*Australian Journal of Educational & Developmental Psychology*. Vol 4, 2004, pp 86-96

# Reading in one's Ethnic Language: A study of Greek-Australian High School Students

James A Athanasou[1]
University of Technology, Sydney
&
Iasonas Lamprianou
University of Manchester

## ABSTRACT

This paper examines reading achievement when the maternal/paternal language has become a de facto second language. The performance of a cohort of Greek-Australian high school students (N=270) on a diagnostic Greek reading test was significantly below that of pupils in second to fourth grades in Greece. The mean item difficulty for Greek-Australian high school students was 0.35 compared with 0.51 for second grade, 0.69 for third grade and 0.80 for fourth grade pupils in Greece. The pattern of responding indicated that the Rasch model fitted the data. The effects of background factors were also examined. Students whose mother or father spoke Greek had statistically significant higher levels of ability than those who spoke Greek and English or English alone. It was also found that the length of the key words (correct responses) had a large effect on the difficulty of the questions; the longer the key word, the more difficult the question was. The implications of the results for assessment of reading in and the acquisition of an ethnic or second language are discussed.

## INTRODUCTION

Multicultural Australia offers a useful large scale natural laboratory for the investigation of second-language reading and comprehension. In this context it may reasonably be hypothesised that second-language acquisition, reading and comprehension would operate normally at a lower level of achievement than one's first language. The reader's thoughts might turn towards the acquisition of English as a second language but our focus is towards those cases where one's original maternal/paternal or ethnic language becomes the *de facto* second language, that is, a language of minor use. Quite reasonably one might expect that there will be a lower level of ability in the areas of reading, writing and speaking when compared with the first language as well as substantial individual variations in literacy. Comparisons might be made against a benchmark and a helpful basis is to consider the language or reading development of a cohort in the country of origin.

The purpose of this paper is to examine the level of Greek reading ability of a sample of Greek-Australian high school students on a standardised Greek reading test. There is a paucity of research in this field and the few previous studies have emphasised the communities of practice associated with teaching or learning (Bradshaw & Truckenbrodt, 2003) or the extent of Greek

[1] James A Athanasou
Faculty of Education
University of Technology, Sydney
PO Box 123, Broadway 2007
Email: Jim.Athanasou@uts.edu.au

language instruction (e.g., Hejek & Nicholas, 2004). Earlier - as part of an ongoing program of research into ethnic identity - Athanasou, Georgoussis and Petoumenos (1996, in Greek) found that levels of reading were correlated 0.235 with interest in Modern Greek language but they did not examine the levels of performance within a coherent model of measurement. This report includes an application of item response analysis to indicate the types of items that are suitable or pose difficulties for such students in Australia and represents the first application of Rasch measurement in this specific domain. The results have implications for the extent of ethnic language acquisition of children through systems of formal education in a multicultural context.

One of the benefits of using Rasch analysis is the transformation of the ordinal raw scores to produce linear measures. Although this was initially advocated by George Rasch (1960), significant research is still being produced towards this direction. For example, Romanoski and Douglas (2002), have shown that untransformed raw scores produce underestimated main effects and spurious interaction effects in the context of Analysis of Variance. Smith and Taylor (2004) elaborated on the generation of comparable interval measures from the raw scores from two instruments that were built to measure the same underlying construct but used different wording of the items and different rating categories. Bezruczko (2004) has also demonstrated that the ordinality of raw scores can obscure the measurement process and cited early work of Cronbach who warned researchers many years ago about the problems of using raw scores.

This study has used a procedure that locates tasks according to their difficulty onto the same dimension as individuals by their level of competence (Rasch 1960). If a set of consistent tasks is used, then the tasks can be located by their level of difficulty along a dimension. Placing a person by the level of his/her performance on the same dimension makes it possible to interpret their performance since he/she is probably able to respond to tasks with difficulty levels below their performance level but not to tasks with difficulty above their ability. Use of this kind of psychometric model has been advocated in educational and psychological measurement (Embretson, 1996, p. 341) and found early application in measuring acquisition of English as a second language (see McNamara, 1996).

