Chapter 7: South Australian teacher judgement assessments: 1997 and 1998

> …profiles function as a framework for assessment and reporting and do not in themselves constitute an assessment method. What they do allow, however, is for teachers, schools and school systems to communicate about student progress and achievement using a language and standards which are consistent across classrooms, schools and school systems.
>
> Hill, 1994, p. 38.

This chapter summarises the key findings from the South Australian teacher judgement assessments using the Statements and Profiles for Australian Schools (SPFAS) approach. The history and detail of the approach are described in Chapter 3. The data collection processes and the assessment role of teachers are summarised here briefly.

Teacher judgement assessments are addressed in their own right without reference to checks against alternative assessments. This is done to appreciate the capacity of teachers to make assessments of individual students on an eclectic loosely specified basis but against general criteria (SPFAS) provided as a map for the development of learning in each strand considered. An overview summary of the assessment results is presented in the adopted metric of those assessments: profile or level units.

The overview then leads to the important question, "How well do the South Australian teacher assessments match test assessments?" To consider that question adequately, the assessment scales need to be converted to a metric common with that of the tests. Processes for converting teacher assessments to the test metric and then comparing them with each other are addressed in the next chapter.

The data collection revisited

Detail of the data collection process is covered in Chapter 3. In brief, the data collections of 1997 and 1998 were identical in most respects. While in 1997 four learning areas were included, only the English learning area data are used in this study. The teacher judgement assessment survey was conducted in October; this timing has a small impact to be considered when test and teacher assessments are compared, as the tests had been conducted in August, two months earlier. In 1998 the remaining four learning areas were included. Of these, the Mathematics learning area is the focus for this study. The teacher judgement assessment survey was conducted in the same month as the tests (for Years 3 and 5) making the timing of the assessments of no further concern in the comparison of data.

Survey software was used to manage the random selection of students and the learning area to be reported for each student. As a result teachers did not know in advance which students they would be required to report on, nor on which learning area. Each teacher reported for five randomly selected students from that class, by indicating which of the eight described levels for a strand in the learning area had most recently been achieved by each student. The teacher then reported on each student's progress towards achieving the criteria for the next level by clicking on a continuous line. This line was segmented but this was not indicated to the teachers. A click on the line activated one of nine segments, leading to an indication of progress in 0.1 segments. Data files for the collections included about 120,000 records in 1997 and over 200,000 in 1998. Each record was one student/strand/rating event. The unique identifier for the student was required to manage the screen presentation to the teacher and was preserved in the collection process to add other identifiers to the file (gender, socio-economic status of the family, language background, and indigenous status) found from the general statistical records of the education department.

Rothman (1998, 1999) analysed strands within learning areas as separate summaries for each strand. In the current analysis records were restructured so that individual strand assessments (for English and Mathematics) were consolidated for each student. The purpose of the record restructure was to allow a mean assessment in the learning area to be made for each student, consistent with the general principle of test design where composite strands are combined in the test design and a general overall score for each student calculated. In the cases of a test the strand equivalent data (e.g., reading and writing in English) can be analysed separately and individual item performance for each student investigated. In the teacher judgement assessment data drilling down below the strand is not possible.

Once the files were structured as consolidated student records it was possible to attach the gender and date of birth to calculate the ages at assessment for each student. For 1997 the restructuring of records resulted in 7871 student cases over 8 Year levels, approximately 1000 cases per Year level, and about 100 cases per age categorised at 0.1 of age (that is about 100 cases for each month of age). On the basis of 5 students assessed per teacher, the data represent the assessments of 1500 teachers. For 1998, 12050 student cases over 8 Year levels were assembled, approximately 1500 cases per Year level, and about 150 cases per age categorised at 0.1 of age. On the basis of 5 students assessed per teacher, the data represent the assessments of 2400 teachers.

The data collection has a number of inherent independent replications; collection period (1997, 1998), different learning areas (Literacy and Numeracy), eight Year levels for each collection, primary versus secondary teachers, and includes 20000 students and almost 4000 teachers overall.

The English Learning Area

The general statistical characteristics of the data for English are listed in Table 7.1. The average ages at assessment differ consistently by one year of increasing age for each increase in Year level. The average age is greater by 0.2 of a year of age than for the Mathematics data reported in Table 7.2 and the data and models developed in Chapter 6, due to the two-month difference in assessment periods. The data are averages of two judgements (Reading and Writing) rather than for each strand separately.

Spread of assessments and scale use

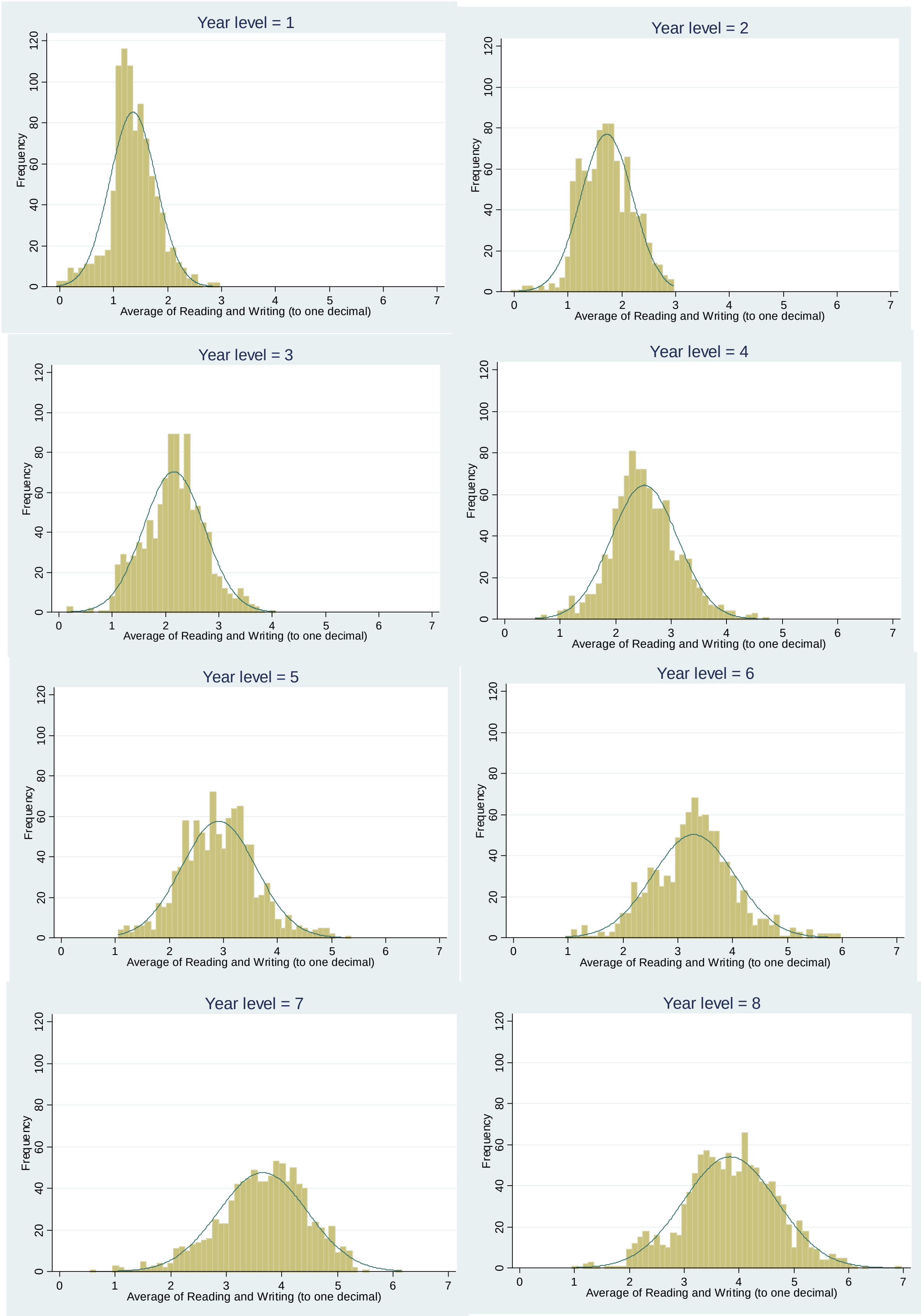
Issues arise from the response format and the history of the design of the teacher assessment collection. Did teachers use the full range of the scale, based on the developmental range of the students for a given Year level? What form does the distribution of the assessment results take?

