

# Application of the Rasch Model and the Bookmark Method in Setting Cut Scores in Mathematics

Ahmad Zamri bin Khairani, Nordin bin Abd Razak, and Hasni Shamsuddin

**Abstract**—Setting cut scores is considered one of the most controversial issues in educational measurement. Not only the procedure is complex, it also relies heavily on the judgments of the panelist. The purpose of this paper is to provide evidence of the adequacy of the Bookmark method in setting cut scores. 15 teachers have been employed as panelist for the purpose. A 50 multiple choice item test in Mathematics has been administered to 588 14-years old students. Three cut scores have been set to distinguish students into different performance levels, namely *basic*, *proficient* and *advanced* cut scores. Result shows that the panelist has high satisfaction of the procedure in terms of practicability and implementation of procedures.

**Index Terms**—Bookmark method, cut scores, mathematics, Rasch Model.

## I. INTRODUCTION

Reporting students' performance is an important aspect in any assessment program. Nevertheless, there are growing concerns over issues of inappropriate procedure in reporting students' performance. One of the important issues is the inappropriate use of setting cut scores. Cut score refers to points on a test scale that are used in separating students into different performance levels [1]. Student who scores more than the cut score is considered as having mastering the requirements whereas student who scores below the cut scores does not. Some assessment programs like credential and licensing provide a single cut score with a pass/fail decision while others such as education provide multiple cut scores to distinguish students into performance levels such as *basic*, *proficient* or *advanced*.

Setting cut scores is always a controversial issue in education. The reason is that schools often adopt arbitrary procedures in setting cut scores. Most schools use students' raw score to evaluate students according to their performance level. For example, a student who gets 80% to 100% mark will receive grade 'A' which implies that he or she has mastered knowledge and skills asked in the test, 65% to 79% is grade B and the level of performance is considered good, 55% to 64% (grade C) as moderate, 45% to 54% (grade D) is considered as weak, while those who get lower than 45% marks earn grade E that indicates failure. Other schools may use different cut scores. The practice is regarded as arbitrary

because no theoretical foundation is employed for setting the cut scores.

Moreover, setting cut scores itself is considered as a complex [2]. Cut scores are set in a process called standard setting. In standard setting participants (called *panelists*) are involved in various tasks. For instance, in the Angoff method which is one of the most widely used standard setting methods before; the panelists need to estimate the probability that a hypothetical *minimally competent candidate* (MCC) will answer a particular item correctly. According to [3], the Angoff method presents an extremely complex cognitive task that panelists are not capable of doing, especially the inexperienced ones. The panelists not only have to think of a *minimally competent candidate* but also the corresponding item difficulty in every performance level. Unable to cope with the task, judges resort to a much simpler task, which is, expressing their own opinion on characteristics of the MCCs [4].

In addition, even though numerous methods have been established for setting cut scores, all methods are considered to be arbitrary as well as politically charged [5]. The setting of cut scores are said to be arbitrary because there are chances of many reasonable cut scores. For instance, in a test, a score of 49, 50 or 51 can all considered as cut scores for passing mark, but test developers can only choose one. The cut scores could always be moved up or down because the choices are matters of judgment. In short, cut scores are to be set but not estimated, therefore, the issue of arbitrariness from standard setting procedure is inevitable.

Thus, the purpose of this study is to provide the adequacy of the Bookmark method in establishing three cut scores namely *basic*, *proficient* and *advanced* cut scores. Adequacy of the procedure is examined based on the practicability, implementation and overall panelists' evaluation.

## II. THE BOOKMARK METHOD

The Bookmark method was developed in 1996 to address problems related to other standard setting methods [6]. Since then it has attract wide range of researchers as well as states in the United States to adopt the method for their assessment program. A survey by [7] finds that the Bookmark method is used by at least 31 states since 2003. The Bookmark method employs item mapping procedure, where all items are presented to participant in a rank order format using Ordered Item Booklet (OIB). The OIB contains a set of items ordered by difficulties, from the easiest to the hardest. The item difficulty is established by item response theory (IRT) calibrations. The difference between student's ability and item difficulty will determine the probability of getting a correct answer. In the Bookmark method, panelists are

Manuscript received August 25, 2013; revised November 15, 2013. This work was supported in part by the Universiti Sains Malaysia Short Term Grant 304/PGURU/631148.