This Rasch approach to measurement is increasingly replacing the traditional approach to assessment based on scores, also known as classical item analysis. The pitfall in classical item analysis is that the difficulty of the tasks and the reading ability of the students are interdependent. That is, the abilities (or scores) of the students are dependent upon the difficulty of the task and the difficulty of a task reflects the abilities of the students. Rasch models use a simple probability equation to describe the relation between a characteristic of the person being assessed and the items, tasks or questions assessing this characteristic (Bond & Fox, 2001). A Rasch analysis of item responses and abilities may appear complex but really it just hypothesises that the probability of answering any question correctly is modelled by a function of the difference between the absolute level of ability of a person and the absolute level of difficulty of a task. For a particular person, the less difficult the item then the greater the probability of answering it correctly. The probabilities are transformed to measurements of log odd ratios or logits. The remaining sections of this paper describe the application of Rasch measurement to a Greek diagnostic reading test.

In this study we looked at the performance of a Greek-Australian cohort on a standardised Greek diagnostic reading test. Of course, it is recognised that items that were suitable in their country of origin may or may not be ideal for assessing the reading ability of a Greek-Australian sample. A standardised test was used as the starting point because it enabled comparisons to be made with data from a Greek cohort and avoided any comparisons based on *ad hoc* measures that did not build upon previous knowledge. The methodology is outlined in greater detail in the following section.

The purpose of this study goes beyond a simple examination of the difference in performance on a standardised reading test between Greek primary school pupils and Greek-Australian high school students. Even a layperson's cursory examination would reveal an obvious difference in Greek-speaking ability between Greek-Australians (for whom Greek is essentially a part-time or casual second language) and native Greek speakers. It is more important to explore the pattern of responses of Greek-Australian students. At the outset it is proposed that the difficulty of items would be fairly similar across samples but that the native Greek speakers would out-perform Greek-Australian students. To summarise, the key research questions are: (a) is there a difference

in reading ability between Greek-Australian high school students and Greek pupils on a standardised reading test?; (b) is the item difficulty of the tasks related across these two groups – in other words do they find the same types of tasks difficult?; (c) what are the patterns of response to reading items in the Greek-Australian sample?

## METHOD

### Participants

The participants in this study were 270 high school students (157 boys, 113 girls) in Years 7 to 9 from the three full-time Greek Orthodox High Schools in Sydney. The average age was 12.7 years (SD = 1.0).

### Instruments and Procedure

Participants completed the *Diagnostic Test of Reading Ability* (Tafa, 1995) which is a standardised 42-item objectively scored, multiple-choice reading comprehension test that requires respondents to complete a statement. Some translated examples of the sample items are listed in Table 1. The test is designed for infants to early primary school and was normed on some 2,518 pupils. The mean item difficulty across second to fourth grade pupils was 0.67 (Tafa, 1995, p. 41) and the internal consistency reliability of the test that is cited (p. 33) for the normative sample in the manual is 0.94 (Cronbach alpha).

**Table 1:** Sample translated example items from the *Diagnostic Test of Reading Ability* (Tafa, 1995)

| |
|---|
| The …….. is an animal.<br>**(milk / cat / grandmother / pot )** |
| I …… many games at home.<br>**(eat / excel / have / wear)** |
| Many years have passed …… I saw him last.<br>**(now / when / again / since)** |
| I went to the neighbourhood post office, because I wanted to buy some ……<br>**(stamps / books / carnations / milk)** |

In addition, students completed separately some demographic background details that related to their Greek ethnic background and identity and language use. Some demographic details of the sample are indicated mainly for the information of the reader in Table 2. The values of these five background factors relating to cultural exposure are defined in Table 2. Students completed the questionnaires during class time. The purpose of the study was explained to them and they were offered a copy of their results. The administration of the questionnaires was supervised by the author and a research assistant.

### Analysis

The results from this testing were analysed for the classical test statistics of item difficulty, item-total correlation and internal consistency. These classical test statistics complemented the analysis in terms of the one-parameter Rasch item response model (BigSteps version 2.82;Linacre & Wright, 2003). The results are presented graphically in an item map on which both items and persons are calibrated on a logit scale. The principal advantage of the Rasch model is that items of Greek reading can be located on a calibrated scale. The scale is invariant across items and groups. Further items can be located on the same scale and the performance of future students can also be determined on the same scale. During the analysis, the software constrained the average difficulty of the items to zero logits.