A view of the spread of the teacher assessments along the assessment dimension, and the likely use of the full spectrum of response possibilities, is obtained in the panels in Figure 7.1. The histograms indicate that the full spectrum of responses appears to have been used by teachers in their assessments. In general terms, the averaged English assessments, based on equal weighting of the Reading and Writing, are spread around the mean and fit the shape, for most assessment values, of the superimposed normal curve. There are exceptions.

For Year 1, scale positions just above 1 are very well used. These are points that indicate that the student has met the criteria for level 1 but has not progressed much further. While these points appear over represented and as a consequence some other points under represented (just below 1 as examples), the panel shows that the full range of assessment points are used. Similar over and under representation are shown in other panels.

For Year 2, a point just past level 2 stands out due to points missing either side, although the segment itself sits close to the super-imposed curve. For Year 3 the early points on the scale from 2 to 3 are over represented. A similar effect is observed for Year 4. At Year 5 the effect has moved to the beginning of level 3 and remains in this segment of the scale for Year 6. For Years 7 and 8 the first segment from 4.0 to 4.5 is over represented relative to a normal distribution. The effect is also obvious for the beginning of 3.0 to 3.5 for Year 8. Year 8 is the first year of secondary school and thus reflects the assessments of secondary teachers as against those of primary school teachers. The data confirm that teachers used the full range of points on the (hidden) underlying 10-point scale (unwittingly since they responded to a line rather than assigning numbers) and did so with a preponderance towards the early segments of each new level.

Figure 7.1 English 1997 – Histograms of score distributions by Year level



Learning status trends with Year level

The means and medians, as shown in Table 7.1 are close, usually differing in the second decimal place only, indicating that the cases are approximately evenly distributed around the mean for each Year level. This is further illustrated in Figures 7.2 and 7.3 where the relationship of each English strand and the combined Reading and Writing strands with Year level/age is shown to be linear with Year level up to Year 7. Year 8 is an exception. This contrasts with the general shape of the mean test scores with Year level/age described in Chapters 5 and 6 where the means of IRT based measures show decelerating growth with Year level/age: IRT measures are not linear with Year level or age.

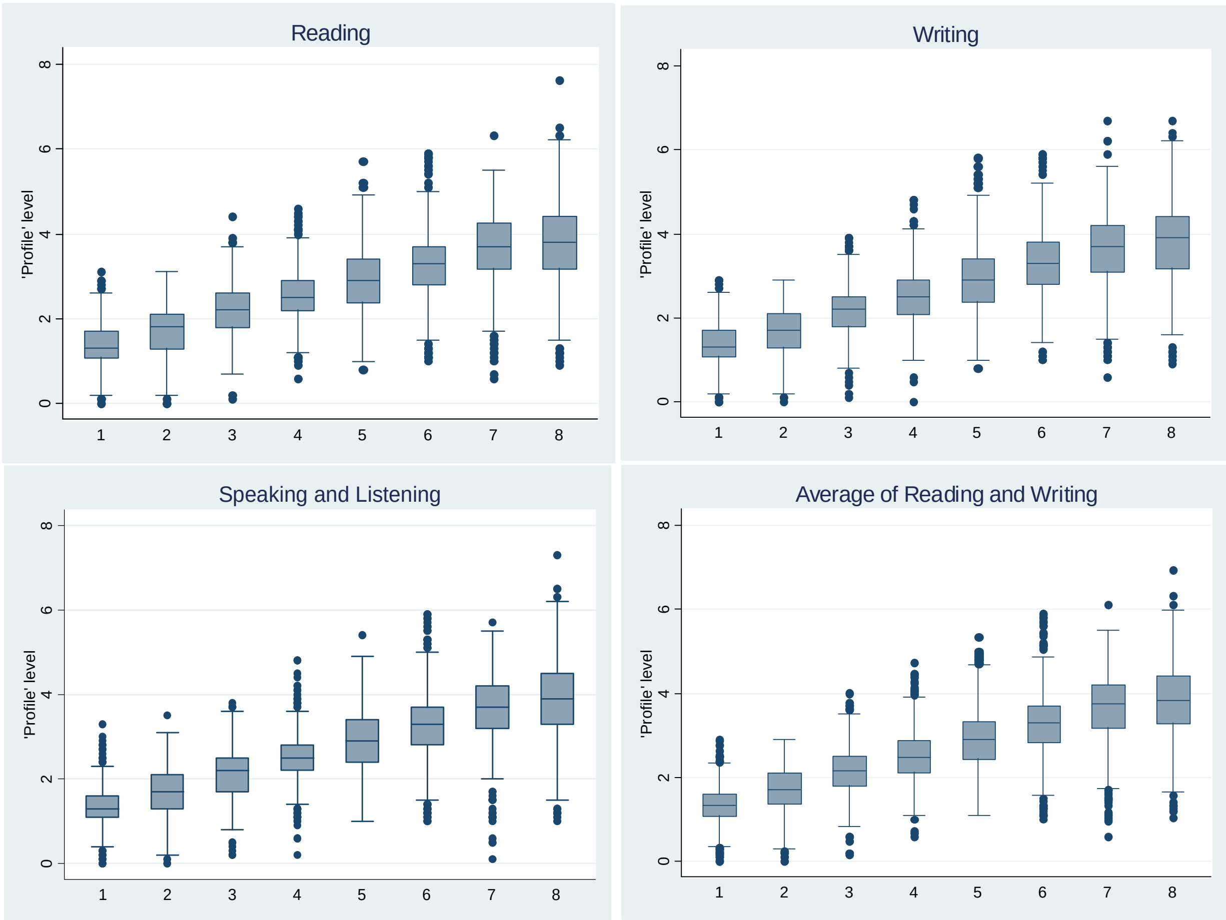
Table 7.1 English Learning Area by Year level–1997: General Statistics

Year level	Average age at assessment (October 97)	Average of Reading and Writing values (in Profile level units) Mean	Median	SD	IQR	SE of mean	Skew-ness	Kurtosis	N
1	6.81	1.36	1.33	0.43	0.50	0.01	-0.04	4.05	923
2	7.79	1.73	1.70	0.48	0.73	0.02	-0.04	3.06	926
3	8.79	2.16	2.17	0.57	0.70	0.02	-0.03	3.23	1005
4	9.78	2.52	2.47	0.60	0.73	0.02	0.25	3.61	969
5	10.79	2.91	2.90	0.69	0.90	0.02	0.22	3.37	996
6	11.78	3.29	3.30	0.75	0.87	0.02	0.15	3.88	945
7	12.81	3.66	3.73	0.80	1.00	0.03	-0.49	3.39	956
8	13.79	3.84	3.83	0.85	1.10	0.03	-0.10	3.17	1151
All	**10.39**	**2.72**	**2.63**	**1.08**	**1.60**	**0.01**	**0.28**	**2.52**	**7871**

The standard deviation (SD) and the inter-quartile range (IQR) increase with Year level and age. This phenomenon is consistent with the Rowe and Hill (1996) observations for teacher judgements assessments in Victoria and consistent with the general linear relationship with grade, often indicative of a grade equivalent rescaling (Schulz & Nicewander, 1997). This key observation will be discussed in more detail following the description of the mathematics teacher judgement assessments. It offers a possible understanding for how teacher judgment assessments are made and why their distributions differ from those of IRT test assessments.

Figure 7.2 shows that the linear relationship of the medians for each Year level is consistent across strands.

