Appendix 12 Summary of Rasch analysis statistics for teacher judgement assessment data

Table A12.1 1997 English

```
     SUMMARY OF 7868 MEASURED (NON-EXTREME) Persons
+--------------------------------------------------------------------------------+
|           RAW                          MODEL          INFIT        OUTFIT     |
|          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD  |
|--------------------------------------------------------------------------------|
| MEAN      80.0       2.9       -1.47     .26       .86    -.8    .86    -.8  |
| S.D.      33.2        .3        1.99     .08      1.54    1.7   1.54    1.7  |
| MAX.     208.0       3.0        3.13    1.13      9.90    9.9   9.90    9.9  |
| MIN.       1.0       1.0      -10.80     .05       .00   -3.9    .00   -3.9  |
|--------------------------------------------------------------------------------|
| REAL RMSE    .33  ADJ.SD    1.96  SEPARATION  6.00  Person RELIABILITY  .97  |
|MODEL RMSE    .27  ADJ.SD    1.97  SEPARATION  7.18  Person RELIABILITY  .98  |
| S.E. OF Person MEAN = .02                                                      |
+--------------------------------------------------------------------------------+
  MINIMUM EXTREME SCORE:      4 Persons
       VALID RESPONSES:  97.9%Person RAW SCORE-TO-MEASURE CORRELATION = .94
(approximate due to missing data)
CRONBACH ALPHA (KR-20) Person RAW SCORE RELIABILITY = 1.00 (approximate due to missing
data)

     SUMMARY OF 3 MEASURED (NON-EXTREME) Items
+--------------------------------------------------------------------------------+
|           RAW                          MODEL          INFIT        OUTFIT     |
|          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD  |
|--------------------------------------------------------------------------------|
| MEAN  209852.3    7705.0         .00     .00       .94   -1.9    .93   -2.5  |
| S.D.    1773.2      47.4         .02     .00       .19    8.2    .19    8.8  |
| MAX.  212360.0    7772.0         .01     .00      1.18    9.4   1.19    9.9  |
| MIN.  208583.0    7671.0        -.03     .00       .72   -9.9    .73   -9.9  |
|--------------------------------------------------------------------------------|
| REAL RMSE    .00  ADJ.SD     .02  SEPARATION  3.47  Item   RELIABILITY  .92  |
|MODEL RMSE    .00  ADJ.SD     .02  SEPARATION  3.58  Item   RELIABILITY  .93  |
| S.E. OF Item MEAN = .01                                                        |
+--------------------------------------------------------------------------------+
UMEAN=.000 USCALE=1.000
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00 (approximate due to missing data)
23115 DATA POINTS. APPROXIMATE LOG-LIKELIHOOD CHI-SQUARE: 100312.54
```

No of iterations = 741

Figure A12.1 1997 English-Distribution of Infit Mean Square values of fit to Rasch model.

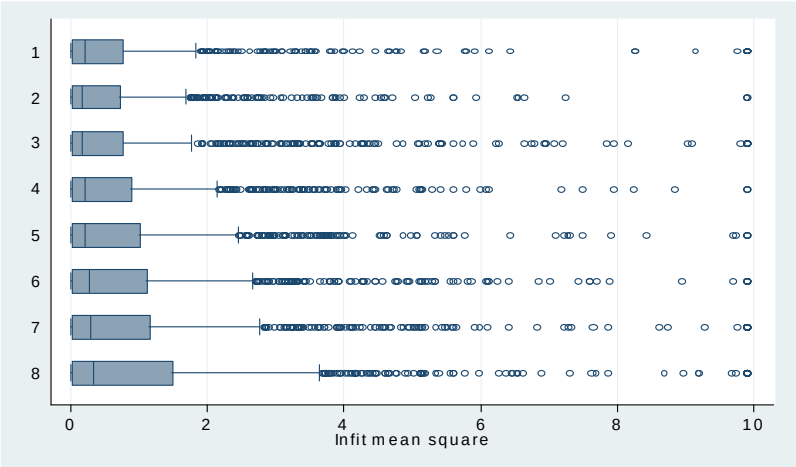


Table A12.2 1997 English subset Years 3 and 5 with both Test and Teacher assessment n=1275- Means and SDs equated for the common subset and then solution applied to all cases.