**Table 2:** Demographic and cultural background of the sample of high school pupils

| Factors | Percentage |
|---|---|
| Born in Australia | 93% |
| Mother born in Australia | 32% |
| Father born in Australia | 14% |
| Language mother mainly speaks to you: | |
| Greek | 25% |
| English | 34% |
| Greek/English | 40% |
| Language father mainly speaks to you: | |
| Greek | 30% |
| English | 27% |
| Greek/English | 42% |

Note: all percentages rounded.

In order, however, for the Rasch model to retain its desirable psychometric properties (e.g. invariance), the fit of the model was investigated by means of the Infit and Outfit Mean Square (Wright & Masters, 1982; Wright & Stone, 1979). Residual-based fit statistics like the Infit Mean Square and the Outfit Mean Square have been successfully used to assess the fit of the Rasch models on datasets for some time (e.g. Smith, 1991; Wright & Mok, 2000; Smith, 2000). One major difference between the Infit and the Outfit Mean Square is that the Outfit Mean Square is more sensitive to unexpected responses when the ability of the respondent is far from the difficulty of the item. For example, the Outfit Mean Square will be inflated when a low-ability person gives a correct response to a very difficult item. Only a few unexpected responses of this kind could inflate the Outfit Mean Square significantly, although the item may, other than that, have a very good overall fit

Because it was the first time that the Rasch model has been applied in this context, it was decided not to use pre-defined cut-off scores but instead to investigate the fit of the model on the data more thoroughly. It was resolved to sort the items and the persons according to their fit statistics (in a descending order) and to examine the most misfitting/most overfitting of them more thoroughly. For the most extreme cases, the sources of the misfit would be studied by means of investigating the residual matrix.

Students with extreme (zero or full) scores were removed from the dataset before the Rasch analysis. This is a well known feature of the unconditional maximum likelihood estimation employed by a large array of Rasch analysis software. Students with a full score are assumed to be more able than the most difficult item but it is not possible to estimate how much more able they are. Students with a zero score are assumed to be less able than the easiest item but it is not possible to estimate how able they are. All in all, 264 students remained in the analysis. No items had extreme scores and all of them remained in the Rasch analysis.

## RESULTS

### The fit of the model

Although a theoretical perfect fit of a statistical model on an empirical dataset can never be achieved, the fit of the Rasch model on the dataset was satisfactory for all practical intents and purposes. Figure 1 illustrates information about the item fit. The items are sorted according to the magnitude of their fit statistics.

Both Infit and Outfit MNSQR (the two statistics we used in the research) have been described as approximately $x^2$–distributed (Wright & Mok, 2000). However, no universally accepted cut-off scores have been suggested. Karabatsos (2000) has shown that the distributional properties of the two statistics could differ significantly from dataset to dataset and rules of thumb are usually used. For example, Wright and Linacre (1994, p.1) suggested that for high-stakes exams the limits for both *v* and *u* should be set to 0.8 (indicating lack of stochasticity) and 1.2 (indicating excessive

"noise" – unpredicted by the Rasch model). Bond and Fox (2001) suggested suggest the use of cut-off scores of 1.3 and 0.7. Moreover, Karabatsos (2000) also explains that only convention suggested the use of a rule-of-thumb of 1.3 for misfit and 0.7 for overfit. He went on to illustrate that the use of any pre-defined cut-off scores is wrong and that different cut-off scores may be more appropriate for different data sets.

The main finding from Figure 1 is that no items have fit statistics outside of the typical rule-of-thumb cut-off scores. For example, all the Infit Mean Square statistics of the items are smaller or equal than the 1.2 cut-off score. All the Outfit Mean Square statistics are within the 1.3 cut-off score. Starting from the most misfitting item and progressing towards the least misfitting, the table of residuals was inspected to identify the sources of aberrance for each of the items. It was realized that the sources of misfit (large residuals) could not be attributed to any apparent reasons. For example, students that were not born in Australia or were not speaking in Greek with their parents were not more/less likely to produce larger residuals. The residuals on this item seemed to be randomly distributed across the sample with no apparent patterns. The same results were found for the rest of the more misfitting items (e.g. items 'Founded', 'Disinterest'). Having addressed the fit of the model, we now consider each research question.

*Is there a difference in reading ability between Greek-Australian high school students and Greek pupils on a standardised reading test?*

As expected the task was clearly more difficult for second-language learners. The average score for Greek-Australian high school students was 15 (*SD* = 6.5) compared with an average of 28.1 for primary pupils in Greece (*SD* = 9.9, N= 2,518; Tafa, 1995, p. 29). The mean item difficulty for Greek-Australian high school students was 0.35 compared with 0.51 for second grade, 0.69 for third grade and 0.80 for fourth grade pupils in Greece (Tafa 1995, p. 41). These differences are substantial and supported the view that the average level of reading ability Greek-Australian high school pupils on this test was well below that of even infants grade levels in Greece.

