PASAA
Vol. 24
December 1994

Analyzing Primary Level Reading Tests Using The Rasch Model And Classical Method [1]

Lim Tock Keng
National Institute of Education
Nanyang Technological University

Abstract

Reading comprehension tests for Primary 3 and 5 levels, constructed initially as part of a package for pupil profiling at the primary level, were analyzed using the Rasch model and classical item analysis. Items were written to test reading comprehension skills in the following areas: following instructions, locating relevant information sources, following a sequence of ideas, anticipating story development, making inferences and reading for main ideas. The items had been first pilot tested in two above-average schools involved in the profiling project. Preliminary analysis of the tests, using both the Rasch model and classical analysis, showed that most of the items were rather easy for the students. The tests were then administered to a much bigger sample of students from eight average and below-average schools to calibrate the items. This paper reports how the Rasch model was applied to analyze the test items of the combined sample, to explore the fit of the items and its inclusion in the next version of the test. In addition, classical item analysis, providing information on the point biserial of the item options, was used to consider the quality of the discriminators.

Introduction

This paper focuses on the development of the Primary 3 (P3) and Primary 5 (P5) reading comprehension tests based on the identification of comprehension skills of effective reading and on the ability to read accurately and efficiently to understand the text (Greenall & Swan, 1968). The reading comprehension skills tested were selected from a repertoire of reading skills, appropriate for P3 and P5 in our schools. The items were constructed, wherever possible, to incorporate diagnostic features for remediation. As the items were first written as part of a package for pupil profiling at the primary level, they were pilot tested in two above-average schools involved in the project (Lim, 1993b). The Rasch and classical analyses of the items indicated that the items were easy for the students of the above-average schools. The items were then tested on a bigger sample of eight average and below-average schools.

This paper reports the analysis of the items using the combined sample of ten schools. The Rasch model was utilized to explore items that had good fit and that could be included in the next version of the test. Classical analysis (difficulty level and point biserial of items and options) was then used to consider the quality of the discriminators.

[1]Paper presented at the 1994 RELC Regional Seminar, Reading and Writing Research: Implications for Language Teaching. I wish to thank Mrs Siok Hoon Foenander for her assistance in the writing of the test items.

Development of the Test

Reading comprehension skills, identified and listed in Table 1, covered the ideas and vocabulary levels. Items were written to test reading comprehension skills in six main areas: following instructions, locating relevant information source, following sequence of ideas, anticipating story development, reading to make inferences and for main ideas. Items in following instructions tested the students' ability to follow a setof directions. The P3 items focused on understanding and following instructions while the P5 items dealt with understanding simple written instructions based on text information.

Students need to be able to scan through a whole text to look for specific information (Greenal & Swan, 1968). As they should know how to locate relevant information source using word association cues, items on this skill are set for both P3 and P5. Understanding text organization is a required reading skill as students must be able to recognise how sentences are joined together to form paragraphs and how this organization is signalled (Greenal & Swan, 1968). Items to test following sequence of ideas include rearranging the order of a set of scrambled sentences to get a coherent paragraph. The P3 test had a story where students had to arrange a set of eight sentences in proper order. P5 students had to arrange a set of sentences based on a given passage.

Table 1. Reading Comprehension Skills at the P3 and P5 Levels.

Ideas level

- follow simple written instructions
- follow a sequence of information/ideas
- read carefully and sort out competing information
- understand the meaning when it is not stated directly
- anticipate and predict information/ideas
- find and put together several pieces of information to reach a conclusion
- identify the main point/idea
- detect feelings in a piece of writing (tiredness, annoyance)

Vocabulary level

- work out the meanings of words from clues in the passage
- understand less common expressions and meanings
- choose a word/words to sum up a piece of writing
- paraphrase information/meaning

A useful reading comprehension skill, according to Riley (1979) is the ability to read and anticipate development and predict outcomes. Before reading a text, students should subconsciously ask themselves what they know about the subject matter. This makes it easier for them to see what information is new to them as they read the passage. As Goodman (1965) put it so aptly, reading is a psycholinguistic guessing game. The P3 and P5 items encourage students to think about what they are reading as they anticipate the next likely event in the given incident. Items are also written on reading to make inferences as a writer may decide to suggest something indirectly rather than state it directly. Students may need practice, as pointed by Greenal and Swan (1986), in inferring this information, which is essential for correct understanding.

