 11 year olds in 1962 was 0.02 percent larger, and the cohort of 10 year olds in 1961 was actually 0.32 percent smaller. This kind of trivial variation suggests that in the early-1960s, very few children dropped out of school between their 10th and 13th birthdays.[10]

---

[10] Three other factors affecting this calculation are immigration, emigration, and death. The average net immigration rate (inflows minus outflows) was 0.75% in 1961-64 and 0.49% in 1971-75. In both periods, young teenagers were probably underrepresented in population

We conducted a similar exercise on the first LSAY cohort, comparing prior years' enrolment with the age cohort who were to become 14 year olds in 1975 (unfortunately, we were not able to obtain enrolment statistics for 1974).[11] By comparison with 14 year olds in 1975, we found that the cohort of 12 year olds in 1973 was 1.26 percent smaller, the cohort of 11 year olds in 1972 was 1.27 percent smaller, and the cohort of 10 year olds in 1971 was 0.68 percent smaller. This implies that in the early-1970s, about 1 percent of children dropped out of school between their 10th and 14th birthdays. To see the largest effect that this attrition could have had on the mean test scores, suppose that those who dropped out of school would have scored a full standard deviation lower than those who remained. If the mean score of those who stayed had been 50, this implies that the true mean score would have been 49.9. Such variation would have had virtually no effect on the observed results.

Social Trends

Another possible factor that might have lowered test scores is if there had been an exogenous increase in violence in schools. Such an increase would have made the job of schools more difficult, potentially counteracting other factors that would have caused scores to rise. To test

---

movements. The annual probability of death for a person aged 10-15 in the 1960s and 1970s was around 0.03-0.04%. Both sets of figures indicate that these factors are unlikely to make a significant difference to our results.

[11] To see whether the absence of 1974 enrolment data affected the results (and to avoid potential problems created by combining enrolment data from different data sources), we replicated the analysis using data from New South Wales (NSW), Australia's largest state. All data were obtained from the annual state yearbooks. By comparison with 14 year olds in 1975, we found that the cohort of 13 year olds in 1974 was 1.05 percent smaller, the cohort of 12 year olds in 1973 was 2.29 percent smaller, the cohort of 11 year olds in 1972 was 2.48 percent smaller, and the cohort of 10 year olds in 1971 was 2.50 percent smaller. This implies that in the early-1970s, about 2 percent of children in NSW dropped out of school between their 10th and 14th birthdays. Supposing that NSW dropouts would have scored one standard deviation below average, this would have biased the true result upwards from 49.8 to 50.0.

this, we would ideally have liked comparable data on school-related violence, which was unaffected by changes in punishment regimes. While we were unable to obtain such data, our analysis of the literature on youth crime rates in Australia suggests that violent and property crime rates have stayed constant through the 1980s, 1990s and 2000s (see eg. Wundersitz 1993, 2005; Carcach 1997; AIHW 2007).[12] It therefore seems unlikely that school-related violence has increased markedly over this period.

Other factors are more difficult to measure. Over the last four decades of the twentieth century, divorce rates and rates of television viewing increased significantly, which could potentially have driven down test scores. It is also conceivable that there might have been a change in students' attitudes to learning over this period.

IV. Expenditure Trends

In Australia, we have been unable to find a consistent series of school expenditure covering the period since 1964.[13] We therefore combine a number of different data sources to produce a consistent series of school expenditure, including both government and private spending. Government expenditure is obtained from various official tabulations, and includes expenditure from all levels of government, provided to both public and private schools. Private expenditure is calculated from household expenditure surveys, from which we estimate private spending as a share of government spending. Taking account of changes in

[12] Wallace (1986) finds a rise in homicides by offenders aged 15-19 between 1958-67 and 1968-81. However, this is partly explained by the higher share of the population in this age band in the latter period.

[13] In their study of schooling productivity in OECD countries over the period 1970-94, Gundlach, Wossmann and Gmelin use education expenditure data reported to UNESCO (see the appendix to Gundlach, Wossmann and Gmelin 1999).

the share of children attending public and private schools, we then estimate private school spending in all years. Appendix 2 provides details of the derivation of our series.

One possible limitation to this estimate is that we are unable to disaggregate spending into primary and secondary school expenditure. Since per-pupil expenditure is typically higher for secondary school pupils than for primary school pupils, and since rising school completion rates mean that the share of pupils in secondary schools is higher in the early-2000s than in the mid-1960s, this will bias upwards the trend in school spending. Empirically, however, the extent of this bias is very small. In 2002-03, per-student spending in government schools was 28 percent higher at the secondary school level than at the primary school level (MCEETYA 2003, Appendix 1: Statistical annex, Table 20). Over the period spanned by our tests – 1975-98 and 1964-2003 – the share of pupils in secondary school grew by 4.6 and 4.4 percent respectively.[14] This means that the upward bias to school spending caused by an increase in the share of students at the secondary school level was only 1.3 percent.