*Is the item difficulty of the tasks related across these two groups?*

The item difficulties are reported in Table 3. The easiest item for the Greek-Australian sample related to 'rain' (76% answering correctly) and the most difficult related to the word 'clear' (8% answered correctly). The correlation of 0.57 between the 42 levels of item difficulty for the Greek-Australian and the normative Greek sample suggested that the pattern of task difficulty was moderately related (but of course different in absolute terms) irrespective of first or second-order language acquisition.

The overall internal consistency of the test was determined using Cronbach's alpha (0.82). This indicated substantial homogeneity of the questions and was consistent with the high alpha of 0.94 reported by Tafa (1995, p. 33). Item-total correlations, however, varied substantially indicating that some items reflected the general content and performance of the task better than others.

However, in passing it was also noted that the difficulty of the items was a function of the length of the key words (the number of the letters of the correct response of each item). For example, the items for which the correct answer was a short word (e.g. But, When, Hid) were generally easier than the items for which the correct answer was a longer word (e.g. Jurisdiction, Direction). The correlation between the number of the letters that made up the correct answer and the difficulty of the item (% correct) for the Greek-Australian sample was -0.47 and -0.32 for the Greek Normative Sample.

*What are the patterns of response to reading items in the Greek-Australian sample?*

In this particular analysis, an average difficulty of zero logits corresponds to item 19 ('Only') where 93 students (35.2%) gave correct responses and 171 gave incorrect responses. Items with positive difficulty estimates (in logits) were more difficult than this item and items with negative difficulty estimates were less difficult. Large positive estimates mean very difficult items and large negative estimates mean very easy items.

| Question Number | Measure - + | Infit Mean Square 0 0.7 1.3 2 | Outfit Mean Square 0 0.7 1.3 2 | Questions |
|---|---|---|---|---|
| 35 | * | : .* : | : . * | Protest |
| 41 | * | : * : | : . * | Founded |
| 22 | * | : .* : | : . *: | Disinterest |
| 20 | * | : .* : | : . *: | Offer |
| 42 | * | : * : | : . *: | Jurisdiction |
| 26 | * | : .* : | : . *: | Valuing |
| 40 | * | : .* : | : .* : | Priceless |
| 29 | * | : * : | : .* : | Direction |
| 15 | * | : * : | : .* : | Inhospitable |
| 39 | * | : * : | : .* : | Distinguish |
| 33 | * | : * : | : .* : | Complied |
| 3 | * | : * : | : .* : | Wagons |
| 16 | * | : * : | : * : | Trustworthy |
| 28 | * | :*. : | : * : | Clear |
| 37 | * | : * : | : * : | Whom |
| 12 | * | : * : | : * : | Proud |
| 1 | * | : * : | : * : | Explorer |
| 30 | * | : * : | : * : | Mankind |
| 8 | * | : * : | : * : | When |
| 21 | * | : * : | : * : | Was suspicious |
| 17 | * | : * : | :*. : | Appearance |
| 31 | * | :*. : | : * : | Protect ourselves |
| 18 | * | :*. : | : * : | Property |
| 10 | * | :*. : | :*. : | Once |
| 24 | * | :*. : | :*. : | Which |
| 25 | * | :*. : | :*. : | Whoever |
| 23 | * | :*. : | :*. : | Totally exhausted |
| 36 | * | :*. : | :*. : | Indecisive |
| 34 | * | :*. : | :*. : | Expression |
| 38 | * | :*. : | :* . : | Rather than |
| 19 | * | :*. : | :*. : | Only |
| 11 | * | :*. : | :* . : | But |
| 13 | * | :*. : | :* . : | Rent |
| 32 | * | :*. : | :* . : | Becoming indignant |
| 27 | * | :* . : | :* . : | Is worthy of |
| 7 | * | :* . : | :* . : | Even if |
| 4 | * | :* . : | * . : | Rained |
| 6 | * | :* . : | :* . : | Reason |
| 2 | * | :* . : | * . : | Respect |
| 14 | * | :* . : | * . : | Promised |
| 5 | * | :* . : | * . : | Earlier |
| 9 | * | :* . : | * . : | Hid |

The asterisk (*) shows the location of the item as well as the infit and outfit value. The colon(:) indicates the 0.7 and 1.3 boundaries.