The most important skill in reading comprehension, as shown by Tay (1979), is the ability to recognize the main idea of a reading selection. Students must be able to read for general sense rather than for the meaning of every word and consequently, be able to distinguish between important and unimportant information. Items set included selecting a topic sentence from a paragraph and selecting the best title of a paragraph from a list of titles.

The P3 and P5 items set in the two tests had a range of formats: multiple choice items, alternate choice items and arranging the sequence of items. Stimulus materials used as context for the questions were collected from primary teachers so that students would be familiar with the situations described in the questions. Graphics was utilized to make the testing materials interesting and attractive.

In the multiple-choice questions, distracters were carefully constructed to encourage pupils to examine each alternative carefully before selecting the best answer. As they had to take account of common areas of errors made by students (Munby, 1968), the P3 and P5 teachers in the pilot schools assisted in the collection of common errors in reading comprehension made by their students. These errors were classified and wherever possible, distracters were set to incorporate these errors.

Analysis of the P3 and P5 Tests

The items were first pilot tested in four P3 classes (N = 153) and four P5 classes (N = 140) in two primary schools in the primary profiling project; both schools have above-average students. One school had students from a higher socio-economic background while the other had students from the middle and lower socio-economic backgrounds. The Rasch model was used to carry out a preliminary analysis to explore the properties of the test items. Reading development, an attribute possessed by the student and assumed to be unidimensional, is the latent trait identfied by the Rasch model.

Rasch analysis of the tests were carried out using the program, Quest (Adams & Khoo, 1993). Quest uses the joint or UCON maximum likelihood procedure to estimate both the item and person parameters with a correction factor for the bias. In examining the model fit of the P3 items and the P5 items, the item and case infit and outfit statistics reported in Quest are the weighted and unweighed residual based statistics described by Wright and Masters (1982). Table 2 indicates that the P3 test has an infit mean square of .98 (with SD = .14) and the P5 test has an infit mean square of .99 (with SD = .21); both data sets fit the model. The reliability of estimate is also good: .97 for the P3 test and .96 for the P5 test. This estimate is the Wright and Masters' (1982) item separation reliability for the proportion of the observed estimate variance that is considered "true".

Table 2.

Summary Table of the Item Estimates of the Reading Tests.

P3 Test				P5 Test			
Item Estimates (Thresholds)				Item Estimates (Thresholds)			
All on All (N = 796 L = 41)				All on All (N = 764 L = 45)			
Summary of item Estimates				Summary of item Estimates			
Mean	.00			Mean	.00		
SD	1.32			SD	1.06		
SD (adjusted)	1.32			SD (adjusted)	1.06		
Reliab. of estimate	.97			Reliab. of estimate	.96		
Fit Statistics				Fit Statistics			
Infit Mean Sq.		Outfit Mean Sq.		Infit Mean Sq.		Outfit Mean Sq.	
Mean	.98	Mean	1.06	Mean	.99	Mean	1.00
SD	.14	SD	.33	SD	.21	SD	.38
Infit tOutfit t		Infit tOutfit t					
Mean	-.26	Mean	.47	Mean	-.11	Mean	.33
SD	3.21	SD	3.11	SD	5.59	SD	4.51
0 items with zero scores				0 items with zero scores			
0 items with perfect scoes				0 items with perfect scores			

Each X represents 1 student

Figures 1A (sample for two schools with 153 students) and 1B (sample for ten schools with 796 students) are variable maps for the P3 test showing the person ability and item difficulty, with the logit scale for the calibration of items and cases being plotted vertically on the map. The advantage of Rasch analysis is that person ability level can be placed on the same scale as the item difficulty level. The validity of the test as a whole can be examined with regards to the ability of the distribution of the students. A test would have more psychometric relevance if the distribution of the item difficulties correspond more or less to the distribution of the person ability level.