To adjust for price changes, we index the expenditure series in three ways. First, we adjust using the all-groups CPI, which assumes that schools' input prices rose at the same price as other goods and services in the economy. Second, we construct a schools price index (an index of the prices of inputs used by schools), to account for the possibility that the price of schools inputs has been growing at a different rate from other prices. And third, we construct an index based on the earnings of professional women, on the assumption that the largest

[14] An eagle-eyed reader may wonder at the fact that the secondary share grew more over the period 1975-98 than over the period 1964-2003. It is important to remember that the figure is not only affected by school completion rates (which have risen steadily), but also by cohort-specific factors, such as the baby boom.

expenditure item for schools is teacher salaries, and around 60-70 percent of teachers were women in this period.[15]

Panel A of Table 3 sets out the proportionate change in total schools spending over time. Using the All Groups CPI, spending increased by 18 percent over the period of the LSAY, and by 333 percent over the period of the IEA tests. Using the Schools CPI, spending increased by 10 percent over the period of the LSAY tests, and by 258 percent over the period of the IEA tests. And indexing schools expenditure by the earnings of professional women, spending declined by 2 percent over the period of the LSAY, and rose by 76 percent over the period of the IEA tests. The substantially lower expenditure growth rates over the period spanning the LSAY data reflects the strong spike in per capita expenditure just prior to 1975, arising from school expenditure decisions made by the Whitlam Government on the advice of the Karmel Committee (1973) report. This spike is captured in the IEA growth rates, but not the growth over the period spanned by LSAY.

With all three price indices, real spending rose over the period of the IEA tests. Using two of the three price indices, real spending rose over the period of the LSAY tests. The exception is when spending is indexed by the earnings of professional women. This is consistent with Leigh and Ryan (2006), who find that – relative to professionals – teachers' earnings declined from the mid-1970s to the mid-2000s. However, this index almost certainly understates the rise in real spending. In 2002-03, teacher salaries comprised 63 percent of government school

[15] To be precise, the share of teachers who were women was 67.4% in 2003, 65.9% in 1998, and 59.5% in 1975. This calculation includes both primary and secondary schools, and the government and non-government sector. We were unable to find the share in 1964, but the share of female teachers was 59.7% in 1970 and 55.0% in 1950. Linearly interpolating, this suggests that the share was 58.3% in 1964.

expenditure.[16] Even if one thought that teachers' salaries should rise at the same rate as professional women, then it would be more appropriate to take a weighted average of the All-Groups price index and the Earnings of Professional Women price index. This approach implies a slightly lower rise than the Schools Price Index. (Using weights of 37 percent for the All Groups price index and 63 percent for the Earnings of Professional Women price index, this approach implies a 5 percent rise in real education expenditure between 1975 and 1998, and a 171 percent rise over the period 1964 to 2003.)

Since we have more precise data on government expenditure than private expenditure, Panel B of Table 3 replicates the results using only government expenditure. Over the periods covered by the LSAY and IEA tests, government expenditure on schools has risen substantially, though not as rapidly as private expenditure. Again, if expenditure is indexed by the earnings of professional women, we observe a decline over the period 1975-98. However, a weighted average of the All Groups price index and the Earnings of Professional Women price index still implies an increase in government school expenditure over this period.

Our preferred estimate uses both private and government expenditure, indexed by the schools price index. This suggests that real per-child school expenditure increased by 10 percent over the period of the LSAY tests, and by 258 percent over the period of the IEA tests.

[16] We calculate this figure using data in MCEETYA (2003, Appendix 1: Statistical annex, Table 19). Since the user cost of capital is not included in our estimates, we omit it when calculating the salary share of Australian government school expenditure.

Table 3: Increase in Schools Expenditure (percent)
Preferred estimates in bold text

Year	Nominal Spending per Child	Indexed by All Groups Price Index	Indexed by Schools Price Index	Indexed by Earnings of Professional Women
Panel A: Private and Government Expenditure				
1997-98 / 1974-75	429%	18%	**10%**	-2%
2002-03 / 1963-64	4,147%	333%	**258%**	76%
Panel B: Government Expenditure Only				
1997-98 / 1974-75	410%	14%	6%	-5%
2002-03 / 1963-64	3,962%	314%	243%	69%

Where has increased educational expenditure gone? One significant factor pushing up costs over this period has been smaller class sizes. In Appendix Table 5, we show student-teacher ratios for 1964-2004. On average, class sizes fell by 20 percent in the period 1975-98, and by 43 percent in the period 1964-2003. These patterns can be directly observed in our data: in the 1964 numeracy test, the average class size was 36. In 2003, the average class size was 26.[17]

Assuming that salaries constitute 63 percent of all school spending, and that cutting class size requires a proportionate increase in funding, these class size cuts would have boosted school spending by 13 percent over 1975-98, and by 27 percent from 1964-2003. Lower student-

[17] Both the 1964 and 2003 IEA numeracy surveys contain data on average class size in the student's grade and school. While the bands are not directly comparable, they give some indication of the class size reduction over this period. In 1964, 7 percent of Australian students were in a class of 24 or fewer students, and 70 percent of students were in a classroom with 35 or more students. In 2003, 31 percent of students were in a class of 24 or fewer students, and 4 percent were in a class of 33 or more students. There is no systematic relationship in either survey between test scores and class size, though it is quite possible that nonrandom sorting might offset causal impacts in either direction.

teacher ratios could therefore account for almost all of the expenditure increase over the period 1975-98, though not over the period 1964-2003.