Figure 7.2 Teacher Judgement assessments - English Learning Area 1997 by strand and Year level



The relationship of the mean teacher assessment with Year level is linear up to Year 7 as shown in Figure 7.3. The median scores vary around the linear trajectory of the means. Both the SDs and IQRs are shown to increase with Year level. This is illustrated again in Figure 7.4 where data points are plotted in 0.1 of a year of age, rather than at the average age of the Year level group. The points representing the mean of all the students in all the age categories of 0.1 of an age follow a linear trajectory with age, on average, with only a few points deviating from the general trajectory. A linear regression of the mean assessment scores with age up to Year 7 has a gradient of 0.374 profile level units per year of age. Using the mean of each age grouping eliminates the variance within age. The line of best fit (up to age 13.5) has a very high R^2 (above 0.99) suggesting a very good fit of the line to the means up to Year 7. Year 8 data indicate that secondary teachers report students at a point lower than the previous primary annual improvement would predict.

Figure 7.3 Teacher Judgement assessments - English Learning Area 1997: means, medians, standard deviations and inter-quartile ranges, by Year level

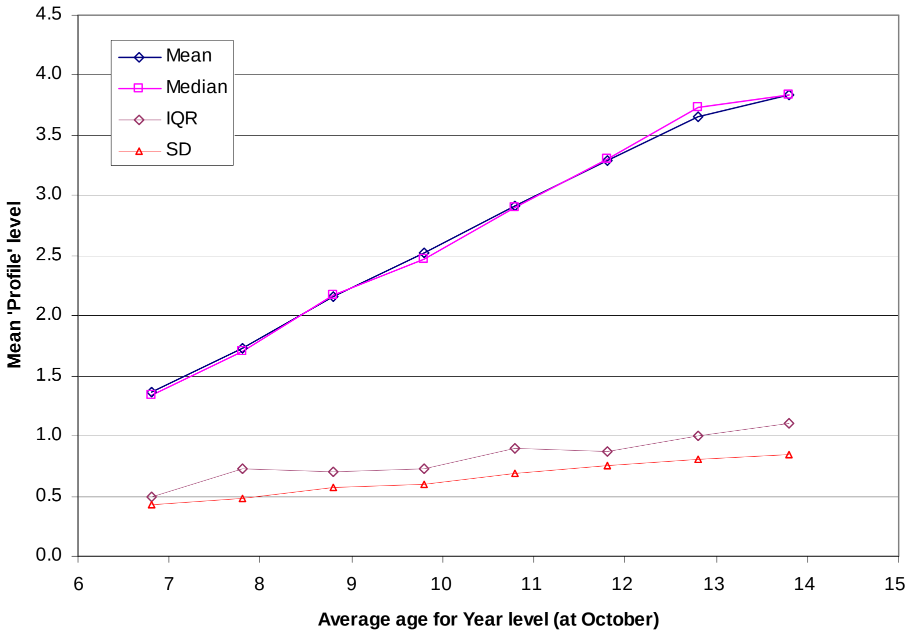
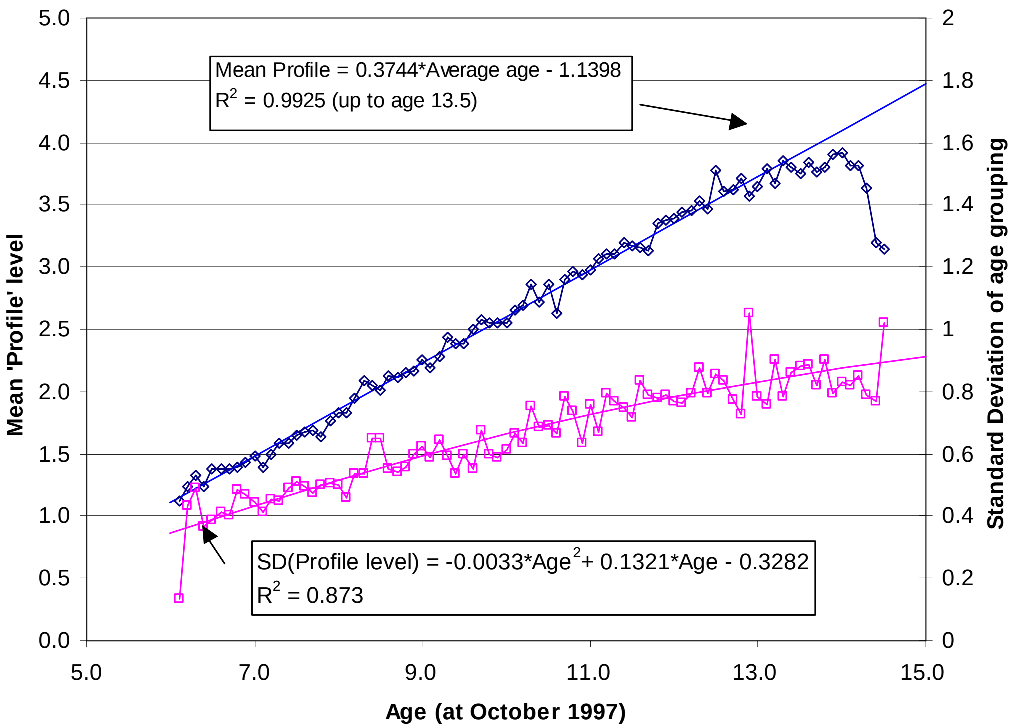


Figure 7.4 Teacher Judgement assessments - English Learning Area 1997: Mean profile level of Reading and Writing strands combined, by age



The increase in the SD with age is essentially linear, with slight levelling out from age 12. This is reflected by a quadratic curve fitting the data points slightly better than a straight-line function. As referenced earlier the phenomenon of linearly increasing means with Year level and age and increasing SDs is consistent with findings for Grade equivalent assessments.

Effect of age at assessment on the relationship of learning with Year level/age

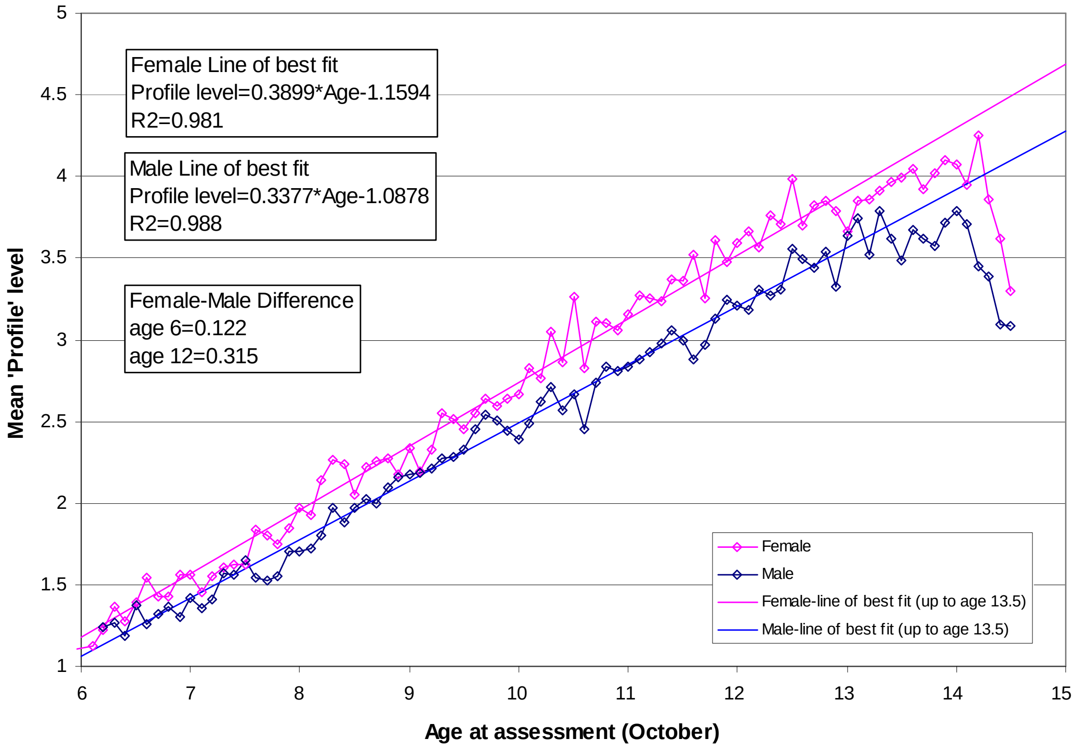
The age effect is shown to apply consistently across age categories of 0.1 of a year. A consistent linear gradient of 0.374 of a profile level unit per year of age from teacher judgement assessments applies up to age 13. The effect of a two to three month difference in the age (time) of assessment can be explored by the use of the regression expression found to fit the age data points in Figure 7.4. The gradient of 0.374 profile level units per year of age is also the growth in the means in successive Year levels assessed at the same time point, since the points are one year apart. A simple estimate of the mean difference between the August assessment date (test) and the October assessment date (teacher judgement) can be based on an assumption of about 2.5 months time difference, equivalent to 0.2 of a year. Based on the linear growth with age expression, a 0.2 difference of a year in age at assessment leads to a 0.07 profile level unit difference. This difference is approaching the pre-set resolution of the teacher judgement scale of 0.1 profile level units.

