2008

Item Banking With Rasch Measurement: an Example for Primary Mathematics in Thailand

Chaowprapha Chuesathuchon
Ubonratchathani Rajabhat University

Russell F. Waugh
Edith Cowan University

Originally published in the Proceedings of the EDU-COM 2008 International Conference. Sustainability in Higher Education: Directions for Change, Edith Cowan University, Perth Western Australia, 19-21 November 2008.

Item Banking With Rasch Measurement: an Example for Primary Mathematics in Thailand

Chaowprapha Chuesathuchon[1] and Russell F. Waugh[2]

[1]Ubonratchathani Rajabhat University, Ubon, Thailand
E-mail: phatom34@hotmail.com

[2]Edith Cowan University, Western Australia
E-mail: r.waugh@ecu.edu.au

ABSTRACT

This study was conducted in Thailand to create a Mathematics item bank and a Computerized Adaptive Test (CAT) for the students to 'interrogate' the bank. First, 290 multiple-choice test items on mathematical equations were created for an item bank. They consisted of nine aspects: (1) identifying an equation; (2) identifying the true equation; (3) identifying equations with an unknown; (4) finding the value of an unknown that satisfies the equation; (5) identifying a method to solve an equation; (6) finding the solutions to equations; (7) finding a solution to an equation related to a given condition; (8) selecting an equation converted from a verbal problem or a verbal problem related to an equation; and (9) solving an equation problem. Seven papers with 50 items each, containing 40 different items and 10 common items, were administered to 3,062 students of Year 6 (Prathom Suksa 6). There were 409, 413, 412, 400, 410, 408, and 610 students taking part in the 1st to the 7th tests respectively. The data were analysed with the Rasch Unidimensional Measurement Model (RUMM 2010) computer program so that all the item difficulties were linked on the same linear scale along with the student measures of mathematical ability. Ninety-eight test items fitted the measurement model and were installed in the item bank. A computer program for CAT was created, tested, and modified after trialling. A controlled experiment involving the use of CAT with 400 Prathom Suksa 6 students from two primary schools in Ubon Ratchathani province, Thailand, was implemented. Thai students were very supportive of the use of CAT with the mathematical item bank. They showed an interest in CAT and in extending the use of CAT to other subject areas with appropriately developed item banks.

INTRODUCTION

Item banks are potentially very helpful for teachers and test developers and they make test-taking easier, faster and more efficient. In the United States, for example, the concept of item banking has been associated with the movements to both individualized instruction and behavioural objectives in the 1960s (Hambleton, 1986; Umar, 1999). Van der Linden (1986, cited in Umar, 1999) viewed item banking as a new practice in test development, as a product of the introduction of Rasch measurement (Rasch, 1960/1980) and the extensive use of computers in modern society. It was suggested that, when a large collection of good items is available to either teachers or test developers, much of the burden of test construction can be removed. The quality of tests used in the schools, for example, could be expected to be better than it would be without an item bank. When a calibrated item bank is developed with Rasch measurement, testing programs can be made more flexible and appropriate, because different groups of students can take different tests which are suitable to each of them and the results can still be compared on the same scale.

What is an Item Bank?

Generally, the words item banks and item pools are used interchangeably in the research literature. Scholars generally identify the term, Item Bank, as a large collection of good test items for which their quality is analysed and known, and which are systematically stored in a computer so that they are accessible to students for measuring their achievement or ability (Choppin, 1981, 1985; Department of Academics, 1991; Millman and Arter, 1984, pp.315-316; Paeratkool, 1975; Rudner, 1998a, 1998b; Wibroonsri, 2005). The items can be stored and retrieved by different aspects, such as subject area, instructional objective measurement, measurement traits, with their accompanying significant item statistics such as item difficulty and discriminating power. Item banks are intended to ease the search and application of various testing procedures in addition to serving the users' needs (Department of Academics, 1991, p.4; Gronlund, 1998, p.130).

Some scholars state that item collection is not only a 'warehouse' or 'storage house' of items but, in a proper item bank, the items are systematically organized through the processes from the start. In a proper item bank, each of the items is codified and classified by subject matter assessed, objectives, and the psychometric traits of the items. The well-selected items are normally stored in the memory unit of the computer so that they can be later easily used when needed (Ebel and Frisbie, 1986, p.927). Ideally, the advancement of item banking could be achieved in that the statistical processes will be applied to differentiate and aggregate the items with the same difficulty level. This contributes to the possibility of the assessment comparison, although the results are gained from different test items (Shoemaker, 1976 cited in Lila, 1996, p.36; Wright and Bell, 1984, p.331).