One factor that might have affected the expenditure growth trends that might not be expected to lead to any improvement in achievement was the decline from the 1960s in religious orders teaching in Catholic (and in some independent) schools. These teachers received a stipend that was less than the wage received by teachers in government schools at the time. As their numbers fell, they had to be replaced by lay teachers (or by themselves now teaching as non-religious staff) who were paid a similar wage to government school teachers. Unfortunately, data on the gap between the stipend paid to religious teachers and the salary paid to lay teachers is limited.[18] However, we do have data on the share of religious teachers in Catholic schools in various years.[19] Accounting for the share of students taught in Catholic schools, and assuming that wages constitute 63 percent of total expenditure, we can see how this factor would have affected total costs on the assumption that the stipend was 25 percent, 50 percent or 75 percent of a lay teacher's salary.[20] From 1964-2003, the increase in spending caused by the shift from religious to lay teachers would have led to a 7 percent spending increase if the stipend for religious teachers was 25 percent of lay salaries, 5 percent if the stipend was 50 percent of lay salaries, and 2 percent if the stipend was 75 percent of lay salaries. Over the period 1975-1998, the impact on total spending would have been 4 percent, 3 percent and 1 percent respectively. While this is clearly non-trivial, the movement away

---

[18] The only estimate we were able to obtain was from the South Australian Catholic Education Office, which informed us that their current religious stipend equates to 62% of the average salary paid to Australian school teachers. We are grateful to Geoff Hallion of Catholic Education SA for providing us with this information.

[19] Flynn and Mok (2002) provide data on the share of Catholic teachers from religious orders in 1965 (72.3%), 1971 (47.8%), 1993 (4.0%), and 2000 (1.6%). We linearly interpolate and extrapolate to obtain estimates for the years 1964, 1975, 1998 and 2003.

[20] The share of students in Catholic schools in our years of interest was 19.6 percent (1964), 17.0% (1975), 19.7% (1998) and 19.9% (2003). Our estimates account for changes in the 'market share' of Catholic schools.

from religious teachers in Catholic schools can only account for a small share of the full increase in spending over either period.

V. Discussion and Conclusion

Depending on which measure we employ, the test scores of young Australian teenagers appear to have fallen slightly from the mid-1960s to the early-2000s. Although our confidence intervals in some cases include zero, we can reject the hypothesis that there has been a statistically significant increase in test scores over the past four decades. Although we do not have as full a set of demographics as we would like, adjusting test scores to account for changes in the available demographic characteristics does not explain the decline; rather, the decline seems even larger. The results are economically significant, implying a drop of between one-fifth and one-tenth of a standard deviation. This suggests, for example, that the numeracy of the typical young teenage student in 2003 was approximately a quarter of a grade level behind his or her counterpart in 1964.

Turning to expenditure data, our preferred estimate is that real per-child school expenditure increased by 10 percent over the period of the LSAY tests (1975-98), and by 258 percent over the period of the IEA tests (1964-2003). Using different measures of expenditure, and different price indices, we obtain similar results.

One way of estimating the change in productivity of Australian schools is to divide the test score trends by the expenditure trends. In effect, this exercise estimates the necessary expenditure required for each point on the literacy and numeracy tests. We find that school productivity has declined by 12-13 percent between 1975 and 1998, and by 73 percent between 1964 and 2003.[21] This contrasts starkly with multifactor productivity across the economy, which rose by 34 percent in the period 1975-98, and by 64 percent from 1964-2003 (ABS 2006).

Although we have done our best to adjust our results given the data available to us, we cannot fully rule out explanations that have nothing to do with school productivity. For example, it is plausible that changing family structure, social norms, and entertainment media may have affected test scores. Alternatively, it might be the case that our measure of school outputs is unduly narrow, and fails to capture factors that do not appear on our tests – such as physical fitness, critical thinking skills, problem solving skills, communication skills or knowledge of science. Since we do not have data on these other outputs, it is not possible to know whether they have risen or fallen over time.