**Figure 1.** Misfit order

**Table 3**: Item difficulties (percentage correct), item-total correlations and difficulty estimates

| Item/word | Greek-Australian sample of high school students (%) | Normative sample of Greek primary school pupils (%) | Item-total correlation (point biserial) | Ability (i.e. probability of being correct is 0.5) |
|---|---|---|---|---|
| 1. Explorer | 25 | 84 | 0.26 | 0.53 |
| 2. Respect | 59 | 84 | 0.54 | -1.14 |
| 3. Wagons | 27 | 82 | 0.27 | 0.42 |
| 4. Rained | 76 | 88 | 0.46 | -2.06 |
| 5. Earlier | 65 | 88 | 0.56 | -1.47 |
| 6. Reason | 29 | 81 | 0.49 | 0.31 |
| 7. Even if | 30 | 80 | 0.48 | 0.25 |
| 8. When | 65 | 78 | 0.33 | -1.45 |
| 9. Hid | 66 | 82 | 0.56 | -1.49 |
| 10. Once | 54 | 80 | 0.38 | -0.93 |
| 11. But | 72 | 80 | 0.41 | -1.85 |
| 12. Proud | 48 | 74 | 0.30 | -0.64 |
| 13. Rent | 43 | 75 | 0.45 | -0.43 |
| 14. Promised | 37 | 86 | 0.56 | -0.15 |
| 15. Inhospitable | 26 | 68 | 0.22 | 0.46 |
| 16. Trustworthy | 29 | 65 | 0.31 | 0.31 |
| 17. Appearance | 34 | 66 | 0.34 | 0.04 |
| 18. Property | 21 | 64 | 0.35 | 0.77 |
| 19. Only | 34 | 76 | 0.42 | 0 |
| 20. Offer | 21 | 65 | 0.20 | 0.74 |
| 21. Was suspicious | 34 | 74 | 0.34 | 0 |
| 22. Disinterest | 36 | 58 | 0.14 | -0.06 |
| 23. Totally exhausted | 26 | 65 | 0.38 | 0.46 |
| 24. Which | 48 | 62 | 0.38 | -0.64 |
| 25. Whoever | 59 | 71 | 0.38 | -1.16 |
| 26. Valuing | 40 | 53 | 0.19 | -0.27 |
| 27. Is worthy of | 28 | 61 | 0.46 | 0.36 |
| 28. Clear | 8 | 64 | 0.23 | 2.06 |
| 29. Direction | 18 | 52 | 0.21 | 0.98 |
| 30. Mankind | 34 | 58 | 0.33 | -0.02 |
| 31. Protect ourselves | 20 | 67 | 0.33 | 0.82 |
| 32. Becoming indigna | 27 | 56 | 0.45 | 0.4 |
| 33. Complied | 26 | 56 | 0.24 | 0.44 |
| 34. Expression | 35 | 53 | 0.41 | -0.02 |
| 35. Protest | 21 | 47 | 0.08 | 0.79 |
| 36. Indecisive | 27 | 49 | 0.39 | 0.38 |
| 37. Whom | 28 | 54 | 0.30 | 0.36 |
| 38. Rather than | 29 | 65 | 0.41 | 0.27 |
| 39. Distinguish | 19 | 37 | 0.21 | 0.92 |
| 40. Priceless | 40 | 41 | 0.15 | -0.29 |
| 41. Founded | 23 | 41 | 0.20 | 0.65 |
| 42. Jurisdiction | 14 | 36 | 0.19 | 1.35 |

Table 3 (Column 5) indicates the items and their difficulties in logits. The positive logit values represent persons with a greater level of achievement and the more difficult items. The equivalent of reliability in the Rasch model for the items is the Item Separability Index and this was computed to be 0.97. The standard error of the items varied from 0.13 (items 2, 10, 12, 13, 24-26, 40) to 0.23 (item 28) logits.