		Test scores	Teacher Rasch Analysis No Anchor	Teacher Rasch Mean and SDs transformed to match test	Teacher Rasch No anchor Original Measurement error	Teacher Rasch No anchor re-scaled Measurement error
Matched Y3 & Y5 only	Mean	1.06	-1.64	1.06	0.26	0.26
	SD	1.37	1.33	1.37	0.06	0.06
	N	1275	1275	1275	1275	1275
All cases	Mean	1.06	-1.47	1.23	0.26	0.27
	SD	1.37	2.00	2.07	0.09	0.09
	N	1275	7872	7872	7872	7872

Table A12.3 All Yrs 1997 Profiles Assessments: Largest Standardized Residual Correlations Used To Identify Dependent Items

RESIDUL CORRELN	ENTRY NUMBER Item	ENTRY NUMBER Item
-.65	2 Writing	3 SpeakListen
-.56	1 Reading	3 SpeakListen
-.27	1 Reading	2 Writing

Table A12.4 All Yrs 1997 Profiles Assessments Table Of Standardized Residual Variance

	Empirical		Modeled	
Total variance in observations =	96.8	100.0%		100.0%
Variance explained by measures =	93.8	96.9%		96.7%
Unexplained variance (total) =	3.0	3.1%	100.0%	3.3%
Unexplned variance in 1st contrast =	1.7	1.8%	57.9%	
Unexplned variance in 2nd contrast =	1.3	1.3%	42.1%	
Unexplned variance in 3rd contrast =	.0	.0%	.0%	
Unexplned variance in 4th contrast =	.0	.0%	.0%	
Unexplned variance in 5th contrast =	.0	.0%	.0%	

Table A12.5 1998 Mathematics

```
     SUMMARY OF 12139 MEASURED Students
+------------------------------------------------------------------------------+
|           RAW                          MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|------------------------------------------------------------------------------|
| MEAN     104.0       3.9        -.73     .19       .71   -1.0    .71   -1.0 |
| S.D.      46.6        .4        1.59     .06      1.19    1.6   1.20    1.6 |
| MAX.     270.0       4.0        3.93    1.12      9.90    9.9   9.90    9.9 |
| MIN.       1.0       1.0       -8.60     .13       .00   -3.9    .00   -3.8 |
|------------------------------------------------------------------------------|
| REAL RMSE    .23  ADJ.SD    1.57  SEPARATION  6.82  Studen RELIABILITY  .98 |
|MODEL RMSE    .20  ADJ.SD    1.57  SEPARATION  7.81  Studen RELIABILITY  .98 |
| S.E. OF Student MEAN = .01                                                   |
+------------------------------------------------------------------------------+
      LACKING RESPONSES:      8 Students
        VALID RESPONSES:  97.9%
Student RAW SCORE-TO-MEASURE CORRELATION = .95 (approximate due to missing data)
CRONBACH ALPHA (KR-20) Student RAW SCORE RELIABILITY = .98 (approximate due to missing
data)

     SUMMARY OF 4 MEASURED Profile levels
+------------------------------------------------------------------------------+
|           RAW                          MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|------------------------------------------------------------------------------|
| MEAN  315703.0   11889.7         .00     .00       .75   -9.9    .73   -9.9 |
| S.D.   14358.9     211.0         .08     .00       .10     .0    .10     .0 |
| MAX.  332136.0   12104.0         .11     .00       .87   -9.9    .85   -9.9 |
| MIN.  293402.0   11543.0        -.10     .00       .59   -9.9    .58   -9.9 |
|------------------------------------------------------------------------------|
| REAL RMSE    .00  ADJ.SD     .08  SEPARATION 24.57  Profil RELIABILITY 1.00 |
|MODEL RMSE    .00  ADJ.SD     .08  SEPARATION 24.57  Profil RELIABILITY 1.00 |
| S.E. OF Profile leve MEAN = .05                                              |
+------------------------------------------------------------------------------+
UMEAN=.000 USCALE=1.000
Profile level RAW SCORE-TO-MEASURE CORRELATION = -.99 (approximate due to missing
data)
47559 DATA POINTS. APPROXIMATE LOG-LIKELIHOOD CHI-SQUARE: 214755.62
```

No of iterations = 275

Note: 89 cases subsequently deleted due to Teacher assessments providing zero data or only one of four strands (items).