In a Rasch item bank, each test item is statistically calibrated to be linked on the same interval level scale. This can be easily processed with a specially developed computer program (such as RUMM, Andrich, Sheridan & Luo, 2003) which shows each item of the test that fits the measurement model. The test is flexible, appropriate to Rasch measurement and its implementation, and it is applicable for school use. This has been explained by van der Linden (1986 cited in Umar, 1999, p.209) who viewed item banking as a new practice in test development, as a product of the introduction of Rasch measurement and the extensive use of computers in modern society. The items which cover every aspect of the domains are categorised and stored into the same domain of knowledge or ability. They are also located on a common, linear scale. In the selection of the items for testing, using Computerized Adaptive Testing, a certain statistical value namely difficulty is considered to be appropriate for the ability or competence level of the student. The result of the test even though different items are used can be compared since each of the test items is on a common, calculated linear scale. An item bank at this level could be considered as a model of a 'measurement system'. In this system, any new items intended for measuring the same attribute could be validated and calibrated onto the existing scale of the bank. Since the items are calibrated, it is possible to compare results from tests consisting of different subsets of items from the bank (Hambleton, Sawaminathan, and Rogers, 1991). As such, a calibrated item bank when developed with Rasch measurement makes the testing programs flexible and appropriate, because different groups of students can take different items which are suitable to each of them and the results can still be compared on the same scale. Together with sophisticated computer software, application of Computerized Adaptive Testing could be made possible at the school or district level (Hambleton et al., 1991).

Potential Benefits of Item Banking

It is believed that item banking can potentially bring many advantages to educational assessment. The students could directly benefit from such an evaluation tool since the well-developed test items can potentially accurately measure and compare their true competence or achievement level. There are ten potential benefits of item banking gleaned from the literature.

(1) Teachers can select good test items which meet the measurement objectives and the content from the item bank to suit their students' abilities in each of the area of testing.

(2) Item banking can reduce time spent on the construction of the test items by teachers. This could result in teachers having more time available for the students and their teaching tasks (Umar, 1990).

(3) The items analysed using Rasch measurement will help create a test which contains items located on a common, linear scale and based on a variety of options or objectives (Rudner, 1998a) which in turn contribute to the comparison of the test results of the students who take the different test items, since the Rasch model used will assure items from multiple tests can be placed on a common scale and indicate the relative difficulty of the items (Rudner, 1998a).

(4) Item banking will enable teachers to build a test which contains items located on a common, linear scale and based on a variety of options or objectives by using a Rasch measurement model which is highly effective in item analysis and unidimensionality assessment (Njiru and Romanoski, 2007a, pp.3-4; Rudner, 1998a,b).

(5) Item banking displays the advancement and standards in a school's measurement of student achievement; that is, valid longitudinal achievement inferences can be made from it.

(6) Teachers and measurement experts will be able to easily improve the item bank either by increasing or improving the test items to update them and make them relevant to the changing curriculum, as is required by State Systems, schools and the public at school and national levels (Njiru and Romanoski, 2007a, pp.3-4).

(7) A well-developed item bank enhances effective measurements because the test items can be improved in both validity and reliability to meet educational higher standards (Umar, 1990). This consequently assures the accuracy and reliability of the measurement.

(8) Security is guaranteed because there are a lot of items in the bank. It is unlikely that the students who take the test can remember all of the items from one or several testings. Item banks can therefore protect item leakage, at least to a large extent (Choppin,1981 cited in Millman and Arter, 1984; Umar, 1999, p.210).

(9) Item banking is a product of a new innovation in measurement, namely Rasch measurement coupled with improvements in computing power (Computerized Adaptive Testing), and is easily applied to school state and national educational assessment; each student can complete different test items but the results from the testing can be compared (Umar, 1999).

(10) Item banking potentially allows for the creation of a test which is adaptive to any group of students who have different learning abilities and for students with disabilities (Umar, 1990).

Item Banking in Thailand

In the case of Thailand, the concept of item banking apparently emerged in 1957 and was widely known in 1982-1984 when Thailand was assigned by her neighbouring ASEAN countries to initiate a testing program for the entire ASEAN education region, but its use in any Asian country is still very limited, probably because of the large cost involved in development (Boonprasert, 1988). Throughout the 1982-1984 project, there were several training seminars and further educational seminars, including the proceedings for the meetings. Since then, the Thai Ministry of Education has been very slowly developing item banking with a view to eventually expanding it to the regional and local levels (Department of Academics, 1991, p.5). At the Provincial level, for example, the Item Banking and Examination Online System Chiang Mai Examination Centre was established in Chiang Mai Province in 2007 (Sangphueng and Chooprateep, 2007). The Project of Item Banking Development of Nong Khai Superintendents was established in 1997 (Srisamran, 1997), but these have not been developed to the stage where they can be used by teachers and students in schools on a continual basis. They are still in the developing and trialling stage.