The most familiar items were 'rained' (-2.06 logits – 'It has not *rained* for more than three months and this is why the water level in the river has fallen greatly') and 'but' (-1.85 logits – 'Yes, we knew it, we guessed it *but* we were not certain that it would occur'). The former may signify an item of information that might be obvious amongst the available choices and the latter may represent a basic word acquired at a fairly early stage in language development and also the most obvious choice for the options available ('instead, but, just, rather'). The most difficult items related to the Greek word 'clear' (+2.06 logits – 'His lecture was very *clear* and to the point') or 'jurisdiction' (+1.35 logits – 'The employee told them to direct their enquiry to the other department because what they were seeking was not his *jurisdiction*'). In both cases these terms while quite common in Greek are unlikely to be used in everyday discussion in Australia.

The students' average estimate was -0.67 logits, indicating that the students were, on average, 0.67 logits less able than required to have a 50% chance to get correct an item of the test with average difficulty. The standard deviation of the students' ability was 0.83 logits, with a maximum ability of 4.03 and a minimum ability of -2.9 logits. Ability of zero logits corresponds (in this analysis) to a score of, approximately, 21 out of 42 items (50% correct). Figure 1 is an item-map that sets out on one calibrated scale student performances and the item difficulties. In this figure the numbers on the extreme left represent the logit scale, ranging from –3.0 (items that require low knowledge) to +3.0 (items that require high knowledge), and these are the values on which the items and persons are calibrated. Zero represents the mean of the item difficulties. Each '#' represents two students. Each '.' The left hand side of the figure represents the distribution of students over the scale; while the right hand side represents the distribution of items plotted according to their value.

The placement of students and items on the same scale allows one to consider how well the different items matched the students' range of knowledge. The chart shows that students did not require much knowledge to answer questions 'But' and 'Rained' but they required an extremely high level of knowledge to be able to answer questions 'Clear' and 'Jurisdiction'. Indeed, about three-quarters of the items were above the general level of knowledge of this cohort of students.