This analysis suggests that an age adjustment should be considered before the test and teacher assessments are directly compared. For a single comparison as applies in this thesis the age/time of assessment adjustment is less necessary, but not applying it will produce a relationship of test and teacher assessments that is slightly displaced. The issue is considered again in Chapter 8.

Gender differences in the English Learning Area

Figure 7.5 illustrates the difference between teacher judgement assessed learning trajectories by gender. Consistent with assessment summaries using test data, female students have assessments consistently higher than do males for any given age.

Figure 7.5 Teacher Judgement assessments - English Learning Area 1997: Mean profile level of Reading and Writing strands combined, by gender of students

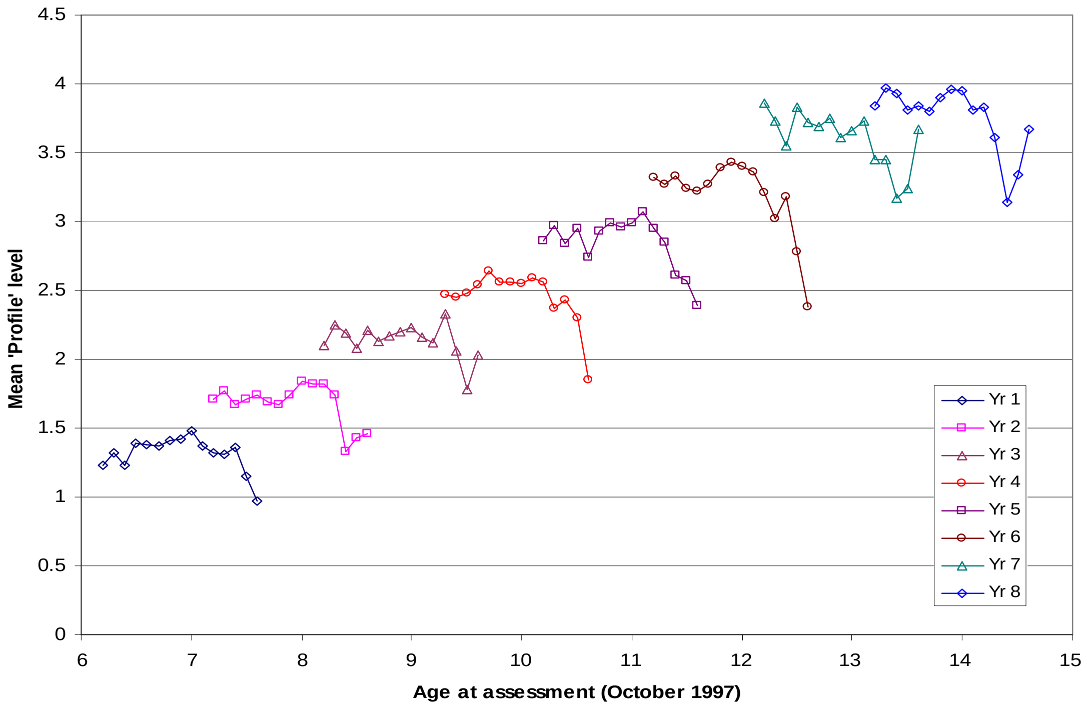


The gender difference based on the teacher assessments appears to increase with age and is reflected in the slopes of the regression lines fitted to the linear segment of the growth from age 6 to age 13.5. The trajectory is clearly different from Year 7 to Year 8 (age 12 to age 13) and is therefore truncated for the line of best fit up to age 13.5, only part way into the data points that represent Year 8. Based on regression lines, the difference at age 6 is 0.122 profile level units. By age 12 the difference is 0.315 units. The general pattern is consistent with that shown in Figure 6.5, where the difference between mean female and male test scores appears to increase with age.

Within Year level trends by age

Figure 7.6 illustrates the trends by age within Year level. At the gross scale of the profile level unit it is difficult to observe how well the trend in test scores with age, shown in Chapters 5 and 6, is reflected in the teacher data. The tails for students beyond the normal age range for the Year level are similar to the pattern for Literacy in Chapter 6. A more detailed comparison is addressed in Chapter 8, once the teacher and test data are brought to an approximately common scale.

Figure 7.6 Teacher Judgement assessments - English Learning Area 1997: Mean profile level of Reading and Writing strands combined, by age within Year level



Since the data were collected in October, the normal range for a Year level is from x.2 years to x+1.2 yrs, where x is the age appropriate to the Year level. The first point for each Year level commences at x.2. The small numbers of cases below this age have been censored for each Year level. For the lower Year levels there appears to be a slight gradient with age until the over-age cases are reached. The pattern is less regular than for the test data and model. There are fewer cases for each age point in the teacher data (i.e. lower *n* relative to the test data) thus potentially larger variation from the general trend pattern at each age point. The within-Year level age effect seems to disappear by Year 7.

The major characteristics of the teacher judgement assessment data for English have similarities with those for Mathematics, based on teacher judgment assessments made one year later.

The Mathematics Learning Area

The general statistical characteristics of the mathematics data are listed in Table 7.1. The mean ages at assessment increases consistently by one year of age for each increase in Year level. The mean age at each Year level is lower by 0.2 of a year of age than the English data reported in Table 7.1 as the assessments all occurred in August 1998. The means for each Year level are consistent with the test data and models developed in Chapter 6.

Spread of assessments and scale use

As for English the first aspect of the data is the distribution of the assessments and the extent to which teachers used the full range of assessment points available to them. The spreads of teacher judgement assessments along the assessment dimension (Figure 7.7) are similar to those for English, and indicate the use of the full spectrum of response possibilities. The data are the means of five judgements per student in each of the mathematics strands rather than for each strand separately (see Figure 7.8 for the individual strands). The histograms indicate that the full spectrum of response possibilities appears to have been used by teachers in their assessments. The mathematics assessments are spread around the mean and fit the shape of the superimposed normal curve for most assessment scale values. As for English there are exceptions. At Year 1 the scale positions just above 1 are very well used. These points indicate that the student has met the criteria for level 1 but has not progressed much further. While these points are over represented, from a normal distribution perspective, some other points are under represented (e.g., just below 1), the panel shows that the full range of assessment points are used. Similar over and under representations are shown in other panels.

On the assumption that the assessments should be normally distributed it would seem that teachers may under report students who have yet to reach the criteria for level 1 (Years 1, 2, 3), level 2 (Years 4, 5, 6), level 3 (Years 6 and 7) or level 4 (Year 8). The distributions for each year level appear to have regions of missing values to the left of the profile boundaries.

The SDs and IQRs increase with increasing year level, as for English, and are shown in Table 7.2.