Figure A12.2 1998 Mathematics-Distribution of Infit Mean Square values of fit to Rasch model

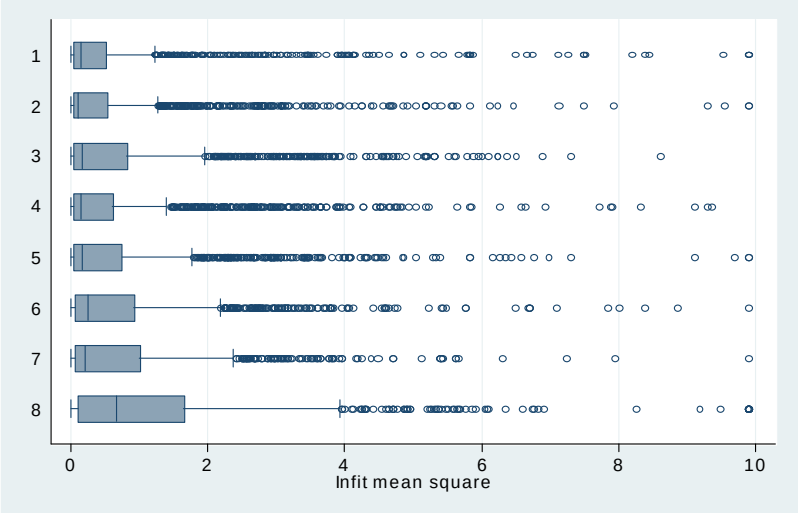


Table A12.6 1998 Mathematics subset Years 3 and 5 with both Test and Teacher assessment n=2105- Means and SDs equated for the common subset and then the solution applied to all cases

		Test scores	Teacher Rasch No Anchor	Teacher Rasch Mean and SDs transformed to match test	Teacher Rasch No anchor Original Measurement error	Teacher Rasch No anchor Re-scaled Measurement error
Matched Y3 & Y5 only	Mean	0.77	-0.70	0.76	0.18	0.27
	SD	1.44	0.99	1.44	0.03	0.04
	N	2105	2105	2105	2105	2105
All cases	Mean	0.77	-0.72	0.74	0.19	0.28
	SD	1.44	1.58	2.28	0.06	0.09
	N	2105	12050	12050	12050	12050

Table A12.7 All Yrs 1998 Profiles Assessments: Largest Standardized Residual Correlations Used To Identify Dependent Items

RESIDUL CORRELN	ENTRY NUMBER Profile lev	ENTRY NUMBER Profile lev
-.47	1 Chance	3 Number
-.37	3 Number	4 Space
-.33	2 Measurement	4 Space
-.30	1 Chance	2 Measurement
-.28	1 Chance	4 Space
-.23	2 Measurement	3 Number

Table A12.8 All Yrs 1998 Profiles Assessments: Table Of Standardized Residual Variance

		Empirical		Modeled	
Total variance in observations	=	155.6	100.0%		100.0%
Variance explained by measures	=	151.6	97.4%		96.5%
Unexplained variance (total)	=	4.0	2.6%	100.0%	3.5%
Unexplned variance in 1st contrast	=	1.5	1.0%	37.8%	
Unexplned variance in 2nd contrast	=	1.3	.8%	32.7%	
Unexplned variance in 3rd contrast	=	1.2	.8%	29.5%	
Unexplned variance in 4th contrast	=	.0	.0%	.0%	