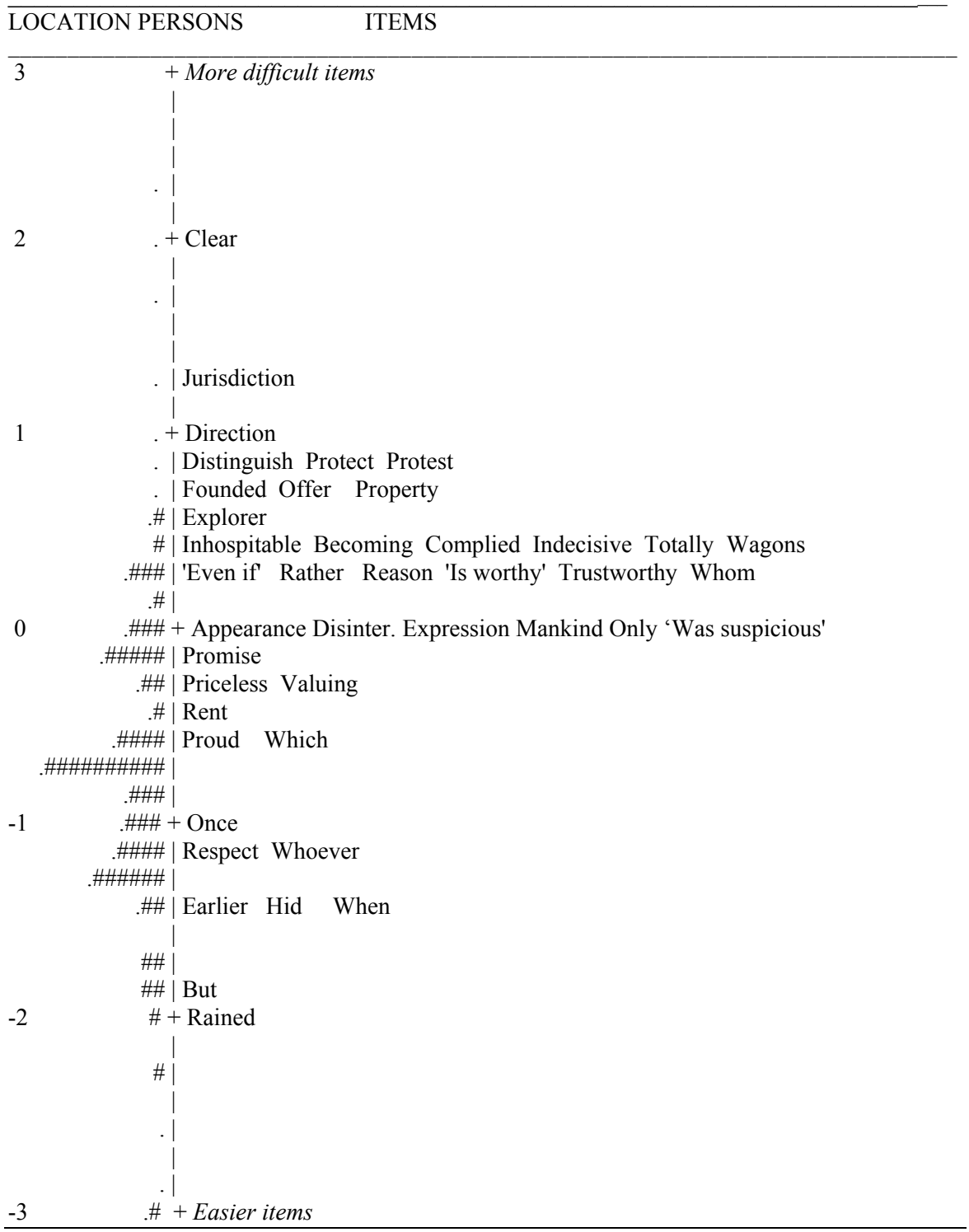
*The effect of background factors*

The effect of background factors were also examined. The increase in ability for those students or their parents who were not born in Australia was not significant. There was, however, a significant effect for the language spoken by the parent. Students whose mother or father spoke Greek had higher levels of ability than those who spoke Greek and English or English alone (Mothers $F(2,247) = 14.1$, $p < .001$; Fathers $F(2,248) = 6.9$, $p < .01$). The mean ability values for mohers are displayed in Figure 3 (a similar pattern also applies for fathers).

## DISCUSSION

The analysis of these responses using a Rasch measurement model provided a helpful means of describing and calibrating students' reading. The results showed relationships between students' overall level of ability and the probability of their answering particular items. In terms of assessing the development of reading there was a substantial need for easier items that match the ability of those 114 students who were low in ability (less than -1 logits).

Ultimately a detailed study of responses might reveal information about the pattern of language acquisition and retention in a multicultural quasi-bilingual context. It was not, however, the purpose of this paper to explore the reasons why second-language acquisition (or more correctly first language retention) affected the pattern of response but one might hypothesise that it mirrors the influence of cultural factors in language acquisition, personal exposure to language, teaching and learning and also the inherent complexities of some words and concepts across languages (see Figure 4 for a tentative model).




Each '#' represents two students; each '.' Represents one student.
The left hand side shows a histogram or bar chart of the distribution of student ability; the right hand side shows the location of the questions in terms of difficulty. The average level is set at zero.