Learning status trends with Year level

The trends with Year level for each strand of mathematic are shown individually in Figure 7.8. All show the same general pattern of linear growth with Year level up to Year 7 and then less growth in Year 8. The data for each student for each strand are combined into a grand average, shown in the lowest right panel. This box-plot averages the variations of the assessment in each strand for each student to an average assessment value, similar to the process that applies in the Numeracy total test score. The average of all strands is plotted by Year level (placed at average age for the Year level) in Figure 7.9 and tabulated in Table 7.2. The trajectory is linear up to Year 7. The medians and means are very close indicating that the distributions are reasonably well centred on the mean.

Figure 7.7 Mathematics 1998 – Histograms of score distributions by Year level

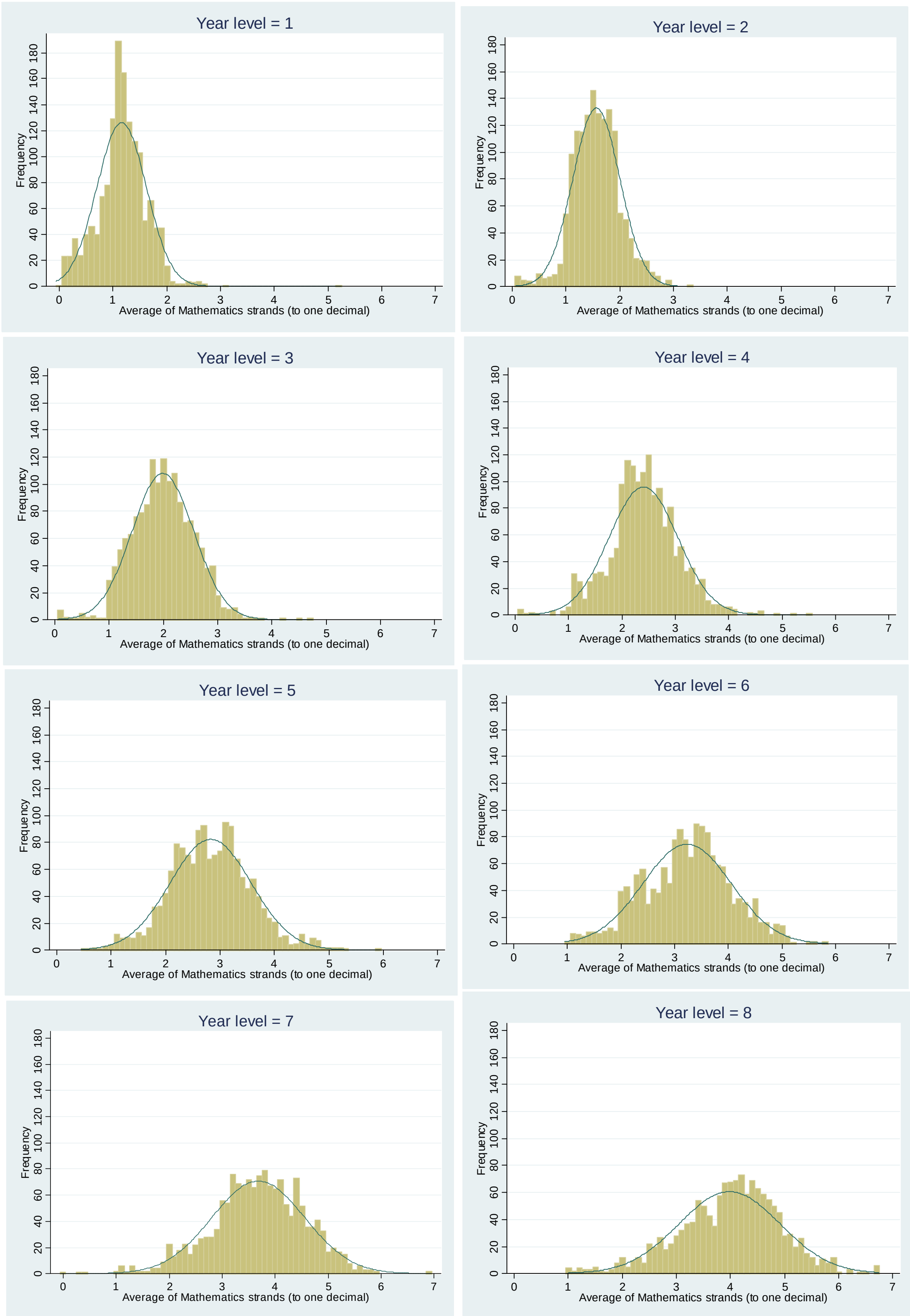


Figure 7.8 Teacher Judgement assessments- Mathematics Learning Area 1998 by strand and Year level

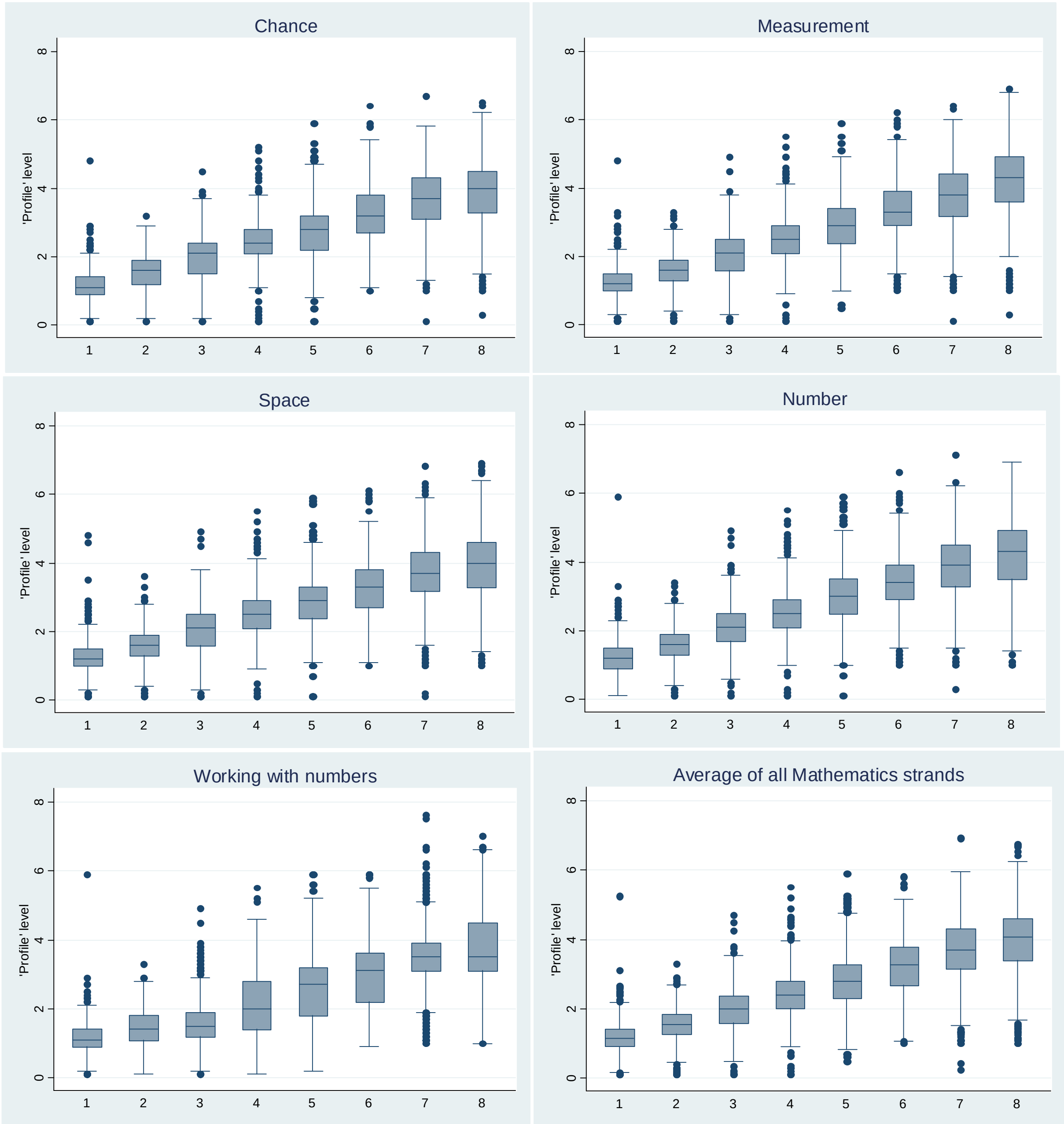
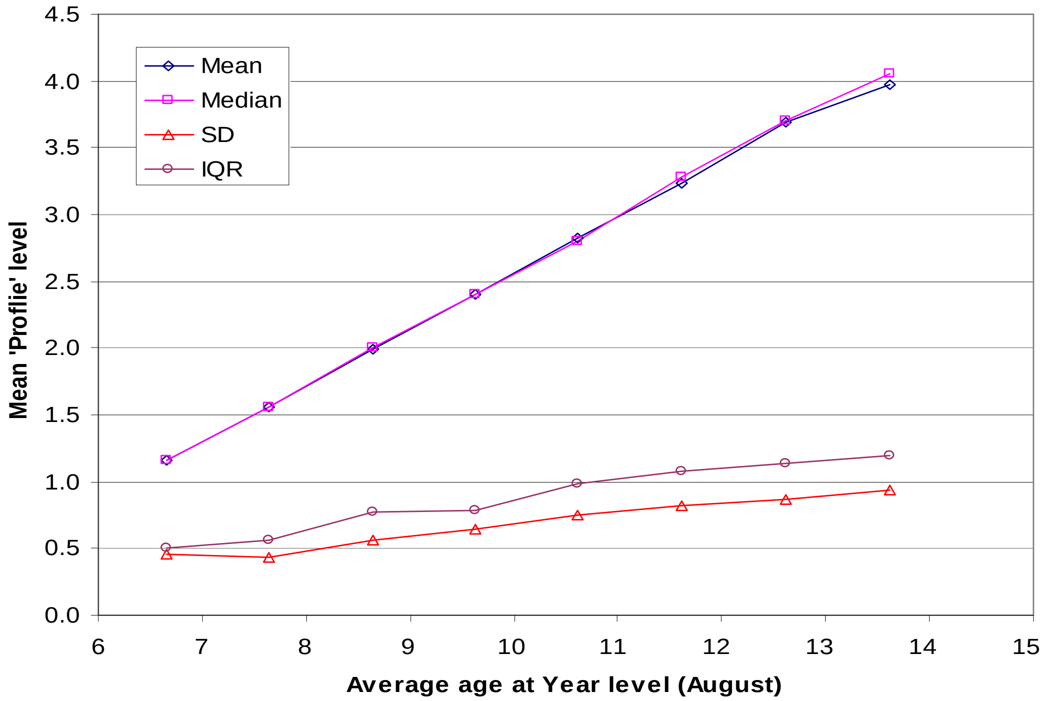