Figure A12.3 Comparisons of 'Item' difficulty relationships

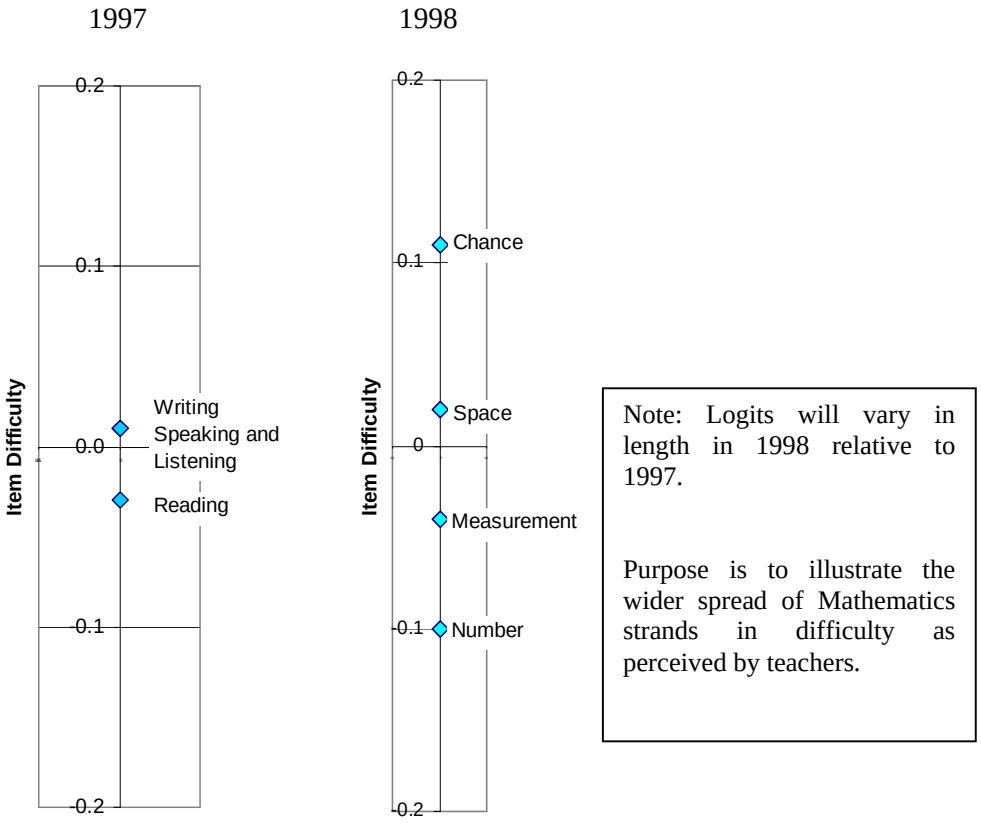
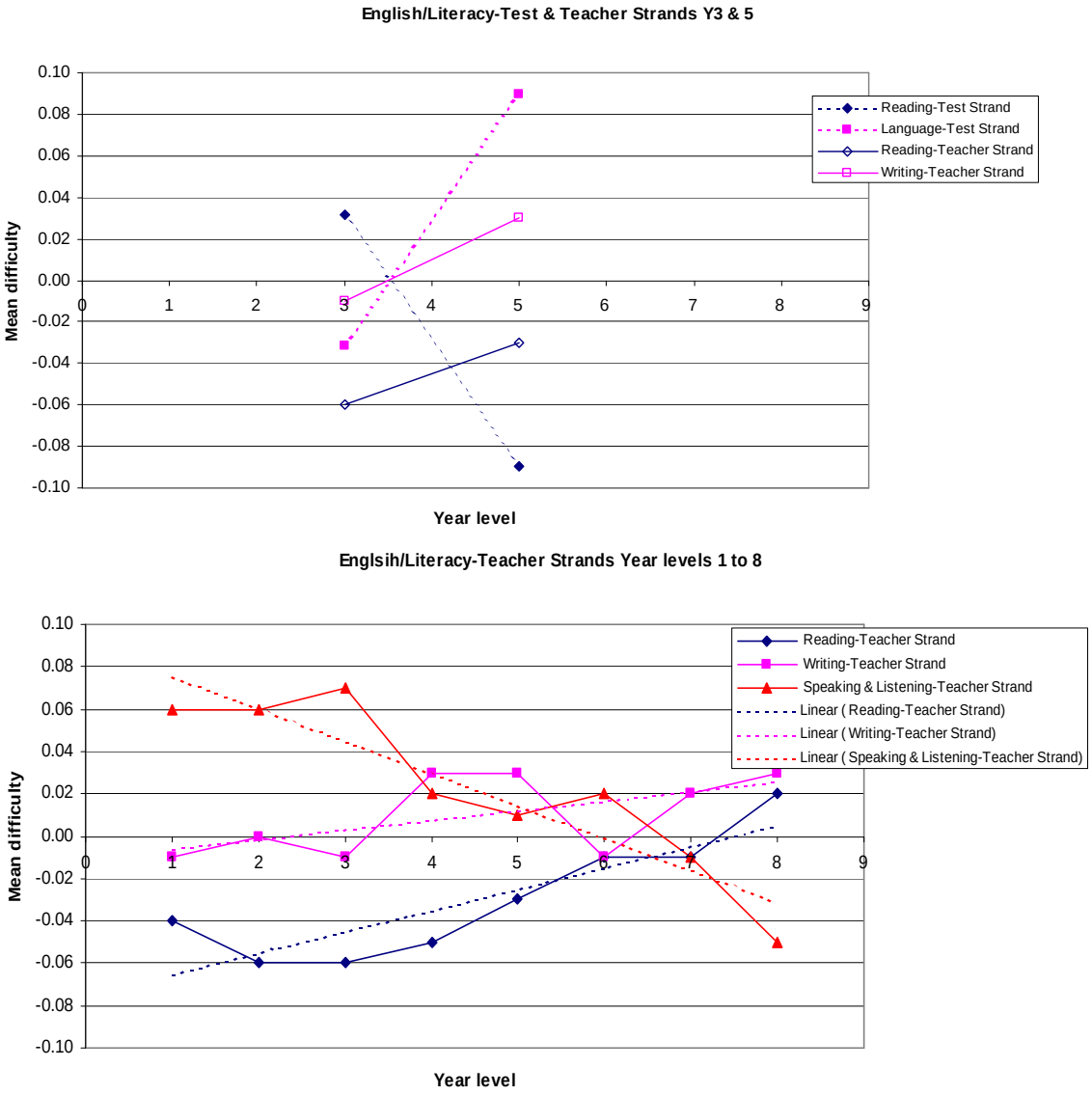