On Thai university campuses, there has been some limited research of item banking such as the Online Test Bank at Sura Nari University of Technology (Chansilp, 2006). The test items in this university were standardized on the basis of Traditional Measurement Theory which can only produce non-linear scores and so it is difficult to see how this item bank project can be useful and it would have been better if the researchers had used Item Response Measurement Theory to create linear measures. Other item bank projects in Thai universities have used Item Response Measurement, but they have used the now discredited so-called 2-parameter model (actually involving three parameters, item difficulty, item discrimination and one parameter of person ability) or the so-called 3-parameter model (actually involving four parameters, item difficulty, item discrimination, a guessing parameter and one parameter of person ability) (see Wright, 1999a for a discussion and discrediting of these models). The best Rasch model to use is the so-called 1-parameter model (actually one parameter of item difficulty and one parameter of person ability) (see Andrich, 1988a, 1988b; Wright, 1999a,b). In Thailand, the 2-parameter and 3-parameter models were used by research students to develop trials of item banks for Mathematics (Maneelek, 1997; Songsang, 2004; Supeesut, 1998, 1999; Tuntavanitch, 2006), English (Phungkham, 1988), and Chemistry (Suwannoi, 1989).

COMPUTERIZED ADAPTIVE TESTING (CAT)

CAT consists of a computer program that allows a student to ‘interrogate’ an item bank (Embreston and Reise, 2000; Weiss, 2004, 19833,1982). The test items are constructed and adapted to the ability level of the individual test-taker and administered using a computer (Beevers, McGuire, Stirling, and Wild, 1995; Lord, 1971, 1980; Nering, 1996; Shermis et al., 1996; Stocking and Swanson, 1998, p.271; Wainer, 1990, 1993; Weiss, 2004, 1983). Examinees do not have to answer exactly the same test items as any other examinees and the number of test items to be answered by different examinees are not equal, they depend on the result of the test items that an examinee chooses to answer (Karnjanawasri, 2002; Lord, 1980; Weiss, 1983; Weiss, 2004).

RASCH ANALYSIS FOR THE ITEM BANK (PRESENT STUDY)

The present study involved an initial analysis with 250 mathematics items involving six tests with 50 items each. For linking the scales, each test contained 10 common items first, and then the six data sets were combined. Responses for the mathematics tests came from 2,452 Prathom Suksa 6 (Grade 6) students in Thailand which were entered into an Excel file, as per the response category codes (zero for wrong and one for right) and then converted to a text file. The data pattern had 254 columns: columns 1-4 were for the ID; columns 5-14 were for 10 answers of common test items; columns 15-54 were for 40 answers of test 1; columns 55-94 were for 40 answers of test 2; columns 95-130 were for 40 answers of test 3; columns 135-174 were for 40 answers of test 4; columns 175-214 were for 40 answers of test 5; and columns 215-254 were for 40 answers of test 6. The data were analysed using the Rasch Unidimensional Measurement Model (RUMM2010, Andrich, Sheridan & Luo, 2003) computer program. The non-performing items of the mathematics test (172 items out of 250) were deleted from the scale, leaving 78 items that fitted the measurement model.

Because 172 items (out of 250) were deleted, as not fitting a Rasch measurement model, only 78 items were stored in the item bank and to improve the bank, a further 50 items were created and analysed. For linking the scales, 10 common items from the 78 set were added to the 50 set for calibration together. Data from 610 students were analysed using the RUMM computer program. Of these 50 items, 30 were deleted as not fitting a Rasch measurement model, leaving 20 good fitting items to be add to the set of 78 items.

In Rasch analysis, the items are designed in a conceptual order by difficulty and this order is tested. The data for the items have to also fit the measurement model in order to create a linear scale and this is tested. The person measures and item difficulties were calibrated on the same scale by the RUMM 2010 program, thus providing the creation of a linear measure of achievement for primary school equations in standard units called logits (log odds of answering positively) (see Figure 1). In Figure 1, the student achievement measures are from low to high on the left hand side and the item difficulties are from easy to hard on the right hand side of the same linear scale.

Rasch Analysis: 78 Item Scale

The final analysis with the RUMM program tested the 78 items (N=2,452) in order to create a linear scale of mathematics achievement from an initial bank of 250 items. The residuals were examined; the residuals being the difference between the expected item score calculated according to the Rasch measurement model and the actual item score of the students. This was converted to a standardized residual score in the computer program. The global item fit residuals and global student fit residuals have a mean near zero and a standard deviation near one, when the data fit the measurement model. In this case, the global item and person fit residuals indicate a satisfactory, but not excellent, fit to the measurement model (see Table 1). The individual probability of fit of items to the measurement model was then checked. Of the 78 items, 71 fitted the measurement model with probability $p>0.04$.

Item Trait Test-of-Fit

The item-trait test of fit examines the consistency of the item difficulties across the student mathematics measures along the scale. This determines whether there was agreement amongst the students as to the difficulties of all items along the scale. The item-trait interaction was not statistically significant at the 0.01 level [Chi-square (df =690) =760.34, p =0.03]. This means that a dominant trait was measured and that overall fit to the measurement is acceptable, but not excellent.