**Figure 2**. Reading ability – item map (measured in logits).


**Figure 3**. Mean differences in ability according to the language spoken by the mother

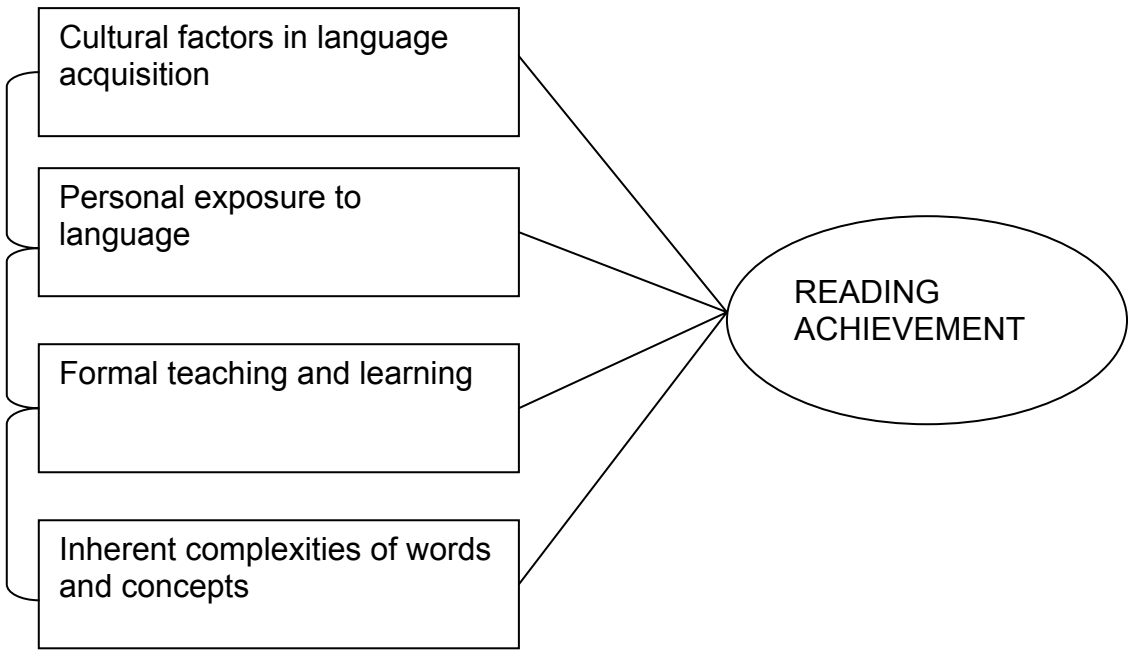



**Figure 4**. Tentative model of reading achievement in an ethnic and now second-language

Using a Rasch model permitted a clearer description of what it means to have a particular level of knowledge. Educators know what specific answers describe a particular score. For example, a logit score of zero means that students are probably able to answer questions 4, 11, 9, 5, 8, 2, 25, 10, 12, 24, 13, 14, 22, 34, 21, 30, 19 and 17. Students with a logit score of zero would probably find the remaining items beyond their ability. The use of a Rasch model also enables results from other forms of assessment to be calibrated on this same logit scale.

The items in the diagnostic reading test provided a unique and meaningful context for an analysis of reading knowledge for Greek-Australian high school students. The results indicated that there was a substantial difference in ability between Greek pupils and Greek-Australians. Simple knowledge of some words was largely missing from the language of these high school students for whom Greek had become a second language. The results also confirmed that there is a possible sequence for the acquisition of achievement and future studies may provide a basis for studying reading development and curriculum reform.

**Acknowledgement**
The helpful comments of an anonymous reviewer are gratefully acknowledged.

## REFERENCES

Athanasou, J.A., Georgoussis, P.N. & Petoumenos, K. (1996). An evaluation of the knowledge and interest of Greek-Australian high school students in the Greek language. [in Greek] *Paidagogikos Logos, 2*, 38-49.

Bezruczko, N. (2004). Raw Score Nonlinearity Obscures Growth. *Rasch Measurement Transactions 18*(2), 973-974.

Bond, T., & Fox, C. (2001). *Applying the Rasch model: fundamental measurement in the human sciences*. New Jersey: Lawrence Erlbaum Associates.

Embretson, S.E. (1996). The new rules of measurement. *Psychological Assessment,* 8, 341-349.

Karabatsos, G. (2000). A critique of Rasch residual fit statistics. *Journal of Applied Measurement*, 1, 152-176.

Linacre, J.M., & Wright, B.D. (2003). *A user's guide to BIGSTEPS Rasch-model computer programs*. Chicago: Winsteps.com.

McNamara, T. (1996). *Measuring second language performance*. Edinburgh Gate, Harlow: Addison Wesley Longman.

Rasch, G. (1960). *Probabilistic models for some intelligence and achievement tests. Copenhagen: Danish Institute for Educational Research. Expanded edition (1980).* Chicago: The University of Chicago Press.

Romanoski, J., & Douglas, G. (2002). Rasch-transformed Raw Scores and Two-Way ANOVA: A Simulation Study. *Journal of Applied Measurement, 3*(4), 421-430.

Smith, R. M. (1991). The distributional properties of Rasch item fit statistics. *Educational and Psychological Measurement*, 51, 541-565.

Smith, R. M. (2000). Fit analysis in Latent Trait measurement models. *Journal of Applied Measurement* 1, 199-218.

Tafa, E (1995). *Diagnostic test of reading ability* [in Greek]. Athens: Ellinika Grammata.

Wright, B. D., & Linacre, M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions 8*(3), 370.

Wright, B. D., & Masters, G. N. (1982). *Rating Scale Analysis. Chicago*: MESA.

Wright, B. D., & Mok, M. (2000). Rasch models overview. *Journal of Applied Measurement*, *1*, 83-106.

Wright, B. D., & Stone, M. H. (1979). *Best Test Design*. Chicago: Mesa Press.