The P3 map for the sample of two schools in Figure 1A demonstrates that the test is rather easy for the pilot sample (N = 153), well within the ability level of the sample; the distribution of the item difficulty appeared to be a little out of phase with the distribution of the ability level of the students. The test appeared to be measuring precisely only for P3 students in the lower end of the spectrum. With the addition of students from average and below-average schools involved in the testing, the item difficulty distribution for the combined sample in Figure 1 B (N = 796) was more in line with the student ability distribution. What was interesting was that items 35 and 37 (following simple instructions) found to be the most difficult items in the pilot sample (see Figure 1A) proved to be the most difficult items in the combined sample (See Figure 1B). These items will be considered below.

Figure 1A

Person-ability Item-Difficulty Map for P3 Reading Test (N = 153)

Ability Distribution		Item Difficulty Distribution			
5.0		35			
4.0	xx				
	xxxxx				
3.0	xxx				
	xxxxx	37			
	xxxxxxxxx				
	xxxxxxxxxxxxxxxxx				
2.0	xxxxxxxxxxx	21	29		
	xxxxxxx				
	xxxxxxxxxxx	25	27	34	
	xxxxxxxxxxxx				
1.0	xxxxxxx	3	28	31	
	xxxxxxxx				
	xxxxxxx	14	20		
	xxxxxxxxxxxxx	4	23	40	
	xxxxxxx				
	xxxxx				
	xxxxx	7	26	33	
	xxx	6			
.0	xxx	36			
		5			
	xxxx	1	39		
	xxxxx				
		18			
	x	38	41		
	x	11	16		
-1.0		15	22	30	32
	xx				
		12	24		
		9			
		13			
-2.0					
		2	8		
		10	17		
		19			

Each X represents 1 student

Figure 1B
Person-ability Item-Difficulty Map for P3 Reading Test (N = 796)

Ability Distribution		Item Difficulty Distribution				
4.0						
	x					
3.0	x	35				
	x					
	x	37				
	xxx					
2.0	xxxxx	34				
	xxx					
	xx	3				
	xxxx					
	xxxxx	21				
	xx	7	25	29	31	
1.0	xxxxx	6	28			
	xxxxxxx	20	27			
	xxxxxxxx	1	5	23	33	
	xxxxxxxxxx	14				
	xxxxxxxxx					
	xxxxxxxxxxxxxxxxxxxxxx	40				
.0	xxxxxxxxxx	36				
	xxxxxxxxxx	4	26			
	xxxxxxxxxxx	39				
	xxxxxxxxxx					
	xxxxxxxxxxxxx	11	30			
	xxxxxxx	15	18	38	41	
	xxxxxx					
-1.0	xxxxxxxx					
	xxxxxx	16				
	xxxx	17				
	xxxxx	2	9	10	12	13
	xxx	19				
	xxx	8				
-2.0	xxx					
	xxxx					
	xx					
	x					
	x					
-3.0	x					
	x					
	x					
-4.0						

Each X represents 4 students

According to the P5 variable map for the pilot sample of two schools (N = 140) in Figure 2A, the P5 test seemed to be far too easy for the above-average students, when compared with their ability level. The distribution of the item difficulty appeared to be out of phase with the distribution of the ability level; the students appeared to form a cluster at the top right quadrant of the map. In comparison, Figure 2B showed that the item difficulty distribution for the combined sample (N = 796) was in line with the student ability distribution. The addition of students from average and below-average schools in the combined sample helped to calibrate the items to reflect the P5 level of reading comprehension.

P5 items in Figure 2A which proved to be difficult for some of the students in the pilot sample of above-average students included items 45 (reading

for main ideas), items 1, 3, 5, and 6 (understanding simple written instructions). Items which proved difficult for the combined sample in Figure 2B included items 1 and 3 (understanding simple written instructions), item 20 (reading to make inferences) and item 43 (reading for main ideas). These items would be examined in a later section.