Table 1

Summary of fit statistics for mathematics achievement scale (78 items)

	Items	Students
Number	78	2,452
Location mean	0.00	0.58
Standard deviation	0.62	1.64
Fit statistic mean	0.63	0.08
Fit statistic standard deviation	1.23	0.73
Item-trait interaction chi square = 760.34, df=690, p=0.03		
Person Separation Index =0.83		
Power of test-of fit: Good (based on the Separation Index)		

Notes on Table 1

1. The item means are constrained to zero by the measurement model.
2. When the data fit the model, the fit statistics approximate a distribution with a mean near zero and a standard deviation near one. The item fit and student fit are satisfactory, but neither is an excellent fit.
3. The item-trait interaction indicates the agreement displayed with all the items across all students from different locations on the scale (acceptable for these data). This means that a dominant trait has been measured.
4. The Student Separation Index is the proportion of observed mathematics variance considered true (in this scale, 83% and is acceptable). It tells us that the measures are well separated compared to the errors.
5. Numbers are given to two decimal places because the errors are between 0.11 and 0.14

Targeting

The item difficulties range from –1.3 logits (SE=0.12) to + 1.6 logits (SE=0.14) and the student measures range from –3.4 logits to +4.2 logits. There are some students (34%) whose mathematics abilities are more than +1.6 logits and less than –1.3 logits and hence not 'matched' against an item location on the scale. In Figure 1, there are no items matching persons at either the lowest end (-1.5 to –3.5 logits) or the highest end (+1.5 to +4.4 logits) of the scale, indicating the improvements that are needed for the test. That is, both easy items and hard items need to be added to improve the targeting of the items for these Prathom Suksa 6 students. There are approximately 600 students who found these test items easy and approximately 180 who found who found them hard. The item difficulties were appropriate for the rest of the students, approximately 1,770 students.

```
----------------------------------------------------------------------------
Location          Persons                Item Difficulties
----------------------------------------------------------------------------
  5.0      High Achievement |           Hard items
                            |
                            |
                            |
                         XX |
  4.0                       |
                       XXXX |
                      XXXXX |
               XXXXXXXXXXXX |
                            |
  3.0                     X |
                         XX |
                        XXX |
                       XXXX |
                     XXXXXX |
  2.0               XXXXXXX |
                      XXXXX |
                      XXXXX |
                       XXXX | 34
                 XXXXXXXXXX | 233 97
  1.0                 XXXXX | 164 112
                 XXXXXXXXXX | 134 73 60
                         XX | 75 121 156 71 182 83
      XXXXXXXXXXXXXXXXXXXXX | 140 236 183
                      XXXXX | 76 74 48 106 104 117 185 212 81 223 142 172
  0.0       XXXXXXXXXXXXXXX | 115 129 171 170  47 29 90 224
                 XXXXXXXXXX | 109 141 103 132 249 127 126 197 130 13 68 248
                              100 180
                XXXXXXXXXXX | 92 85 150 125
         XXXXXXXXXXXXXXXXXX | 241 220 218 45 202 157
                   XXXXXXXX | 116 135 200 14 87 122 53 91
  1.0        XXXXXXXXXXXXXX | 16 139 88 51 174 119
                   XXXXXXXX | 138 214
                    XXXXXXX | 59
                    XXXXXXX |
                         XX |
  2.0                    XX |
                        XXX |
                         XX |
                          X |
                            |
  3.0                       |
                            |
                          X |
```

```
                                              |
                                              |
  4.0     Low Achievement     |              Easy items
--------------------------------------------------------------------------
```
Figure 1   Person measures of achievement and item difficulty map for mathematics test (N=2,452, I=78)

Notes on Figure 1 (each X represents 11 students)
1. The scale is in logits, the log odds of answering positively.
2. Measures are ordered from low to high on the LHS and item difficulties are ordered from easy to hard on the RHS.

Category Response Curves

The RUMM program provides a category response curve for each item, which makes it possible to check whether the category responses are being answered consistently and logically. A perusal of the category response curves for the 78 items indicates that the students answered the response categories consistently and logically. The items contained two response categories, 0 for wrong and 1 for correct. Figure 2 shows the category response curve for the item 180, a moderately difficulty item (difficulty = -0.01 logits) that doesn't fit the measurement model as well as one would like. Nevertheless, the Response Category Curve is good showing that the marking for this item is consistent and logical.