Table 7.2 Mathematics Learning data by Year level –1998: General Statistics

Year level	Average age at assessment (August 98)	Average of all Mathematics Strand values (in Profile units) Mean	Median	SD	IQR	SE of mean	Skew-ness	Kurtosis	N
1	6.66	1.16	1.16	0.46	0.51	0.01	0.45	7.52	1452
2	7.64	1.56	1.56	0.44	0.56	0.01	0.00	3.96	1456
3	8.64	1.99	2.00	0.57	0.78	0.01	0.08	3.71	1537
4	9.63	2.40	2.40	0.64	0.78	0.02	0.09	4.20	1548
5	10.62	2.83	2.80	0.75	0.98	0.02	0.18	3.43	1541
6	11.62	3.23	3.28	0.82	1.08	0.02	-0.08	2.89	1540
7	12.62	3.69	3.70	0.87	1.14	0.02	-0.26	3.34	1554
8	13.62	3.98	4.06	0.94	1.20	0.03	-0.35	3.39	1422
All	**10.14**	**2.61**	**2.48**	**1.17**	**1.74**	**0.01**	**0.36**	**2.50**	**12050**

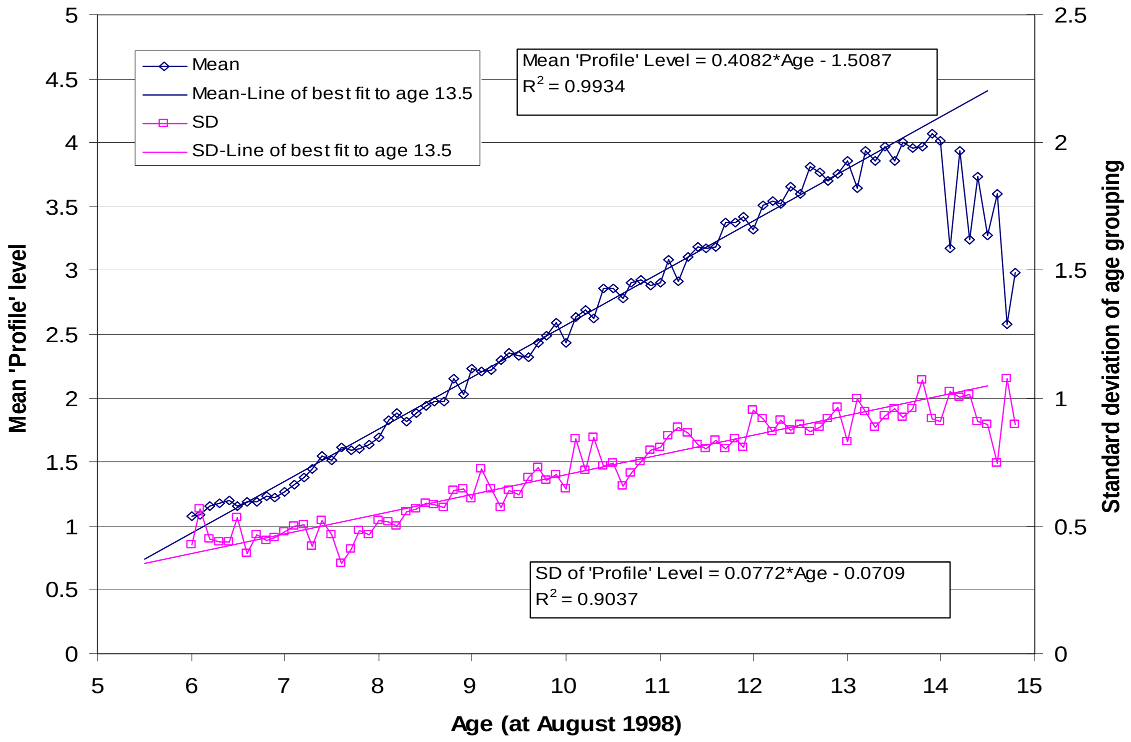
The consistency of the gradient of improvement in assessed learning indicates that many teachers, over multiple Year levels perceive the mean performance of students as having increased by a constant amount for each year of schooling. The spread of the assessments increases with Year level/age, reflected in SDs and IQRs. As for English, the Year 8 teachers assess students to be, on average, at a lower point than the continuation of the primary teacher gradient would expect.

Figure 7.9 Teacher Judgement assessments - Mathematics Learning Area 1998: means, medians, standard deviations and inter-quartile ranges by Year level



Clearly there are different perceptions of learning status by secondary teachers relative to primary teachers. Whether it is a reflection of the actual learning status, a cultural difference between how secondary and primary teachers see learning, or any of a range of other factors cannot be determined from this data. What is known is that from a test perspective, accepting the general model in Chapter 6, the amount of increase in mean learning status diminishes with Year level and age. This matter is taken up again in Chapter 8.