Ahmad Zamri bin Khairani and Nordin bin Abd Razak are with the School of Educational Studies, 11800 Universiti Sains Malaysia (e-mail: ahmadzamrir@usm.my, norazak@usm.my).

Hasni Shamsudin is with Sekolah Menengah Sains Kepala Batas, 13200 Kepala Batas, Penang (email: emel\_hasni@yahoo.com).

provided with probability (such as .50) and are asked to determine whether a MCC has this probability of answering each item correctly. Using IRT model, relationship between student’s ability and item’s difficulty, the Bookmark method helps to determine an ability level that represent minimal competency in terms of item difficulty. The advantages offered by IRT calibration provide significant advantages compared to estimation by panelists [8], [9].

The Bookmark method typically proceeds in three rounds. Round 1 includes the orientation and training the panelists. At the initial stage, panelists review the OIB and discuss the knowledge, skills and abilities perform successfully on every item [10]. It is also important to clarify why an item is more difficult than the previous one. At the end of Round 1, panelists place bookmark on the OIB. The bookmark represents judgment of the division between items that the minimally qualified student should master from those items that are not. For selected response item (such as the multiple choice items), “mastery” is defined as having at least the probability of answering the items correctly [11]. For constructed response items (such as partial credit item), “mastery” is defined as having at least a probability of receiving a particular score point or higher. A particular cut score is determined by obtaining the scale value corresponding to a probability of answering the item correctly. The choice of probability particular likelihood used is referred to response probability (RP). Several RP have been suggested, but the two most widely used are 0.67 [12] and 0.50 [13]. RP of .67 can be interpreted as follows: “For a given cut score, a student with the test score at that point will have a .67 probability of answering an item at the given cut score correctly”.

In Round 2, panelists work in small groups (5 to 6). They are provided with feedback from Round 1 results with attention focused on the range of the Bookmark. The group discuss the skills and content that student should master for placement in every performance levels within the bookmark range. At the end of Round 2, if the panelists would like to change their bookmark due to inputs from the discussion, they can place their second set of bookmark. If not, they will put the bookmark on the same place on the OIB together with the previous bookmark. Median of the second set of bookmark for all panelists is determined. Median is preferable than mean to provide group’s statistics because of it ability to deal with extreme scores, which is expected from panelists that have different backgrounds.

Round 3 begins with presentation of impact data to large group. One of the main feedbacks from the impact data is the percentage of students who fall into (or below) each performance levels [14]. To promote further consensus, discussion will involve large groups involving all panelists. At the end of Round 3, judges make their final bookmark placements. The final bookmarks are tabulated for each performance levels, medians are calculated, and the final cut scores and impact data estimates are presented to all panelists. It should be noted that the bookmarks are not cut scores. The cut scores are referred to student ability on the measured scale based on the IRT calibration used. This procedure will be explained later in the methodology section.

### III. METHODOLOGY

The present study uses a 50 multiple choice items based on the Malaysian Mathematics curriculum specifications [15]. The test was administered to 588 14 years-old students. Data analysis was conducted using the Rasch Model. Methodology of both the development and analysis of the test is available in the present journal at volume 2 (3) and published in May 2012 [16] and will not be covered in this paper. The following Table I shows statistics (in *logits*) of each item based on the Rasch Model analysis. Item 29 (−1.65 *logits*) is the easiest item and is placed at the first page of the OIB while item 50 (2.18 *logits*) is the hardest and placed last.