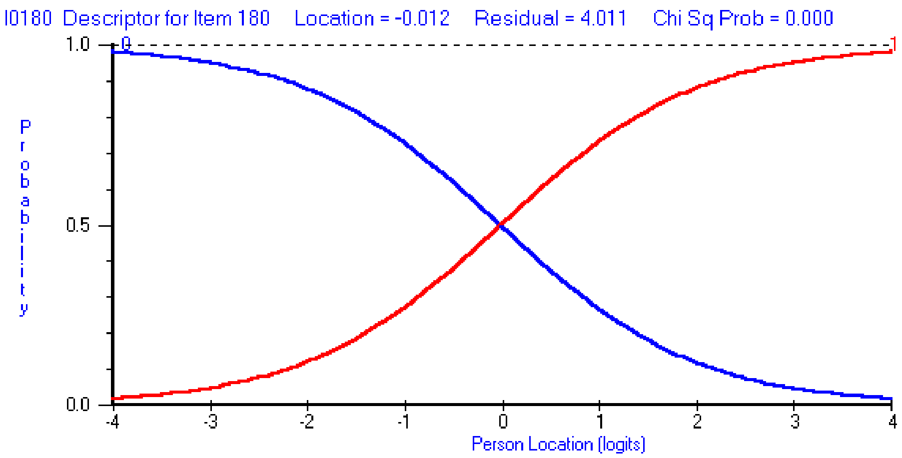

Figure 2: Response category curve for item 180 (not-so-good fitting item)

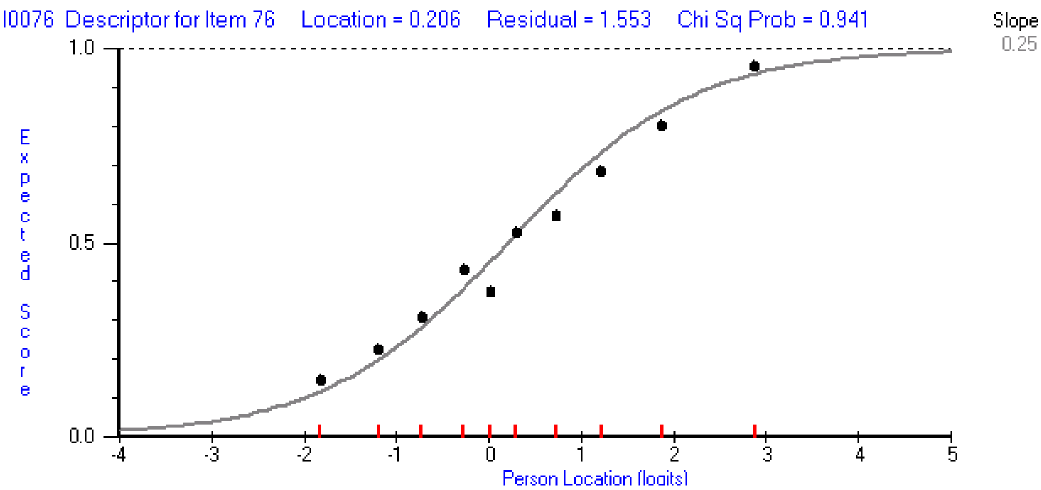

Figure 3: Characteristic curve for item 76 (a Good-Fitting Item)

Item Characteristic Curves

The item characteristic curve for Item 76 (good-fitting item) of the mathematics scale is shown on Figure 3. The line indicates the expected score of mathematics ability groups, ranging from the lowest to highest ability groups, for each observed mean measure (dots) of a student ability group. When the observed scores closely follow the curve of expected values, the group is performing as expected on the item (as shown for item 76).

Item Difficulties

After the Rasch analysis, the items were ordered in terms of their calibrated item difficulties and 'placed' in the bank. Two examples are given in Tables 2 and 3.

Table 2

Item difficulties for finding a method to solve the equations (I= 17, N=2,452)

Item Number	Item Content	Difficulty
1(150)	Finding the method to solve the equation J ÷65 = 130	-0.28
2(109)	Finding the method to solve the equation X ÷ 29 = 174	-0.15
3(141)	Finding the method to solve the equation P + 100 = 200	-0.13
4(103)	Finding the method to solve the equation 96 + L = 386	-0.12
5 (68)	Finding the method to solve the equation 16 × Q = 64	-0.04
6(100)	Finding the method to solve the equation X + 45 = 90	-0.02
7 (29)	Finding the method to solve the equation Z ÷73 = 365	+0.14
8(224)	Finding the method to solve the equation 56 + B = 168	+0.19
9(106)	Finding the method to solve the equation Z × 35 = 140	+0.24
10(104)	Finding the method to solve the equation J – 35 = 105	+0.24
11(185)	Finding the method to solve the equation L – 47 = 188	+0.27
12(223)	Finding the method to solve the equation 80 + F = 240	+0.29
13(142)	Finding the method to solve the equation 75 + D = 375	+0.37
14(140)	Finding the method to solve the equation Y + 40 = 80	+0.47
15(183)	Finding the method to solve the equation 125 + E = 250	+0.54
16(182)	Finding the method to solve the equation X + 61 = 122	+0.73
17 (60)	Finding the method to solve the equation X + 100 = 100	+0.95

Table 3

Item difficulties for finding a solution to an equation (I=9, N=2,452)

Item Number	Item content	Difficulty
1(119)	Find the solution of Q × 24 = 168	-0.82
2(116)	Find the solution of Y + 14 = 140	-0.77
3(200)	Find the solution of 21 + Z = 63	-0.70
4(122)	Find the solution of 25 × F = 25	-0.62
5(241)	Find the solution of 7 + R = 84	-0.54
6(202)	Find the solution of 11 × D = 88	-0.45
7(157)	Find the solution of A – 10 = 100	-0.41
8(197)	Find the solution of M – 38 = 152	-0.06
9(117)	Find the solution of 175 = E – 5	+0.25