However, it is possible to identify explanations that relate directly to school productivity. First, part of the spending increase over the period in question resulted from class size cuts. If smaller classes have little or no impact on test scores (Hanushek 1998; Hoxby 2000, but cf Krueger 1999, 2003), then this policy change would have led to a reduction in school productivity. A second possibility is that falling teacher quality led to a disproportionate drop in student performance. If a 10 percent reduction in real teacher salaries reduces student performance by more than 10 percent, then falling teacher salaries could lower school productivity. A third productivity-related explanation is that shifts in the way that schools were managed led toa decline in school productivity. For example, Australian schools of

[21] An alternative (narrower) measure of the increase in school inputs would be the reduction in student-teacher ratios, which fell by 20 percent over the period 1975-98, and by 43 percent over the period 1964-2003 (Appendix Table 5). If this were regarded as the only input, it would imply a reduction of school productivity in the order of 19-20 percent between 1975 and 1998, and a reduction of 32 percent between 1964 and 2003.

education shifted towards a whole-language approach to teaching reading in the 1970s (van Kraayenoord and Paris 1994). To the extent that this was less effective than other methods of reading instruction, it might also have led to a decline in school productivity.[22] Although our approach does not allow us to distinguish between these explanations, we hope it will help encourage further analysis.

[22] Similarly, it is possible that the mainstreaming of students with disabilities reduced mean test scores. This might occur for two reasons. The first is if the positive peer effects for disabled students were lower than the negative peer effects for non-disabled students. We are not aware of any research that bears directly on this issue. The second is if a significant proportion of very low scoring students were added to the group sitting the tests. None of the individual distributions in Figures 3 and 5 show such an addition. Rather, the decline in performance occurred across the entire distributions.

References

Afrassa, T. and Keeves, J. 1999. “Changes in students’ mathematics achievement in Australian lower secondary schools over time” *International Education Journal,* 1(1):1-21

Atkinson, A.B. 2005. *Atkinson review: Final report - measurement of government output and productivity for the national accounts*. Palgrave MacMillan: London

Australian Bureau of Statistics (ABS). 2006. *Australian System of National Accounts, 2005-06*, Cat No 5204.0. ABS: Canberra

Australian Institute of Health and Welfare. 2007. *Juvenile justice in Australia 2004–05*. AIHW cat. no. JUV #2. AIHW: Canberra.

Boyle, R. 2006. “Measuring Public Sector Productivity: Lessons from International Experience”. Committee for Public Management Research Discussion Paper 35. Institute of Public Administration: Dublin

Brown, G., Micklewright, J., Schnepf, S.V. and Waldmann, R. 2005. “Cross-National Surveys of Learning Achievement: How Robust are the Findings?”. IZA Discussion Paper No. 1652. IZA: Bonn.

Carcach, C. 1997. “Youth as Victims and Offenders of Homicide”. Australian Institute of Criminology, Paper presented at the Australian Institute of Criminology Conference, Juvenile Crime and Juvenile Justice: Toward 2000 and Beyond, Adelaide, 26-27 June 1997.

Cardak, B.A. and Ryan, C. 2006. “Why are High Ability Individuals from Poor Backgrounds Under-represented at University?”, La Trobe University Economics Discussion Paper A06.04. La Trobe University: Melbourne.

Clements, K.W. and Izan, H.Y. 1985. “Australian consumer price data, 1951-1981”, Discussion Paper 85.01, Department of Economics, University of Western Australia.

Cook, M. D. and Evans, W.N. 2000. “Families or Schools? Explaining the Convergence in White and Black Academic Performance” *Journal of Labor Economics*, 18(4): 729-754.

Department for Education and Skills. 2005. “Measuring Government Education Output in the National Accounts”, UK Government: London

Douglas, J. 2006. “Measurement of Public Sector Output and Productivity”, Policy Perspectives Paper 06/09, New Zealand Treasury: Wellington

Flynn, J. R. 1987. “Massive IQ gains in 14 nations: What IQ tests really measure”. *Psychological Bulletin*, 101, 171-191.

Flynn, J. R. 1999. “Searching for justice: The discovery of IQ gains over time”. *American Psychologist*, 54, 5–20.

Flynn, J. R. 2006. “Efeito Flynn: Repensando a inteligência e seus efeitos [The Flynn Effect: Rethinking intelligence and what affects it]”. In C. Flores-Mendoza & R. Colom (Eds.), *Introdução à Psicologia das Diferenças Individuais* (pp. 387-411). [*Introduction to the psychology of individual differences*]. Porto Alegre, Brazil: ArtMed. (translation available on Flynn's website)

Flynn, M. and Mok, M. 2002. *Catholic Schools 2000: A Longitudinal Study of Year 12 Students in Catholic Schools 1972-1982-1990-1998*. Sydney: Catholic Education Office

Gundlach, E., Wossmann, L. and Gmelin, J. 1999. "The decline of schooling productivity in OECD countries"'. Kiel Institute of World Economics, Working Paper No. 926,

Gundlach, E., Wossmann, L. and Gmelin, J. 2001. "The Decline of Schooling Productivity in OECD Countries". *Economic Journal*, 111: C135-C147.

Hanushek, E. A. 1997. "The productivity collapse in schools". In (W. Fowler, Jr., ed.), *Developments in School Finance* 1996, pp. 183-95. Washington, DC: US Department of Education, National Center for Education Statistics

Hanushek, E. A. 1998. "The Evidence on Class Size", W. Allen Wallis Institute of Political Economy Occasional Paper Number 98-1, University of Rochester, Rochester, NY

Hanushek, E.A. and Kimko, D.D. 2000. "Schooling, Labor-Force Quality, and the Growth of Nations" *American Economic Review*, 90(5): 1184-1208.