Figure 7.10 Teacher Judgement assessments - Mathematics Learning Area 1998 Mean profile level all strands combined, by age



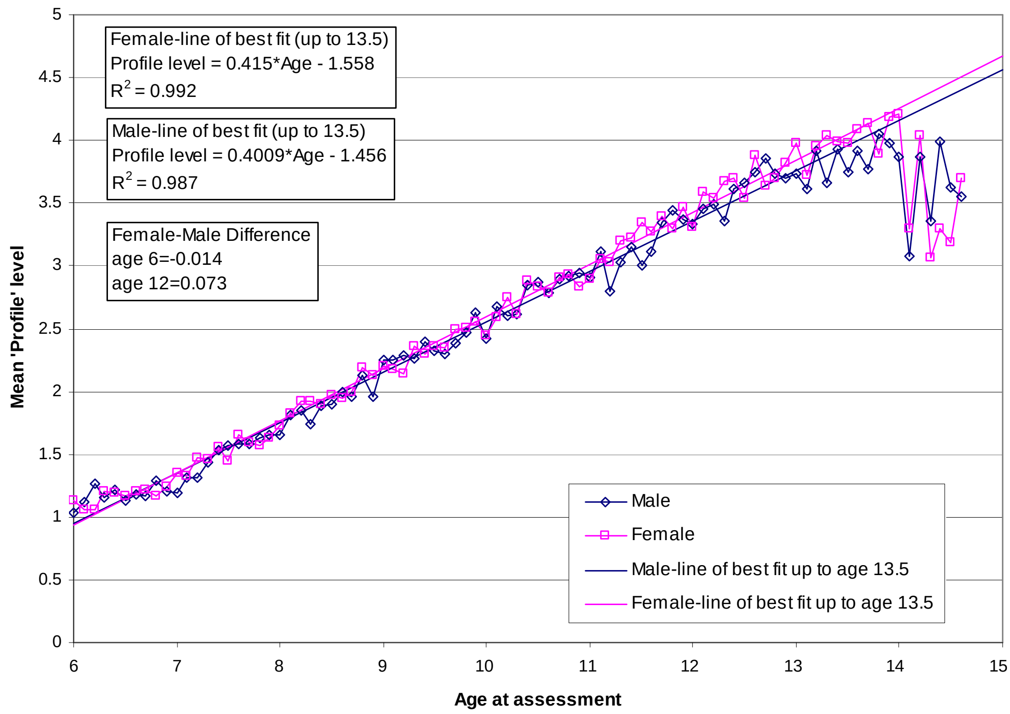
The pattern of higher mean learning status for age groupings at 0.1 of an age appears to hold for the mathematics assessments as it does for English. A straight line regression of mean learning status in profile level units up to age 13.5 has a gradient of 0.4 profile units per year of age, and an R^2 of 0.9934. There is clearly a strong relationship of improving learning status with age. The SDs also show a linear trend with age (to 13.5), with a linear regression as good a fit to the points as a quadratic - in mild contrast to the English data in Figure 7.4 where a flattening of the SD curve occurs after age 12.

Gender differences in the Mathematics Learning Area

Identifying the data by gender (by age), as illustrated in Figure 7.11, reveals a clear difference between teachers' perceptions of mathematics learning compared with teachers' perception of English learning. In Figure 7.11 the trajectories for males and females are intertwined. A regression of the means on age suggests a very slight advantage to females, and a slightly greater variability in the assessment of males (slightly lower R^2 for males). Compared with the clear gender difference found for English, in Figure 7.5 and confirmed in the test model in

Chapter 6, the difference in teachers' assessments by gender for mathematics are trivial. There is a hint that teachers at secondary level might perceive a small mean difference in favour of females, in contrast to the test model which suggests a small mean difference in favour of males. In both cases the differences are very small. Teacher judgment assessments in the Victorian VELS/CSF (Chapter 4) show the same tendency for teachers in higher Year levels to judge female students on average to be slightly ahead of males of the same age in their learning. In the next chapter the test data do not support this difference.

Figure 7.11 Teacher Judgement assessments- Mathematics Learning Area 1998 – Mean profile level of all strands combined, by gender of students



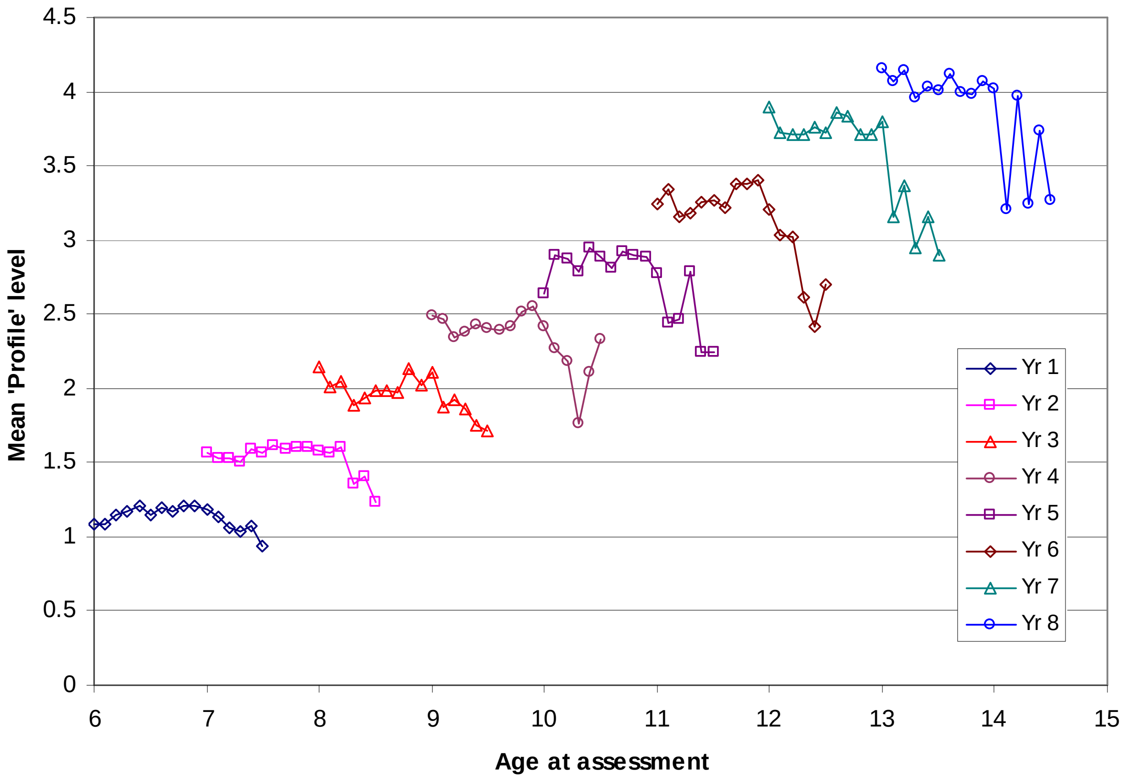
As for English the assessments for students up to age 14.6 are shown for completeness,

The general upward trend applies up to age 14.0. The older students in Year 8 up to age 14.6 are included but a very small number of students older than this are censored. Up to age 14.0 the students are of normal age for the Year level. As shown below in Figure 7.12, older students outside the normal age range for the Year level generate a tail effect. The proportion of students in the tail is small. Data in consecutive Year levels, summarised without considering Year level, do not show the tail due to the cancelling effect of the small numbers of under normal age high scores on the over normal age lower scores, leading to the age means following the general trend.

Within Year level trends by age

The age patterns within each Year level are shown in Figure 7.12. At Year 1 there is a positive gradient with age within the Year level up to the limit of the normal age group for the Year level (7.0 years in August), and then a tail of reducing means for the older students. However the negative gradient of the tail is much less than for the higher Year levels. This pattern is plausible, as mathematic development may not show such a large between-student variation in the early stages. The overall mean for Year 1 (1.16 profile level units from Table 7.2) is very low on the profile level scale and as shown in Figures 7.8 and 7.9, there appears to be a reticence to assess students as "not yet at level 1". These two aspects might explain the slighter tail. In the model in Chapter 6 the tail is similar to that of the upper Year levels, an artefact of being a direct derivation from the Year 3 test shape and the stretching of the distribution on the learning axis to match the steepness of the assumed learning trajectory.