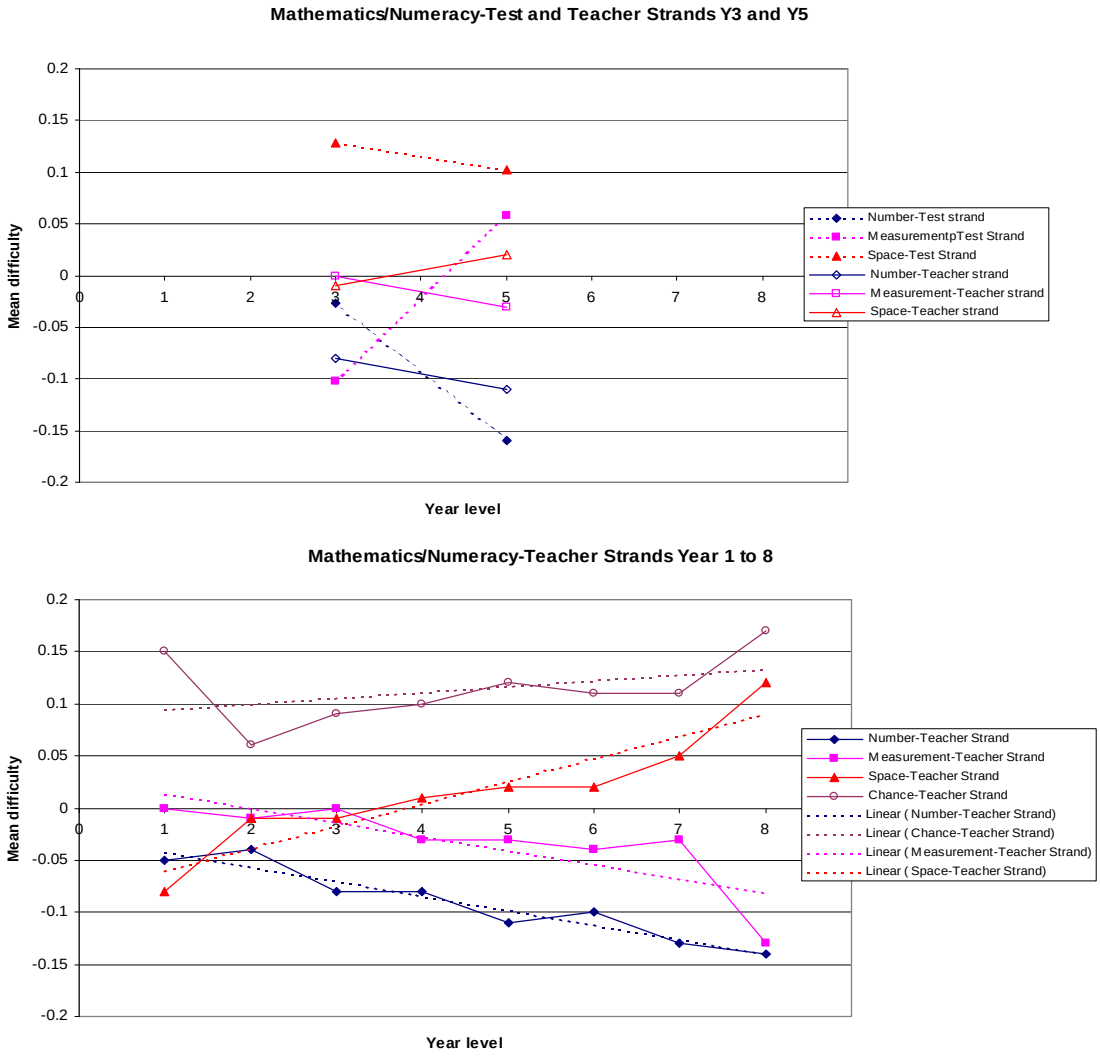
Figure A12.4 English Teacher and Test assessments compared by Strand difficulty