Holzer, M. and Lee, S (eds). 2004. *Public Productivity Handbook*. 2nd Edition, CRC Press: Boca Raton, FL

Hoxby, C.M. 2000. "The effects of class size on student achievement: New evidence from population variation," *Quarterly Journal of Economics*, 115(4): 1239-1285.

Hoxby, C.M. 2001. "School Choice and School Productivity (or Could School Choice be a Tide that Lifts All Boats?" *mimeo*, Harvard University, Cambridge, MA.

Interim Committee, (1973), *Schools in Australia: Report of the Interim Committee for the Australian Schools Commission*, (Peter Karmel, Chair), Australian Government Publishing Service, Canberra.

Jamison, E.A, Jamison, D.T, and Hanushek, E.A. 2006. "The Effects of Education Quality on Income Growth and Mortality Decline". NBER Working Paper No. 12652. NBER: Cambridge, MA.

Krueger, A.B. 1999. "Experimental Evidence of Education Production Functions," *Quarterly Journal of Economics*, 114(2): 497-532

Krueger, A.B. 2003. "Economic Considerations and Class Size", *Economic Journal* 113(485): F34–F63.

Marks, G. and Ainley, J. 1997. "Reading Comprehension and Numeracy Among Junior Secondary School Students in Australia", Research Report 3. ACER: Melbourne

Marks, G. and Fleming, N. 1998a. "Factors Influencing Youth Unemployment in Australia 1980-1994", Research Report 7. ACER: Melbourne

Marks, G. and Fleming, N. 1998b. "Youth Earnings in Australia 1980-1994: A Comparison of Three Youth Cohorts", Research Report 8. ACER: Melbourne;

Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA). 2003. *2003 National Report on Schooling in Australia*. MCEETYA: Carlton South, Vic.

Mukherjee, S. 1997. "Juvenile Crime: Overview of Changing Pattern", Paper presented at the Australian Institute of Criminology Conference, Juvenile Crime and Juvenile Justice, Towards 2000 and Beyond, Adelaide, 26-27 June.

Nettlebeck, T. and Wilson, C. 2004. "The Flynn effect: Smarter not faster". *Intelligence*, 32, 85-93.

Oaxaca, R. 1973. "Male-female Wage Differentials in Urban Labor Markets." *International Economic Review* 14(3): 693-709.

Polk, K. 1995. Youth violence: myth and reality" in D. Chappell and S. Eggers (eds.) *Australian Violence: Contemporary Perspectives II*. Australian Institute of Criminology: Canberra

Rosen, E. 1993. *Improving Public Sector Productivity: Concepts and Practice,* Sage Publications: New York

Rothman, S. 2002. "Achievement in Literacy and Numeracy by Australian 14-Year-Olds, 1975-1998" Research Report Number 29, ACER: Melbourne.

Ryan, C. and Watson, L. 2004. "Year 12 completion and retention in Australia in the 1990s" *Australian Journal of Labour Economics*, 7(4), 481–500

SCRGSP (Steering Committee for the Review of Government Service Provision) 2007, *Report on Government Services 2007*, Productivity Commission: Canberra

Snooks, G.D. *Portrait of the family within the total economy. A study in longrun dynamics, Australia 1788-1990*, Cambridge University Press: Cambridge

van Kraayenoord, C.E. and Paris, S.G. 1994. "Literacy instruction in Australian primary schools". *The Reading Teacher*, 48(3): 218–228.

Wallace, A. 1986. *Homicide: The Social Reality*. NSW Bureau of Crime Statistics and Research: Sydney

Weale, M. 2007. "Following the Atkinson Review: the quality of public sector output" *Economic and Labour Market Review*. 1(7): 22-26

Wundersitz, J. 1993, "Some trends in officially recorded youth offending: a state-by state Comparison", *National Conference on Juvenile Justice, Conference Proceedings No. 22*, eds L. Atkinson & S.A. Gerull, Australian Institute of Criminology, Canberra, 53-66

Wundersitz, J. 2005, *Juvenile Justice in South Australia: A 2004 Update*. Information Bulletin 47. South Australian Office of Crime Statistics and Research: Adelaide

Appendix 1: Common Questions Across IEA Numeracy Tests

Below, we show some examples of the common questions used on the IEA tests. In comparing results, readers should bear in mind that the age and grade level differed across tests (see text for details).

(i) In the division 24.56/0.04, the correct answer is

a. 0.614   *b. 6.14*   *c. 61.4*   **d. 614*   *e. 6140*

Fraction correct - 1964: 39%; 1978: 38%; 1995: 23%.