TABLE I: ITEMS STATISTICS

Item	Measure	SE	Infit MNSQ	Outfit MNSQ
50	2.18	.15	.99	1.00
5	2.00	.14	.95	.94
4	1.77	.13	1.00	.96
48	1.71	.13	1.14	1.29
14	1.44	.12	.97	1.01
37	1.37	.12	.98	1.08
38	1.22	.11	.97	1.09
20	1.09	.11	1.19	1.30
6	.90	.11	1.09	1.15
8	.67	.10	1.15	1.23
19	.63	.10	1.03	1.09
31	.56	.10	1.03	1.07
42	.33	.10	1.06	1.12
43	.30	.10	1.02	1.02
41	.29	.10	1.13	1.16
30	.26	.10	1.09	1.10
45	.23	.10	1.15	1.28
10	.10	.09	1.19	1.24
36	.08	.10	1.06	1.09
23	.07	.10	.86	.84
18	.02	.09	1.16	1.28
39	.02	.10	.98	.96
40	−.05	.09	.97	.98
34	−.08	.09	1.03	1.02
32	−.09	.09	.95	.95
25	−.17	.09	1.11	1.13
9	−.19	.09	.93	.90
22	−.20	.09	.88	.87
12	−.21	.09	.99	1.02
7	−.26	.09	.98	.94
27	−.27	.09	.93	.96
46	−.32	.09	1.13	1.18
35	−.38	.09	1.01	1.06
26	−.52	.09	1.02	1.01
33	−.53	.09	.97	.94
15	−.59	.10	.92	.92
17	−.60	.10	1.05	1.10
16	−.61	.10	.85	.79
24	−.64	.10	.84	.77
11	−.70	.10	1.09	1.19
49	−.71	.10	.81	.76
21	−.80	.10	.81	.72
2	−.82	.10	.90	.87
47	−.86	.10	1.07	1.21
44	−.92	.10	.93	.91
13	−.98	.10	.89	.82
1	−1.13	.10	.95	.87
3	−1.41	.10	.86	.74
28	−1.54	.11	.89	.79
29	−1.65	.11	.90	.82

In the standard setting procedure, 15 teachers with 1 to 25

years of experience between them are employed as panelists. The choice of teachers with different experiences is important to provide wide range of backgrounds since it will influence the cut scores [17]. The task of the panelist is to provide cut scores that will differentiate student into different performance levels. Because of different background, it is important to provide the panelists with a framework that provide descriptions of what students at each performance levels know and able to do. This framework is called performance level descriptors (PLD) and is given in Table II. Since the purpose of study is to established three cut scores for *basic*, *proficient* and *advanced* level, four performance levels will be established based on the exercise, namely, *below basic*, *basic*, *proficient*, and *advanced* levels.

TABLE II: PERFORMANCE LEVEL DESCRIPTORS

Performance Level	Performance Level Descriptions
Basic	1. Have basic fundamental knowledge of concepts and skills
	2. Able to solve questions that require direct application of concepts and procedures
	3. Made minimal and ineffective Mathematics communication
Proficient	1. General understanding of basic concepts and skills
	2. Able to select strategies consistently to solve questions
	3. Able to apply concepts and skills in routine questions
	4. Effective communication
Advanced	1. General understanding of basic concepts and skills
	2. Able to select strategies consistently to solve questions
	3. Able to apply concepts and skills in routine questions
	4. Effective communication

Besides PLD, the panelists are also presented with the OIB. Originally, each page of the OIB contains the item and statistics such as its difficulty. Nevertheless, since all panelists are not familiar with the OIB, additional information such as 1) topic, 2) learning outcome, and 3) learning objective are also included. This is to assist the panelists to get more information about the items, thus provide better ratings. Example of a page in the OIB is provided in the following Fig. 1.

Item No	2	Raw Score	351/566	Difficulty	-0.82 logits
Topic	Directed Numbers				
Learning Objective	Perform computations involving combined operations of addition, subtraction, multiplication, and division of integers to solve problems				
Learning Outcome	Students will be able to solve problems involving combined operations of addition, subtraction, multiplication, and division of integers including the use of brackets				
In a school, every student was given -8 points for each offence committed and +15 points for each good behavior shown. If a student commits 7 offences and shows 12 good behaviors, calculate the total number of points collected					
A	-7	C*	124		
B	-84	D	236		

Fig. 1. Example of a page in OIB.

The first piece of information was the item’s position in the test (it was item number 2 in the test), followed by the raw score, which was the sum of the scored responses to an item

by the students (351 students successfully answered this question). Next, was estimation of the difficulty of item 2 (-0.82 logits) followed by the topic. The Learning Objective and Learning Outcome sections were added specifically for this standard setting study in order to facilitate panelists in aligning the set of items with the curriculum specifications. An asterisk (\*) indicates the correct answer. In reviewing the OIB, it was important for each panelist to understand what made each item progressively more difficult than the previous items in the booklet. In essence, each panelist needed to know the knowledge, skills, and abilities for each of the performance level because they were the critical aspects in setting the cut scores.