The items relating to the identification of the method to solve the equations were found to be ordered from easy (item 150) to very hard (item 60) (see Table 2). Some examples are given now. Item 109

(Find the method to solve the equation X ÷ 29 = 174) and item 103 (Find the method to solve the equation 96 + L = 386) were found to be easy. Item 224 (Find the method to solve the equation 56 + B = 168) and item 104 (Find the method to solve the equation J – 35 = 105) was found to be of moderate difficulty. Item 183 (Find the method to solve the equation 125 + E = 250) and item 182 (Find the method to solve the equation X + 61 = 122) were found to be very difficult.

The items relating to finding the solutions to equations are ordered in difficulty from very easy (item 119) to moderately hard (item 117) (see Table 3). For example, the students found it very easy to find the solutions to the equations Q × 24 = 168 (item 119), Y + 14 = 140 (item 116), and 21 + Z = 63 (item 200), 25 × F = 25 (item 122), 7 + R = 84 (item 241), 11 × D = 88 (item 202), and A – 10 = 100 (item 157). They found it moderately easy to find the solution to the equation M – 38 = 152 (item 197) and they found it moderately hard to find the solution to the equation 175 = E – 5 (item 117).

ITEM BANK CONTENT

The item bank for mathematics on equations for the year 6 (Prathom Suksa 6) students contained 98 items which fitted the measurement model and consisted of:

1. Seven items relating to the identification of an equation, ordered from very easy (difficulty = -0.85) to moderately hard (difficulty = +0.39);
2. Eleven items relating to the identification of the true equation, ordered from very easy (difficulty = -1.07) to very hard (difficulty = +1.57) ;
3. Three items on identifying equations with an unknown, were all very easy (difficulties from -0.96 to -0.66);
4. Eight items on finding the value of an unknown that satisfies the equation, ordered from very easy (difficulty = -1.27) to very hard (difficulty = +1.37);
5. Seventeen items relating to Identify the Method to solve the Equation, ordered from very easy (difficulty = -0.28) to extremely hard (difficulty = +0.95);
6. Twelve items relating to finding the solutions to equations, ordered from very easy (difficulty = -0.82) to moderately hard (difficulty = +0.25);
7. Twenty-three items relating to finding a solution of an equation which related the given condition, ordered from moderately easy (difficulty = -0.23) to very hard (difficulty = +1.01);
8. Nine items on selecting an equation converted from a verbal problem or a verbal problem related to an equation, ordered from very easy (difficulty = -0.86) to hard (difficulty = +0.22);
9. Seven items on problem solving, ordered from very easy (difficulty = -0.66) to moderately hard (difficulty = +0.16).

ATTITUDE TO CAT

A Rasch analysis was conducted to create a linear scale of student attitudes to CAT involving five aspects: (1) Interest in CAT, (2) Confidence in CAT, (3) CAT as Modern and Useful, (4) CAT as Reliable, Fair and Good, and (5) CAT Recommendations. The analysis used 30 items and 400 students. Detailed results are not presented here due to lack of space, but some items were found to be very easy to answer positively. Students were very supportive of the use of CAT (item difficulty = -0.53 logits) and they wanted CAT to be used for other subjects (item difficulty = -0.41 logits).

CONCLUSION

Item banking and CAT are the assessment future. That is, assessment of achievement and ability in school subjects, using item banking and CAT, are increasing in western countries. This trend is likely to occur in Asian countries too. This is because of the increasing use of computers and Rasch measurement modelling in assessment and certification of student achievement. The ‘old’ measurement model of True Score Theory (just adding the scores on a set of questions, not ordered by difficulty, with measures not calibrated on the same scale) is unsuitable for large scale comparisons across schools or states. Computers now allow assessment across schools and they now allow schools to help each other in assessment of achievement by pooling and calibrating items on the same scale.

Implications for Students and Teachers

With regard to teachers, computerized adaptive testing is likely to be accurate in assessing individual student's ability in any tested situation. The teachers can use it with individual students or groups without worrying about cheating in the examinations. Computerized Adaptive Testing could help prevent examinees from getting bored with having too many test items. Also, through Computerized Adaptive Testing, each examinee does different test items and a different number of items. This depends upon an individual's ability. In addition, data gained from the test can be used for many purposes, such as, to follow up an individual's learning progress, to diagnose deficiencies in each student, and to assess students' achievement. Student's weaknesses in any subject matter can consequently be remedied. Computerized Adaptive Testing is an efficient and authentic assessment of student's learning. It is recommended that teachers prepare more examples of item banks and Computerized Adaptive Testing for use in primary schools in Thailand in different subjects.