(ii) Joe had three test scores of 78, 76, and 74, while Mary had scores of 72, 82, and 74. How did Joe's average compare with Mary's?

a. *Joe's was 1 point higher.*
b. *Joe's was 1 point lower.*
**c.* *Both averages were the same*
d. *Joe's was 2 points higher.*
e. *Joe's was 2 points lower.*

Fraction correct - 1964: 88%; 1978: 83%, 1995: 74%; 1999: 74%; 2003: 68%

(iii) If P=LW and if P=12 and L=3, then W is equal to

a.3/4   *b.3*   **c.4*   *d.12*   *e.36*

Fraction correct - 1964: 68%; 1978: 63%; 1995: 66%.

(iv) If $x/2 < 7$, then

a. $x<7/2$ *b.* $x<5$ **c.* $x<14$ *d.* $x>5$ *e.* $x>14$

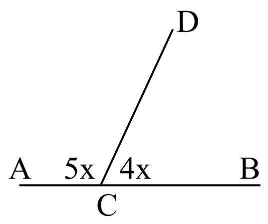
Fraction correct - 1964: 37%; 1978: 38%; 1995: 41%.

(v) Which of the following is false when a and b are different real numbers:

a. $(a+b)+c = a+(b+c)$
b. $ab=ba$
c. $a+b=b+a$
d. $(ab)c=a(bc)$
e. $a-b=b-a$ *

Fraction correct - 1964: 41%; 1978: 46%; 1995: 33%.

(vi) If AB is a straight line, what is the measure in degrees of the angle BCD in the figure?



a. 20 *b. 60* *c. 50* **d. 80* *e.100*

Fraction correct - 1964: 62%; 1978: 60%; 1995: 75%.

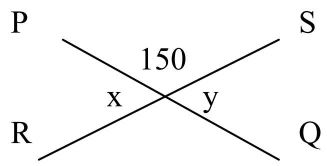
(vii) If x = -3, the value of -3x is
a. -9 *b. -6* *c. -1* *d. 1* **e. 9*

Fraction correct - 1964: 29%; 1978: 37%, 1995: 38%; 1999: 52%.

(viii) A runner runs 3,000 metres in exactly 8 minutes. What was his average speed, in metres per second?
a. 3.75. **b. 6.25* *c. 16.0.* *d. 37.5.* *e.62.5*

Fraction correct - 1964: 47%; 1978: 28%, 1995: 28%; 1999: 30%.

(ix) If, in the given figure, PQ and RS are intersecting straight lines, then x+y is equal to
a. 15. *b. 30.* **c. 60.* *d.180* *e.300*



Fraction correct - 1964: 41%; 1978: 32%, 1995: 51%; 1999: 55%; 2003: 50%

Appendix 2: Expenditure on Primary and Secondary Education

Government Expenditure

Our total expenditure series is created as follows:

- 1963-64 to 1979-80: We use data on expenditure on education published by Commonwealth Bureau of Census and Education (the publication that subsequently became *Expenditure on Education, Australia*, ABS Cat No 5510.0). Our series sums expenditure by state and federal government on primary and secondary schools. For 1963-64 to 1967-68, federal expenditure on schools is not separately tabulated, so we use total expenditure on education.
- 1980-81 to 1997-98: We use data from *Expenditure on Education, Australia*, ABS Cat No 5510.0.
- 1998-99 to 2003-04: We use data on operating expenses on primary and secondary education, from *Government Finance Statistics, Education*, ABS Cat No 5518.0.

In recent years, school expenditure series have also included a notional user cost of capital. Since that information is not available for earlier years, we do not include it in our series.

Student enrolment data are obtained from the *Australian Yearbooks* (various years).

Private Expenditure

We are not aware of any consistent series of private expenditure on schools. Our calculations are therefore based on data from the *Household Expenditure Survey* (HES). In the years 1984, 1988-89, 1993-94, and 1998-99, the HES contains comparable data on school fees, tabulated by primary/secondary and government/non-government schools (figures from the 1974 and 2003-04 HESs were not directly comparable).

Using data on the number of students enrolled in each of these categories (from the *Australian Yearbooks*), and the number of households in Australia (from *Australian Demographic Statistics*, ABS Cat No 3101.0), we estimate the average annual expenditure per student in the following categories: government primary, government secondary, non-government primary, non-government secondary. We then calculate the ratio of fees to government expenditure in each of these years. In three of the HESs (1984, 1988-89 and 1998-99), the ratios are very similar. However, in the 1993-94 HES, the ratio of private school fees to government expenditure differs significantly from the other surveys, so we opt not to use this year. The ratio of private spending to government spending is listed in the Appendix Table 1.

Appendix Table 1: Ratio of Private Expenditure to Government Expenditure

HES Year	Govt Primary	Govt Secondary	Non-Govt Primary	Non-Govt Secondary	Used for years
1984	1.1%	3.3%	21.8%	62.9%	1963-64 to 1987-88
1988-89	1.4%	3.8%	20.6%	62.9%	1988-89 to 1997-98
1993-94	1.4%	5.2%	20.1%	76.1%	Not used
1998-99	1.4%	4.2%	25.8%	62.7%	1998-99 onward

We estimate per-student private expenditure for all years by combining the estimates in Appendix Table 1 with data on the share of students in each of the four types of schools (government primary, government secondary, non-government primary, non-government secondary), and with government expenditure per student.