The standard setting process is conducted in three rounds. In short, at Round 1 panelists placed their bookmark at a particular page based on their judgments. At Round 2, in a group of 5, they discussed their ratings. The discussion is mainly on the different of pages they put their bookmark. Median bookmark for the group is calculated by the leader. After the discussion the panelists placed their second set of bookmark for each performance levels. At Round 3, the median for each three groups are presented. The panelist then discuss whether they agree or disagree with the bookmark. At the end of Round 3, the panelists placed their third and final bookmark. During the standard setting process, the panelists is constantly remind to provide their rating individually based on their discussions in Round 2 and Round 3.

In order to establish the cut scores using the item mapping procedure, the standard setting process employs RP of 0.67. The value is considered easier and more responsible for setting cut scores [18]. The procedure of setting *basic* cut score is given on the following example. The median of the *basic* bookmark is placed at page 14 which correspond with item Q17 with difficulty,  $\delta = -0.60$  logits. Both the value of  $RP = 0.67$  and  $\delta = -0.60$  are inserted into the Rasch Model equation of

$$P_{ni} = \frac{\exp(\beta_n - \delta_i)}{1 + \exp(\beta_n - \delta_i)} \tag{1}$$

Thus, it becomes

$$0.67 = \frac{\exp[\beta - (-0.60)]}{1 + \exp[\beta - (-0.60)]} \tag{2}$$

By solving the equation (2), the ability of MCC at *basic* cut score,  $\beta$  is -0.098 logits. The same procedure is repeated for establishing *proficient* and *advanced* cut scores.

Besides establishing cut scores, the panelists are also asked to provide their response on the adequacy of the Bookmark method. The adequacy is examined with regards to three aspects namely practicability, implementation and overall panelists’ evaluation [19]. Practicability includes whether the method is easy to implement. Implementation of procedures deals with how confident the panelists are with the overall procedure of standard setting. Panelists’ feedback, meanwhile, refers to the extent to which judges feel comfortable with the process, the performance level descriptors as well as the cut scores. Panelists were given statements such as, “I felt that this procedure was fair”. The

panelists were also ask for feedbacks in a given rating scale such as “How do you rate the definition of mastery” to which they were asked to respond “very difficult”, “somewhat difficult”, “somewhat easy” or “very easy”. Feedback from judges is essential as it is an important part of establishing validity of the established cut scores.

#### IV. FINDINGS

The summary for Round 3 ratings is given in the following Table III. Based on the PLD, median bookmark placing for *basic* cut score is at page 14 (Item Q17, difficulty =  $-0.60$  logits). One panelist rate that in her opinion, *basic* students can have .67 probability of getting correct answer for easier item, that is Q49 (difficulty =  $-0.71$  logits) at page 9 of the OIB. This is the minimum rating. Meanwhile, two panelists rate for page 22 (Q22, difficulty =  $-0.20$  logits), which is the maximum rating. Using the same interpretation, bookmarks for both *proficient* and *advanced* levels are placed at page 32 (Q10, difficulty =  $-0.20$  logits) and 45 (Q14, difficulty =  $1.44$  logits) respectively in the OIB. Based on the standard deviation, it shows that more variations for *proficient* ratings compared with *basic* or *advanced*. It can also be seen that the panelists have different ratings at every performance levels and some of the ratings are extreme. Thus, the use of median over mean as statistics during the standard setting process is justified.

TABLE III: SUMMARY OF ROUND 3 RATINGS

Statistics	Rating (Page on the OIB)		
	Basic	Proficient	Advanced
Median	14	32	45
Minimum	9	24	39
Maximum	22	47	50
Std Deviation	3.17	4.94	2.83

Using the item mapping procedure mentioned in the methodology section before, the correspondent ability for every performance levels are given as follows: *Basic* =  $-0.098$  logits, *proficient* =  $0.79$  logits and *advanced* =  $1.93$  logits. Nevertheless, both the researcher and leaders felt that that cut scores are too high. For example, if the value of  $-0.098$  logits is used as cut scores, then more than half of the students (53.57%) fall under *below basic* level. This is certainly not the true picture of the students' Mathematics ability at the schools. Based on the discussion with the leaders, the cut scores are later adjusted and lowered to 1 standard error of measurement based on suggestion by [20]. The final cut scores and percentage of students at every level are presented in the following Table IV.