Implications for Schools and Schools Administrators

In relation to school networking, it would be useful for members of the network (teachers from many different schools) to access the item banks available through Computerized Adaptive Testing. The school network could develop a bank containing tests for different subject areas or different banks for different subject areas. This can be done by establishing one school as the item bank, equipped with a central computer, while other member schools in the network can access the bank through the networking computers in their schools. This can save time and school resources in preparing tests and conducting examinations whenever it is needed. Regarding the development of the test items, teachers in every school network could cooperate to construct, try out, analyse, and select qualified items to store in the item bank. If this process is continuously done, the item bank will become large with thousands of well-calibrated items by difficulty equated on the same scale. The pooling of resources between different schools might be launched by provincial administrators. The provincial administrators could run in-service courses on CAT and item banks, with items appropriate to many school subjects. Moreover, Computerized Adaptive Testing is a new approach for learning assessment and evaluation which is likely to be the future of assessment. There is a large monetary cost to implement this, but it would be well worth it, and probably necessary in the future.

Implications for Academics

Academics are encouraged to learn more about Rasch measurement modelling, item banking and CAT, because the increasing use of these tools will eventually move to universities. Academics will also be involved in the development of item banking and CAT programs at schools and for their own research.

REFERENCES

Andrich D. (1988a) A General Form of Rasch's Extended Logistic Model for Partial Credit Scoring. Applied Measurement in Education 1(4), pp.363-378

Andrich, D. (1988b) Rasch Models for Measurement, Sage University Paper on Quantitative Applications in the Social Sciences, Series Number 07/068, Sage Publications, Thousand Oaks,

Andrich, D., Sheridan, B., and Luo, G. (2003) RUMM2010: A Windows-Based Item Analysis Program Employing Rasch Unidimensional Measurement Models. Perth, Western Australia: RUMM

Laboratory. Beevers, C. E., McGuire, G. R., Stirling, G., and Wild, D. G. (1995). Mathematical Ability Assessed by Computer. Computers & Education, 25(3), pp.123-132.

Boonprasert, U. (1988). The Construction of Item Banking, Chulalongkorn University, Bangkok

Chansilp, S. (2006). Online Test Bank SUT. Retrieved 20/5/2007, from http://library.sut.ac.th/central/HeaderFrame.html

Choppin, B. (1981). Educational Measurement and the Item Bank Model. In C. Lacey and D. Lawton (Eds.), Issues in Evaluation and Accountability. Methuen, London.

Choppin, B. (1985). Principles of Item Banking. Evaluation in Education, 9, pp.87-90.

Department of Academics. (1991). Local Item Bank for Schools, Khurusapha Press, Bangkok

Ebel, R. L., and Frisbie, D. A. (1986). Essentials of Educational Measurement (4th ed.), New Jersey: Prentice-Hall, Englewood Cliffs

Embreston, S. E., and Reise, S. P. (2000). Item Response Theory for Psychologists, Erlbaum, Mahwah, NJ

Gronlund, N. E. (1998). Assessment of Student Achievement (6 ed.), Allyn & Bacon, Boston, MA

Hambleton, R. K. (1986). The Changing Conception of Measurement : A Commentary. Applied Psychological Measurement, 10, pp.415-421.

Hambleton, R. K., Sawaminathan, H., and Rogers, J. H. (1991). Fundamental of Item Response Theory. Sage Publications, Newbury Park, CA

Karnjanawasri, S. (2002). Modern Test Theories (2 ed.), Chulalongkorn University Press, Bangkok.

Lila, S. (1996) The Development of Computerized Item Banking System. Unpublished Doctor of Education, Srinakarinwirote University.

Lord, F. M. (1971). Robins-Monro Procedures for Tailored Testing. Journal of Educational and Psychological Measurement, 31, pp.80-120.

Lord, F. M. (1980). Applications of Item Response Theory to Practical Testing Problems, Erlbaum, Hillsdale, NJ

Maneelek, R. (1997). The Effect of Some Variables on Concurrent Validity and Item Number of Computerized Adaptive Testing. Unpublished Ed.D, Srinakharinwirot, Thailand.

Millman, J. and Arter, J. A. (1984). Issues in Item Banking. Journal of Educational Measurement, 21(4), pp.315-330.

Nering, M. L. (1996). The Effect of Person Misfit in Computerized Adaptive Testing. Unpublished PhD, University of Minnesota.

Njiru, J., and Romanoski, J. (2007a). Development and Calibration of Physics Items to Create an Item Bank Using a Rasch Measurement Model. Paper presented at The International Conference on Learning, Johannesburg, South Africa, June 26-29, 2007.

Njiru, J., and Romanoski, J. (2007b). Development and Calibration of Physics Items to Create an Item Bank Using a Rasch Measurement Model. The International Journal of Learning, 14 (2), pp.19-29.

Owen, R. J. (1969). A Baysian Approach to Tailored Testing (Research Bulletin No. 69-92), Educational Testing Service, Princeton, NJ.