Our nominal expenditure series are presented in Appendix Table 2.

Price Indices

We use three price indices to adjust the expenditure series.

1. The All Groups CPI is simply obtained from *Consumer Price Index, Australia*, ABS Cat No 6401.0.

2. We construct a Schools CPI, by splicing together the following series:
   - 1963-64 to 1965-66: Miscellaneous Groups CPI, from ABS CPI, Cat No 6401.0
   - 1966-67 to 1977-78: Other Services CPI, from Clements and Izan (1985)
   - 1978-79 to 1992-93: Schools Prices Index, from the Statistical Annex to the *National Report on Schooling in Australia* (various years) prepared by the Department of Education, Science and Training.
   - 1993-94 to 2003-04: Education CPI, from ABS CPI, Cat No 6401.0

3. We construct a price index based on the earnings of female professionals. This is based on the assumption that the main expenditure item for schools are teacher salaries, and that the education sector must compete against other professional occupations. Our index splices together the following series:
   - 1963-64 to 1974-75: Based upon an index of women's award wages, constructed by Snooks (1994).
   - 1975-76 to 2003-04: Based upon the average weekly ordinary time earnings for full-time female professionals, from the *Employee Earnings and Hours* survey.

   Note that the index is very similar if we use the earnings of all female workers instead of professionals.

Where price indices were reported quarterly, we calculate an annual average.

While the Australian financial year runs from July to June, some of our series are only available on a calendar year basis. In these instances, we match to the second half of the financial year (eg. *Employee Earnings and Hours* data for 1976 are used to construct the female professionals' earnings index in financial year 1975-76).

The three price indices are presented in Appendix Table 3. The three real expenditure series are presented in Appendix Table 4.

Student-Teacher Ratios

Student-teacher ratios are estimated from the following sources:

- 1964-75: We combine data on the number of teachers and students, sourced from the *Australian Yearbook*

- 1976-2001: Sourced from *Education and Training Indicators, Australia, 2002*, ABS Cat No 4230.0, Data Cubes Topic 11.
- 2002-2004: Sourced from *Schools, Australia*, ABS Cat No 4221.0, Table 54.

Student-teacher ratios are presented in Appendix Table 5.

Appendix Table 2: Nominal Education Expenditure
(Spending per child per year, in current dollars)

Year	Government Spending	Private Spending	Total Spending
1963-1964	$165	$18	$183
1964-1965	$178	$19	$197
1965-1966	$188	$20	$208
1966-1967	$201	$21	$222
1967-1968	$221	$23	$244
1968-1969	$247	$30	$277
1969-1970	$281	$28	$309
1970-1971	$327	$32	$359
1971-1972	$382	$37	$419
1972-1973	$444	$44	$488
1973-1974	$573	$57	$630
1974-1975	$899	$89	$987
1975-1976	$1,074	$106	$1,180
1976-1977	$1,200	$118	$1,318
1977-1978	$1,367	$135	$1,502
1978-1979	$1,449	$146	$1,595
1979-1980	$1,570	$162	$1,732
1980-1981	$1,525	$162	$1,687
1981-1982	$1,713	$189	$1,901
1982-1983	$2,000	$228	$2,228
1983-1984	$2,193	$259	$2,452
1984-1985	$2,442	$298	$2,740
1985-1986	$2,597	$325	$2,922
1986-1987	$2,739	$349	$3,088
1987-1988	$2,830	$365	$3,195
1988-1989	$3,135	$410	$3,545
1989-1990	$3,261	$428	$3,688
1990-1991	$3,506	$459	$3,965
1991-1992	$3,740	$489	$4,229
1992-1993	$3,831	$504	$4,335
1993-1994	$3,898	$518	$4,416
1994-1995	$3,983	$536	$4,520
1995-1996	$4,188	$570	$4,759
1996-1997	$4,372	$603	$4,975
1997-1998	$4,585	$639	$5,224
1998-1999	$5,350	$802	$6,152
1999-2000	$5,644	$855	$6,498
2000-2001	$5,967	$915	$6,882
2001-2002	$6,305	$978	$7,284
2002-2003	$6,720	$1,058	$7,778
2003-2004	$7,169	$1,129	$8,297

Appendix Table 3: Price Indices (2003-04=100)