Figure 2A
Person-ability Item-Difficulty Map for P5 Reading Test (N = 140)

Ability Distribution	Item Difficulty Distribution
4.0	
	45
x	
	1
3.0 xx	
	5
xxxxxxx	
xxxxxxxxxxxxxx	
	3 6
xxxxxxxxxxxxx	43
2.0 xxxxxxxxxxxx	31
xxxxxxxxxxxxxxxx	
xxxxxxxxxxxxxx	9 20 23
xxxxxxx	
xxxxxxxxxxxxxxx	
x	42
xxxxxxxxxxxx	12 27
1.0 xxxxxxxxx	
x	
xxx	22 40
xxxxx	24 44
xx	39
.0	
x	14 30
	10 16 19
	13
	7 25 34
x	2 8
x	
	17
-1.0	15 26
	11 18
	41
x	29 35 37
	32 33
	4 36
-2.0	
	38
	21
-3.0	28

Each X represents 1 student

Figure 2B
Person-ability Item-Difficulty Map for P3 Reading Test (N = 764)

Ability Distribution	Item Difficulty Distribution				
4.0					
3.0					
	3				
x					
xxxxxx					
	1				
2.0 xxxxxxxxx	20	43			
xxxxxxxxx					
xxxxxxx					
xxxxxxxxxx					
xxxxxxxxxxxx					
xxxxxxxxxxxx					
xxxxxxxxxxxx	23				
1.0 xxxxxxxxxx	5				
xxxxxxxxx	27	31			
xxxxxxxxxxxxxxxxxxxx	9	14	19	42	45
xxxxxxxxxxxx	6	39			
xxxxxxxxxxxx	22				
xxxxxxxxxxxxxxxxxx	13	26	40	44	
.0 xxxxxxxxx	8	12	16		
xxxxxxx	2	17	24		
xxxxxxxxxxxxxxxxx	10	15	18		
xxxxxx	30	34			
xxxxxxxx	25	41			
xxxxxx					
xxxxxxxx	37				
-1.0 xxxx	35				
xxx	4	11	28	36	
xxxxxx	29	33			
x	32	38			
xxx					
xx					
xxxx					
-2.0 x					
	7				
x					
	21				
x					
-3.0					
x					
-4.0					

Each X represents 3 student

The fit of the items is displayed in Figure 3 (P3 item fit map for the combined sample) and Figure 4 (P5 item fit map for the combined sample). The horizontal scale on the fit map is the infit mean square; the asterisks indicate the magnitude of the fit statistic for each item. The items of both test fitted adequately, as most of the infit mean squares of the items in both the P3 and P5 test lie within 30% above and below its expected value. The P3 items in Figure 3 which did not fit well were items 10, 13 and 19 (locating relevant information source). As for the P5 items in Figure 4, items that did not have a good fit included items 5, 6 and 9 (understanding simple written instructions based on text information) item 18 (following sequence of ideas) and item 36 (locating relevant information source). These items would be examined below.

Figure 3
Item Fit Map of the P3 Reading Test (N = 796)