The top panel compares Year 3 and Year 5, the only Year levels for which data from both sources occurred. Test strand data were created from test population means of sets of items designated as Reading or Language. Language includes elements of writing, spelling and grammar and is most likely not directly comparable to Writing. Strand difficulties are created by the difference between population means with the mid point in difference set to 0 and the general scale reversed so that 'easy' is lower on the scale. On this basis Test Strand difficulties were about 0.05 logits apart at Year 3 with Language easier. By Year 5 they were 0.18 logits apart and with Reading now easier than Language. For Teacher assessments Reading was easier than Writing and while both became more difficult they stayed in the same general relationship.

The lower panel shows the trends in strand difficulty by Year level based on Differential Item Function. (In this analysis strands are items.) As Year level increases Reading and Writing as seen by teachers appears to get harder; Speaking and Listening becomes easier.

Figure A12.5 Mathematics Teacher and Test assessments compared by Strand difficulty



The top panel compares Year 3 and Year 5, the only Year levels for which data from both sources occurred. Test strand data were created from test population means of sets of items designated as Number, Measurement and Space (Chance is not identified in the Test design). Strand difficulties are created by the difference between population means with the mid point in difference set to 0 and the general scale reversed so that 'easy' is lower on the scale.

On this basis Test Strand difficulties were about 0.22 logits apart at Year 3 with Measurement the easiest. By Year 5 the spread has become 0.26 logits apart and with Number now easier than Measurement. Space is hardest in both periods. For Teacher assessments Number was easier than either Measurement or Space. Number and Measurement become less difficult by Year 5. Space is consistently the hardest in both assessment processes.

The lower panel shows the trends in strand difficulty by Year level based on Differential Item Function. (In this analysis strands are items.) As Year level increases Number and Measurement as seen by teachers appears to get easier; Space and Chance become harder and Chance remains the strand seen as hardest to achieve a high assessment.