Year	All Groups Price Index	Schools Price Index	Earnings of Professional Women Price Index
1963-1964	10.0	7.9	4.0
1964-1965	10.3	8.1	4.1
1965-1966	10.7	8.5	4.3
1966-1967	10.9	8.9	4.6
1967-1968	11.4	9.1	4.9
1968-1969	11.7	9.5	5.3
1969-1970	12.0	10.1	5.6
1970-1971	12.5	10.5	6.6
1971-1972	13.4	11.3	7.3
1972-1973	14.2	12.8	9.2
1973-1974	16.1	13.9	12.9
1974-1975	18.7	15.4	15.3
1975-1976	21.2	19.7	17.7
1976-1977	24.1	21.4	20.0
1977-1978	26.4	21.6	21.6
1978-1979	28.6	23.0	23.1
1979-1980	31.5	24.5	25.7
1980-1981	34.4	27.1	28.8
1981-1982	38.0	30.3	32.4
1982-1983	42.4	35.1	36.1
1983-1984	45.3	37.0	38.6
1984-1985	47.2	39.2	41.1
1985-1986	51.2	40.8	47.2
1986-1987	56.0	43.1	48.3
1987-1988	60.1	44.7	51.6
1988-1989	64.5	47.2	55.1
1989-1990	69.7	50.2	57.8
1990-1991	73.4	53.6	61.5
1991-1992	74.8	56.6	65.3
1992-1993	75.5	57.9	67.4
1993-1994	76.9	60.2	70.8
1994-1995	79.4	62.7	73.4
1995-1996	82.7	65.8	75.0
1996-1997	83.8	69.9	78.8
1997-1998	83.8	74.2	82.6
1998-1999	84.9	77.9	84.8
1999-2000	86.9	81.7	87.0
2000-2001	92.1	85.7	89.9
2001-2002	94.8	89.6	92.8
2002-2003	97.7	94.0	96.4
2003-2004	100.0	100.0	100.0

Appendix Table 4: Real Education Expenditure
(Spending per child per year, in 2003-04 dollars)

Year	Indexed by All Groups Price Index	Indexed by Schools Price Index	Indexed by Earnings of Professional Women Price Index
1963-1964	$1,838	$2,308	$4,575
1964-1965	$1,909	$2,430	$4,792
1965-1966	$1,947	$2,446	$4,789
1966-1967	$2,031	$2,510	$4,829
1967-1968	$2,150	$2,679	$4,962
1968-1969	$2,366	$2,907	$5,204
1969-1970	$2,577	$3,068	$5,513
1970-1971	$2,864	$3,437	$5,405
1971-1972	$3,118	$3,696	$5,705
1972-1973	$3,433	$3,804	$5,316
1973-1974	$3,915	$4,538	$4,870
1974-1975	$5,267	$6,429	$6,439
1975-1976	$5,568	$5,992	$6,672
1976-1977	$5,467	$6,162	$6,601
1977-1978	$5,686	$6,963	$6,944
1978-1979	$5,584	$6,930	$6,911
1979-1980	$5,498	$7,058	$6,729
1980-1981	$4,900	$6,215	$5,866
1981-1982	$4,997	$6,271	$5,861
1982-1983	$5,249	$6,348	$6,167
1983-1984	$5,413	$6,629	$6,352
1984-1985	$5,800	$6,987	$6,669
1985-1986	$5,705	$7,158	$6,191
1986-1987	$5,512	$7,168	$6,400
1987-1988	$5,313	$7,143	$6,189
1988-1989	$5,494	$7,509	$6,435
1989-1990	$5,293	$7,346	$6,385
1990-1991	$5,404	$7,402	$6,446
1991-1992	$5,656	$7,467	$6,476
1992-1993	$5,738	$7,480	$6,434
1993-1994	$5,740	$7,332	$6,234
1994-1995	$5,694	$7,214	$6,159
1995-1996	$5,753	$7,228	$6,348
1996-1997	$5,935	$7,122	$6,316
1997-1998	$6,231	$7,044	$6,326
1998-1999	$7,248	$7,895	$7,257
1999-2000	$7,478	$7,955	$7,472
2000-2001	$7,470	$8,029	$7,658
2001-2002	$7,686	$8,133	$7,851
2002-2003	$7,961	$8,271	$8,070
2003-2004	$8,297	$8,297	$8,297

Appendix Table 5: Student-Teacher Ratio

Year	Student-Teacher Ratio	Year	Student-Teacher Ratio
1964	25.4	1985	15.2
1965	25.0	1986	15.3
1966	25.6	1987	15.1
1967	25.1	1988	15.1
1968	24.3	1989	15.3
1969	23.8	1990	15.3
1970	23.1	1991	15.4
1971	22.6	1992	15.3
1972	21.8	1993	15.3
1973	21.0	1994	15.5
1974	20.2	1995	15.4
1975	19.2	1996	15.4
1976	18.6	1997	15.3
1977	18.0	1998	15.3
1978	17.6	1999	15.0
1979	17.0	2000	14.9
1980	16.9	2001	14.7
1981	16.8	2002	14.7
1982	16.5	2003	14.5
1983	16.1	2004	14.3
1984	15.6		

Proportionate change in class size from:	
1975 to 1998	-20%
1964 to 2003	-43%
Impact of class size reduction on spending (assuming salaries constitute 63% of total school expenditure)	
1975 to 1998	+13%
1964 to 2003	+27%