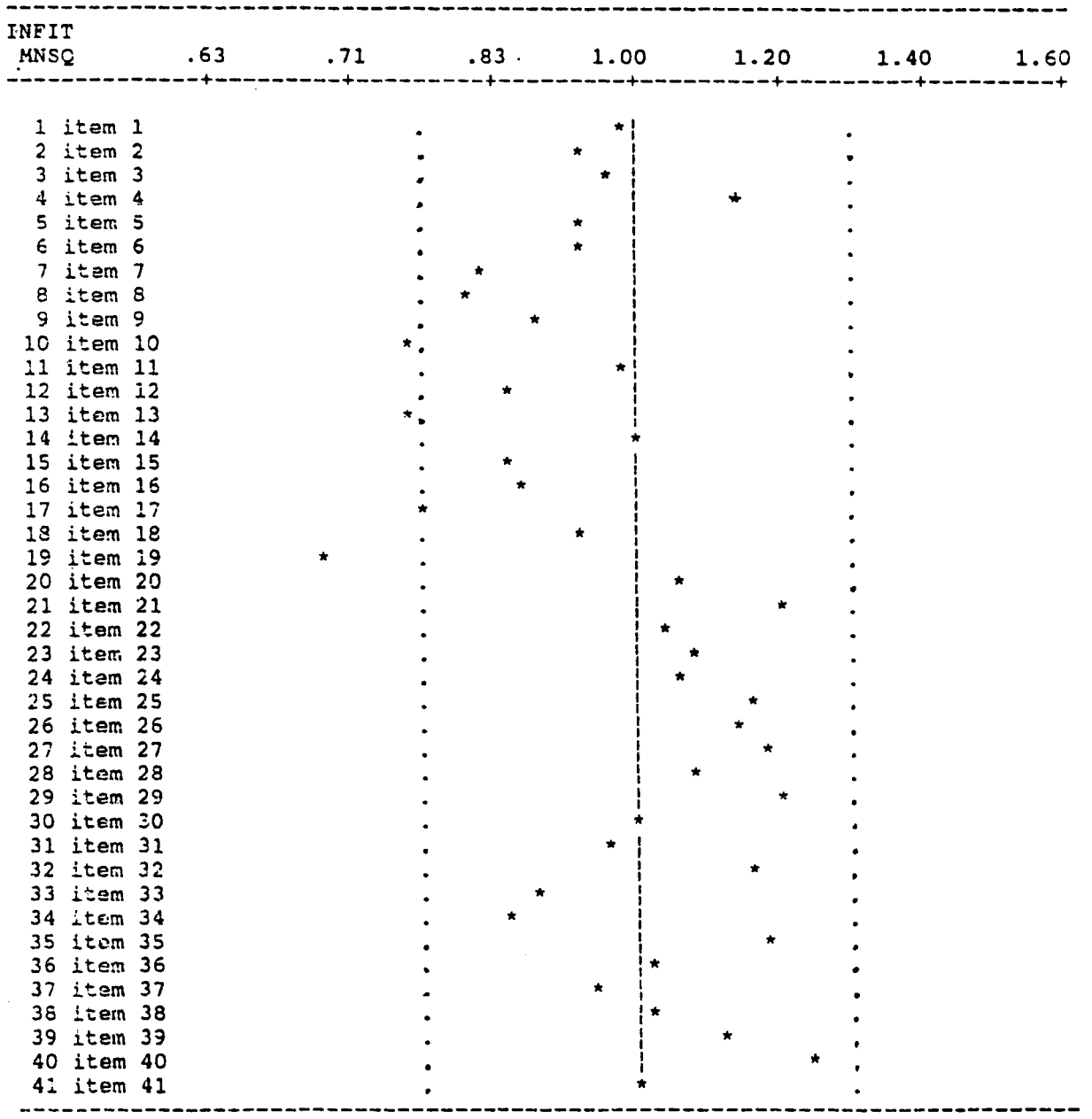
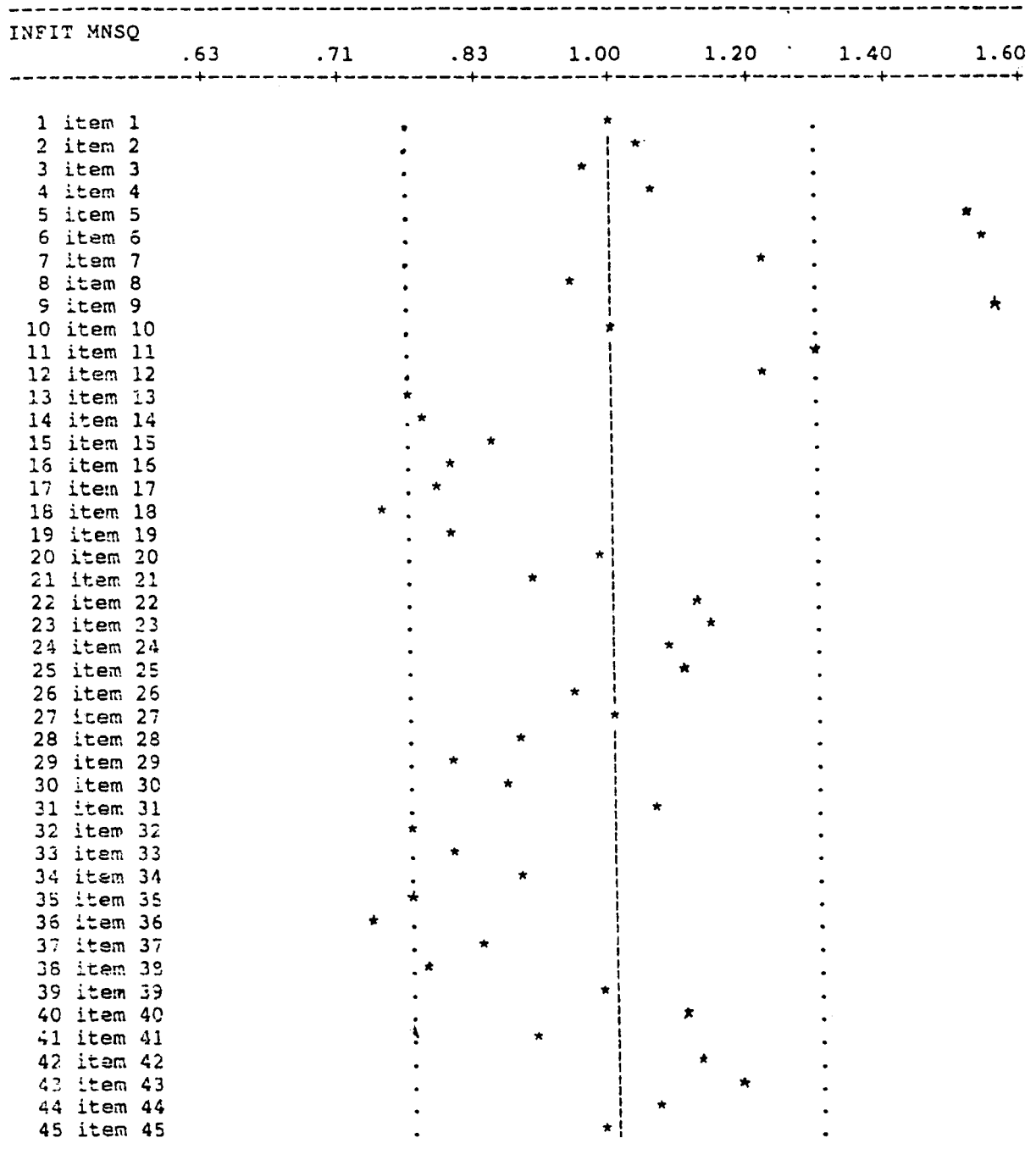