Figure 7.12 Teacher Judgement assessments - Mathematics Learning Area 1998: Mean profile level of all strands combined by age within Year level



The pattern with age over the Year levels (above Year 1) while variable, indicates a general positive trajectory with age within a Year level up to the oldest within normal age point and then a tail for the above normal students. As for English, the data are censored with cases of very low frequency below normal age and above 0.5 above normal age not shown. At Year 8 it would appear that there is no relationship of mean learning status with age, within the normal age range for the Year level; that is the age effect appears to have disappeared. The older student tail still occurs, reflecting the generally lower mathematics skill of the over-age students.

Common findings across two data collection periods and learning areas

There are some strong common features of the view of student learning development with Year level and age as seen by teachers. The perceived shape of learning for the mean of Year level cohorts is the same whether the learning area is English or Mathematics. The shape holds when the assessments are summarised by age at a highly refined unit of age (0.1) approximating a month of age. The shape is linear with Year level/age until the end of primary school. This is in contrast to the shape indicated by test based IRT measures of ostensibly the same learning, which is curvilinear (Chapter 6).

A linear trajectory is typical in a grade-equivalent approach (Schulz & Nicewander, 1997) hinting that teachers may be using their expectations for a given Year level as the basis for their judgement. Teacher assessments referenced to the Profiles scale, replicated over time and subject, produce similar linear patterns. SDs increase with Year level/age. In contrast, IRT descriptions of learning development with age (Chapter 5) and the IRT related model for tests in South Australia (Chapter 6), the SDs generally reduce with age. The NAPLAN (2009) data for 2008 have this property of diminishing SD for all subjects except Writing.

The implication in the apparent grade-equivalent pattern is that teachers' assessment processes appear to draw on the general perception that teachers have of the standard for the Year level. This applies even though the assessments are strongly referenced to the level criteria framework to assist in the judgement of where in the framework a given student is placed. It is remarkable that this grade-equivalent standard, averaged over about 180 teachers per Year level in 1997 and 300 teachers per Year level in 1998, has such a strong linear relationship with Year level over the full primary Year level spectrum. The trend holds across the two subjects under investigation with slightly different gradients and based on general patterns found for the 6 other learning areas surveyed (Rothman, 1998; DECS Curriculum Bulletins 1998-1999), the same trend appears to hold generally across the other six learning areas.

Teacher judgement assessments of student learning have two subtle elements. The first is the apparently strong relationship of learning with decimal age. It is assumed that age differences of the order of a month of age, were not considered by teachers in making their assessments. The age effect applies to the data undifferentiated by Year level as well as within Year levels, at least below Year 7.

The second element is the clear perception by teachers that females, on average, at a given age have a slight performance advantage over males in English language development. Contrasting with this is the different perception of mathematics learning where there appears to be hardly any gender effect (notwithstanding the consistent but subtle difference at higher Year levels between test and teacher judgement assessments). It is unlikely that gender would have been consciously considered when the assessments were being made. The gender differences for English, when compared in Chapter 8 with the patterns identified by test summaries, follow approximately the same patterns implying the effects are not merely the result of a bias. Slightly higher results for females, however, appear to apply in teacher judgements of Mathematics at higher Year levels.

The data also indicate that teachers generally view learning at a fine degree of refinement. How well this refinement matches the scale from the testing process generally and for individual students is addressed in the next chapter. It is observed here that teachers used the full spectrum of available points (at 0.1 of a Profile level) for the population of students in a Year level. The response format (clicking on a continuous line) did not allow a teacher to see that any judgement was at this level of refinement nor is it possible to resolve from this data collection how teachers would feel about applying a numerical value to learning progress. However the principle is clear that teachers might be able to estimate learning status at the refinement of about the equivalent 2 to 4 weeks of learning development (though not necessarily at this implied frequency).

Acceptability of teacher judgement assessment to teachers

As documented in Chapter 3 the introduction of the Statements and Profiles for Australian Schools (SPFAS), and its precursor in South Australia, achievement levels, was controversial and led to teacher industrial concerns about workload and the possible misuse of assessments. Teachers' confidence in the their judgements was considered to be important to establish and thus data on confidence was collected as a condition for union and teacher participation. Teachers were asked two confidence related questions at the completion of their judgements: teachers' confidence in the process generally and their confidence in the assessment for each student individually. A five-point rating scale was used with a rating of 5 indicating the greatest confidence.

Table 7.3 reports the percentage of responses at each point on the 5-point scale for both questions, along with the 'no response' rates. Confidence in the general process appears to be less than the confidence in the assessment for individual students, based on the combined percentage of responses at rating points 3, 4 and 5. Confidence in the process was around 60%, assuming that the highest ratings of 3 to 5 reflect a positive view of the process. The more specific confidence in each individual student assessment was higher. Over the combined ratings 3 to 5, confidence was around 70%. Given the industrial concerns about the introduction of the assessment reporting and the directly-connected national curriculum (covered in Chapter 3), the confidence ratings appear surprisingly positive.

Table 7.3 Ratings by teachers of their confidence in the process and in their specific assessments.

	Confidence in the process		**Confidence in specific student assessments**	
	1997	**1998**	**1997**	**1998**
No rating	20.7%	32.3%	20.6%	18.6%
1	7.3%	2.9%	5.9%	3.5%
2	7.9%	6.0%	5.7%	7.5%
3	17.5%	22.8%	17.9%	25.2%
4	29.5%	27.2%	31.3%	34.5%
5	17.1%	8.7%	18.6%	10.7%
3+4+5	**64.1%**	**58.7%**	**67.8%**	**70.4%**

Concluding comments

The introduction of a profiles approach to curriculum and assessment in schools in South Australia was driven by curriculum leaders rather than assessment advocates. The focus of curriculum leaders was essentially and appropriately on how the general profiles framework might help schools and teachers match and refine their existing curriculum arrangements to the described structure of developmental learning. The relationship of the statements and profile structures to assessment and recording of student learning was not fully addressed. The ways in which individual and personal assessments of learning status using a developmental map might be applied to add refinement to a profile level assessment were not often considered. The purpose of the collection of data was not clearly resolved and delayed a number of years. The mechanics for the collection were addressed only in the negotiations to conduct the collection (as described in Chapter 3). In particular the level of resolution possible for a teacher judgement assessment was not considered in a way that could have led to pretested process for the possible degree of resolution. As a result the collection of data was an imperfect process.

A key concern was the perception by teachers that a profiles-based assessment might apply only irregularly. For many teachers, assessments were made only for the two data collections and not embedded as general classroom records. This is unsurprising given that the record of learning status was of such low resolution that keeping records would be seen as a waste of teacher effort. A personal student history in profile level units would not contribute to the day-to-day learning support for students. However, the data summarised above suggest that teachers' judgements provide a very comprehensive overview of what learning development looks like over 8 years of schooling. The consistency of the annual increment in the means of teacher judgement assessments suggests that some powerful underlying perception of learning growth is understood by a sufficient number of teachers to ensure the means grow as observed. Not all teachers need to have this ability for the pattern to exist, and persist, over two collection periods. However it seems for all 8 of the learning areas described in the curriculum description (SPFAS) the same general linear trajectory pattern of the mean of the Year level assessments applied.

The data collected in 1997 and 1998 indicate that teachers can articulate an on-balance judgement assessment using a well-described logical framework, even if the framework has many ambiguities. The allocation of a numerical value to learning status, notionally the same process that applies with formal testing, is feasible and, in operation, the data generated are consistent with what is expected in learning development. The major difference, in broad terms, between learning development as seen by teachers and as described by vertically scaled IRT tests is the shape of the general trajectory of learning. Teachers using the SPFAS framework generate data that describe a linear trajectory with increasing spread. IRT data describe trajectories that are non-linear, with mean growth per period reducing as upper segments of the learning scale are approached. The spread of students around the trajectory reduces rather than increases. What is not clear from examining the teacher data in isolation is the extent to which assessments of individual students by teachers and tests produce equivalent assessments for a student. The next chapter explores a range of ways that data above and data from test sources can be compared as grouped data and for individual students.