TABLE IV: ADJUSTED CUT SCORES

Statistics	Adjusted Cut Scores (logits)			
	Below Basic	Basic	Proficient	Advanced
		$-0.034$	$0.37$	$1.50$
% of students	48.99	13.94	25.85	11.22

Adequacy of the Bookmark method is assessed with regards to three criteria, namely, practicability, implementation of procedures as well as panelists' judgments. In terms of practicability, all panelists agree that the procedure is practical. This is not unexpected since their tasks

were further eased since the OIB had been prepared with the inclusion of knowledge, skills, and abilities for every item that was in line with the curriculum specifications. Furthermore, the panelists were not involved in identifying and describing item descriptions that served as the concise performance level descriptors. This task was completed by the researcher through discussion with group leaders. Hence, the panelists' task was narrowed down to conceptualize the PLDs into items in the booklet.

The panelists also express satisfaction in the implementation of procedure. 96% of the panelists agree that the goal of the procedure, which is to set cut scores for *basic*, *proficient*, and *advanced* performance levels is clear. The same percentages of panelists also agree that the training materials are useful to help them during the standard setting process. Nevertheless, one important lesson learned in this study is that the panelists are not very comfortable with the concept of mastery. Most of them (77%) felt that the concept is difficult. Nevertheless, the present takes several steps to ensure the quality of ratings. Firstly, the Bookmark method allowed panelists to have 3 rounds of rating. Inconsistent panelists were given an opportunity to revise their ratings or explain the basis for their rating during group discussion. However, panelists were not required to revise their ratings if they were comfortable with them. Secondly, between Round 2 and Round 3, panelists were encouraged to discuss their ratings in order to promote consistency in the data. Impact data were also presented in Round 3 in order for panelists to see the consequences of their rating. If the impact data was not consistent with panelists' expectation, they may want to raise or lower the cut scores. All these measures should be able to enhance the quality of standard setting judgments.

The overall panelists' evaluation is also an indication of the adequacy of the Bookmark method in this present study. All panelists felt that they understood how to place the bookmarks. They also reported that the training made their task easier. A majority of panelists (96%) agreed that the training materials were helpful as well as that the procedure was well described (89%). In addition, the panelists are also understand and comfortable with the tasks given. For example, 96% of panelists reported that they placed their Round 1 ratings without consulting others. Also, most of the panelists (92%) indicated that they improved their rating as the standard setting process continues. 89% of the panelists agreed that the time allocated was enough. They also expressed satisfaction that their opinions were considered important. All of the panelists also felt comfortable during their discussion.

Another indication of the success of the standard setting study may be found in panelists' perceived validity of the standard setting procedure. A majority of the panelists (94%) reported that the procedure produced valid standards. Most of them (95%) also expressed their satisfaction on their group's final ratings. Therefore there was substantial evidence to conclude that the Bookmark standard setting study was carried out successfully.

#### V. CONCLUSION

Its limitation notwithstanding, the present study has

employed some degree of sound procedure; where various checks on the consistency, reasonableness as well as explicitness of the study have been done, thus encouraging the researcher to have faith in the results. Nevertheless, based on the literature on standard setting study, several other important issues need to be address. Firstly, study must be conducted to examine the extent of classification error, where a particular student is placed at a lower performance level than he or she supposes to be (and vice versa). Secondly, since the panelists' ratings take centre stage of the study, more study need to be conducted to provide evidence of consistency of the panelists, especially both the intra-panelist and inter-panelist consistency. Finally, results from the Bookmark method need to be compared with results from different standard setting method as a measure of external validity.