Figure 4
Item Fit Map of the P3 Reading Test (N = 764)



Besides Rasch analysis, classical analysis of the items was carried out using Microcat (Assessment System Corporation, 1989). The program Iteman was used to generate the item difficulty discrimination index and options analysis of the P3 and P5 items to provide teachers with information about the percentages of students making different kinds of errors. The detailed analysis is in Lim (1993a). Special attention was paid to the options analysis of the difficult items and items with poor fit, as indicated by the Rasch model. Information provided by both the Rasch and classical analysis would be used to improve the items of the two tests.

In both the tests, generally the Rasch and classical analyses revealed that the skills of understanding and following instructions appeared to be difficult for some P3 and P5 students. In the difficult P3 items (items 35 and 37), students were required to either underline or circle words that indicate a certain stated fact. The options analysis revealed that more than half the students tended to underline or circle too many words, probably because they were not focused on the key words. In the difficult P5 items (items 1 and 3), students were asked to shade or mark named objects if certain conditions exist in the given passage. The options analysis again showed that P5 students appeared to have a problem with paying attention to both the passage and to what exactly they had to do.

Some P3 students seemed to have difficulties with the misfit items (items 10, 13 and 19) dealing with the location of relevant information source. Six information sources, different kinds of books, were given to the students to locate a certain information. The options analysis of these items showed that the items were not discriminating between the high and low ability students. The answers of students who keyed the wrong options were spread out among the 5 wrong options indicating that students were not able to locate which book they had used to locate the information.