Paeratkool, C. (1975). Measurement Techniques (6 ed.), Wattanapanich, Bangkok.

Phungkham, N. (1988). A Comparison of the Quality of CAT and Classical Testing in English Vocabulary Ability of Mathayom Suksa 3 students. Unpublished Master in Education, Chulalongkorn University, Thailand.

Rasch, G. (1980/1960). Probabilistic Model for Some Intelligence and Attainment tests, The University of Chicago, Chicago, IL (original work published in 1960)

Rudner, L. (1998a). Item Banking (No. EDO-TM-98-04). USA: Eric Clearinghouse on Assessment and Evaluation, Washington ,DC.

Rudner, L. (1998b). Item Banking: Practical Assessment, Research and Evaluation, 6(4). Retrieved June 9, 2007, from http://PAREonline.net/getvn.asp?v=6&n=4 .

Sangphueng, S., and Chooprateep, J. (2007). Item Bank in Examination Online System Chiang Mai Examination Center. Retrieved 20/05/2007, from http;//www.chiangmaiexam.com

Shermis, D., Stemmer, M., and Webb, M. (1996). Computerized Adaptive Skill Assessment in a Statewide Testing Program. Journal of Research on Computing in Education, 29, pp.49-67.

Songsang, K. (2004). The Information Function of Computerized Adaptive Testing. Unpublished Doctor of Education Thesis, Srinakharinwirot University, Thailand.

Srisamran, P. (1997). The Evaluation of the Project of Item Banking Development of the Years 1995-1997, Sakonnakorn, Thailand

Stocking, M. L., and Swanson, L. (1998). Optimal Design of Item Banks for Computerized Adaptive Tests. Applied Psychological Measurement, 22, pp.271-279.

Supeesut, N. (1999). Construction of Tailored Test Package and Test Administering Through Microcomputer in Mathematics for Mathayom Suksa 1 Level. Unpublished Master Degree Thesis, Chiang Mai University, Thailand.

Suwannoi, P. (1989). A Computerized Paramedical Testing in Chemistry for Mattayom Suksa 5 Students. Unpublished Master Degree, Khon Kaen University.

Tuntavanitch, P. (2006). An Estimation of Elementary Students Efficacy in Division Skill Using Computer Programing Method and Usual Testing Method, Faculty of Education, Surin Rajabhat University, Thailand.

Umar, J. (1990). Development of an Examination System Based on Calibrated Item Bank Networks (Unpublished Project report): SIDEC, Stanford University.

Umar, J. (1999). Item Banking. In G. N. Masters & J. P. Keeves (Eds.), Advances in Measurement in Educational Research and Assessment, Pergamon Press, New York van der Linden, W. J. (1999). Computerized Educational Testing. In G. N. Masters & J. P. Keeves (Eds.), Advances in Measurement in Educational Research and Assessment (pp. 138-150). Pergamon Press, New York

van der Linden, W. J. and Pashley, P. J. (2000). Item Selection and Ability Estimation in Adaptive Testing. In W. J. van der Linden & C. A. W. Glass (Eds.), Computerized adaptive testing: Theory and practice (pp. 1-25), Kluwer Academic, Dordrecht, the Netherlands

Wainer, H. (1990). Introduction and History. In H. Wainer, N. J. Dorans, R. Flaugher, B. F. Green, R. J. Mislevy, L. Steinberg and D. Thissen (Eds.), Computerized Adaptive Testing: A Primer (pp. 1-22), Lawrence Erlbaum Associates, Hillsdale, NJ.

Wainer, H. (1993) Some Practical Considerations When Covering a Linearly Administered Test to an Adaptive Format. Journal of Educational Measurement, 12, pp.15-20.

Weiss, D. J. (1982). Improving Measurement Quality and Efficiency with Adaptive Testing. Applied Psychological Measurement, 4, pp.273-285

Weiss, D. J. (Ed.). (1983). New Horizons in Testing: Latent Trait Test Theory and Computerized Adaptive Testing, Academic Press, New York

Weiss, D. J. (2004). Computerized Adaptive Testing for Effective and Efficient Measurement in Counseling and Education. Measurement and Evaluation in Counseling and Development, 37, pp70-84

Wibroonsri, Y. (2005). Measurement and Achievement Test Construction (4th ed.), Chlalongkorn University Press, Bangkok

Wright, B. D. (1999a). Fundamental Measurement of Psychology. In S. E. Embretson & S. L. Hershberger (Eds.), The New Rules of Measurement (pp.65-104), Lawrence Erlbaum Associates, Mahwah, NJ

Wright, B. D. (1999b). Rasch Measurement Models. In G. N. Masters & J. P. Keeves (Eds.), Advances in Measurement in Educational Research and Assessment (pp. 85-97), Elsevier Science, Oxford, UK

Wright, B. D., and Bell, S. R. (1984). Item Banks: What, Why, How. Journal of Educational Measurement, 21, pp.331-345.