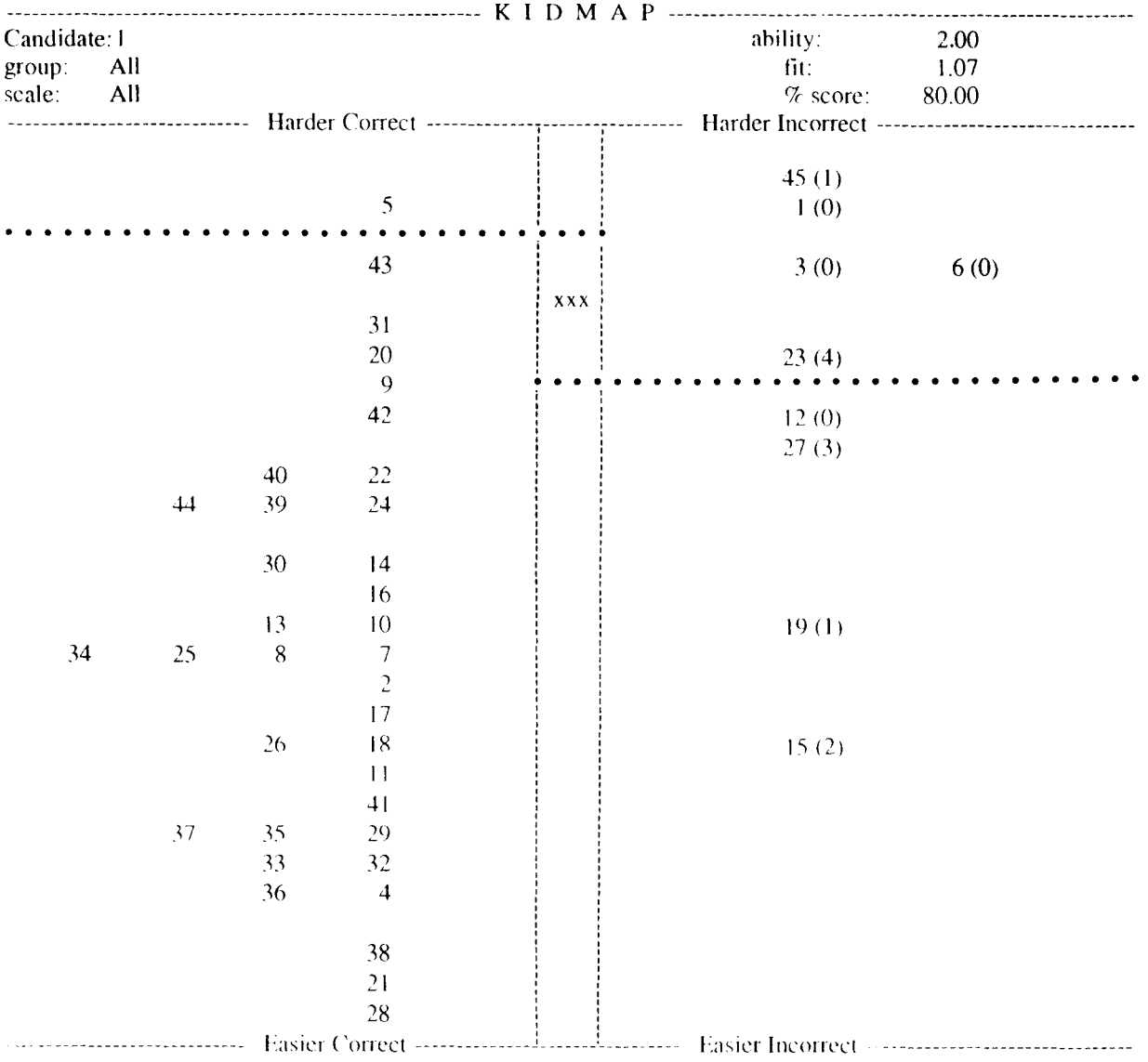
As for P5, some students had difficulties with reading to make inferences and reading for main ideas. Item 20 (reading to make inferences) tested on a common error, whether the students understood what it was to sleep in. The options analysis indicated that many students misunderstood it as going to bed late. In item 43 (reading for main ideas), the options analysis showed that students did not read the short passage properly: it was the girl and not the dog which was lost. Thus for the skill of reading for main ideas, some students failed to distinguish between the main point and supporting details.

As for the P5 items which did not fit well, items 5, 6 and 9 were on understanding simple written instructions. These three items, which required students to shade or mark named objects if certain conditions exist in the given passage, have minus point biserial values, indicating that students who gave the correct answers to these items scored relatively low on the test as a whole. Other items that did not fit well were item 18 (following sequence of ideas) and item 36 (location of relevant information source). The options analysis of the two items also indicated that they were not discriminating between the high and low ability students. Possibly, in the sequence of ideas, the P5 students fail to work out the sequential relationship. In addition, P5 students appeared to have the same problems as P3 students in the location of relevant information source using word association clues.

Feedback to Teachers and Students

An important objective of the study was to provide teachers with information to help them diagnose the strengths and weaknesses of individual students as well as groups of students, so as to provide remediation. The students' total and skill scores were generated for the teachers. In addition, using Rasch model techniques, individual student profiles known as kidmaps provided graphical representations of individuals' estimate and their pattern responses. As illustrated in a kidmap of a P5 student in Figure 5, the kidmap is constructed such that items are plotted in order of difficulty on the left hand side of the profile if the student has answered correctly and on the right hand side if he or she has answered them incorrectly (Adams & Khoo, 1993). As shown by the student's kidmap, when an individual's pattern of responses conform to the model, we expect the majority of items below the individual to be plotted on the left of the figure and the majority of items above the individual to be plotted on the right side of the figure.

Figure 5
A Kidmap of the Primary 5 Reading Test



Teachers were guided as to how to make use of the information provided by Rasch analysis of the items and the kidmaps. To help the teachers understand the kidmaps properly in terms of each student's strengths and weaknesses, skill scores of each student were provided. They were also given information on the item difficulty and item discrimination index and the skills that they could help their students in. The error analysis carried out revealed that the students need to pay proper attention to simple instructions and to key facts. Other errors made included failing to follow relationships of thought and failing to distinguish between main point and supporting details. Teachers could then carry out remediation sessions for the appropriate students.

Conclusion

Preliminary analysis of the tests, both Rasch and classical analysis, showed that most of the items were rather easy for the students from the pilot sample of the two above-average schools. When students from eight average and below-average schools were added to the sample, the Rasch and classical analysis of the items indicated that many of the items were of the correct level for P3 and P5. It also identified the items that needed to be improved. In item calibration, it was important to get a stratified sample of students from good, average and below-average schools.

Items had been written for each of the six skills and analysis of the items might help to develop bands within each of the skills. After the skill bands are developed, it would be possible to put the students in the different bands for each skill. This would further help teachers to provide appropriate remedial lessons for students who need them. There is also a need to select linking items to be used as anchor items for further development of the tests, so that the items would be part of an item bank that can be used for students from primary 3 to primary 5.

References

Adams, R.J. & Khoo, S.T. (1991) *Quest: the interactive test analysis system.* Hawthorn, Victoria: ACER.

Assessment Systems Corportion. (1989). *User's manual for the Microcat.* St Paul, MN: Author.

Department of Education & Science. (1988). *Task group on assessement and testing: A report.* London: HMSO.

Goodman, K. S. (1970). Reading: a psycholinguistic guessing game. In H. Singer & R. Ruddel (E.d.) *Theoretical models and processes of reading,* Newark, Del.: International Reading Association.

Greenal, S. & Swan, M. (1968) *Effective Reading, Teacher's Book,* Cambridge, UK: Cambridge University Press.

Lim, T.K. (1993a). *Analysis of the P3 and P5 Reading Tests of the Primary Pupil Profiling Project: Pilot and Combined Samples,* Unpublished Manuscript, National Institute of Education, Centre for Applied Research in Education, Singapore.

Lim, T. K. (1993b). *Development of Reading Tests at the Primary Level,* Paper presented at the Seventh Intenational Objective Measurement Workshop, Atlanta, April.

Masters, G., Lokan, J., Doig, B., Khoo, S. T., Lindsey, J., Robinson, L., & Zammit, S. (1990) *Profiles of Learning: The basic skills testing program in New South Wales, 1989,* Hawthorn, Victoria: ACER.

Munby, J. (1968). *Read and Think,* Essex, UK: Longman.

Riley, P. (1979) Practising specific language skills. In Guidelines for Teaching Reading Skills, *RELC Journal Supplement,* 2, 12-21.

Tay, M. W. J. (1979). Teaching reading comprehension: a "skills approach." In Guidelines for Teaching Reading Skills, *RELC Journal Supplement,* 2, 7-11.

Wright, B. D. & Masters, G. (1990) *Rating Scale Analysis: Rasch measurement,* Chicago: MESA Press.