#### REFERENCES

- [1] R. K. Hambleton, "Setting Performance Standards on Educational Assessments and Criteria for Evaluating the Process," in *Setting Performance Standards: Concepts, Methods, and Perspectives*, G. J. Cizek, Ed., Mahwah, NJ: Lawrence Erlbaum, 2001, ch. 4, pp. 89–116.
- [2] G. J. Cizek, "Conjecture on the Rise and fall of Standard Setting: An Introduction to Context and Practice," in *Setting Performance Standards: Concepts, Methods, and Perspectives*, G. J. Cizek, Ed., Mahwah, NJ: Lawrence Erlbaum, 2001, ch. 1, pp. 3–18.
- [3] L. A. Shepard, "Implications for standard setting of the national academy of education evaluation of the national assessment of educational progress achievement levels," in *Proc. Joint Conference on Standard Setting for Large-Scale Assessment*, Washington, DCY, 1994.
- [4] M. J. Zieky, "So Much Has Changed: How the Setting of Cut Scores has Evolved since the 1980s," in *Setting Performance Standards: Concepts, Methods, and Perspectives*, G. J. Cizek, Ed., Mahwah, NJ: Lawrence Erlbaum, 2001, ch. 2, pp. 19–52.
- [5] S. A. Livingstone and M. J. Zieky, "A comparative study of standard setting method," *Applied Measurement in Education*, vol. 2, no. 2, pp. 121-141, 1989.
- [6] C. J. Cizek, "Standard setting guidelines," *Educational Measurement: Issue and Practice*, vol. 15, no. 1, pp. 13-21, 1996.
- [7] Wisconsin Department of Public Instruction. (2003). On The Bookmark standard setting procedure. [Online]. Available: <http://www.dpi.wi.us/oea/ctbbmrk03.html>.
- [8] K. R. Ricker. (2003). On Setting cut scores: Critical review of Angoff and Modified-Angoff Method. [Online]. Available: <http://www.education.ualberta.ca/educ/sych/crame/RickerCSE2003.pdf>.
- [9] J. C. Impara and B. S. Plake, "Standard setting: An alternate approach," *Journal of Educational Measurement*, vol. 34, no. 4, pp. 353–366, 1998.
- [10] A. Karantonis and S. G. Sirecci, "The Bookmark standard-setting method: A literature review," *Educational measurement: Issues and practice*, vol. 25, no. 1, pp. 4–12, 2006.
- [11] G. J. Cizek, M. B. Bunch, and H. Koons, "Setting performance standards: Contemporary methods," *Educational Measurement: Issue and Practice*, vol. 23, no. 4, pp. 31–50, 2004.
- [12] H. Hyunh, "On score locations of binary and partial credit items and their application to item mapping and criterion referenced interpretation," *Journal of Educational and Behavioral Statistics*, vol. 23, pp. 3–56, 1998.
- [13] N. Wang, "Use of the Rasch IRT model in standard setting: An item mapping method," *Journal of Educational Measurement*, vol. 40, pp. 427–446, 2003.
- [14] L. Jie. (2003). The bookmark standard setting procedure: Strength and weakness. [Online]. Available: <http://www.education.ualberta.ca/educ/sych/crame/files/standard-setting.pdf>
- [15] Curriculum Development Centre, "Sukatan pelajaran kurikulum bersepadu sekolah menengah," *Integrated Curriculum for Secondary Schools*, 2000.
- [16] A. Z. Khairani and N. Abd. Razak, "Advance in educational measurement: A Rasch Model analysis of Mathematics Proficiency Test," *International Journal of Social Science and Humanity*, vol. 2, no. 3, March 2012.
- [17] J. J. Norcini and J. A. Shea, "The credibility and comparability of standards," *Applied Measurement in Education*, vol. 10, no. 1, pp. 39-59, 1997.
- [18] N. J. Williams and E. M. Schulz, "An investigation of response probability (RP) values used in standard setting," in *Proc. National Council on Measurement in Education*, Montreal, Canada, 2005.
- [19] M. T. Kane, "So Much Remains the Same: Conception and Status of Validation in Standard Setting," in *Setting Performance Standards: Concepts, Methods, and Perspectives*, G. J. Cizek, Ed., Mahwah, NJ: Lawrence Erlbaum, 2001, ch. 3, pp. 53–88.
- [20] R. K. Hambleton and M. J. Pitoniak, "Setting Performance Standards," in *Educational Measurement*, 4<sup>th</sup> ed., L. Brennan, Ed., Westport, CT: American Council on Education and Prager Publishers, 2006, ch. 8, pp. 355–366.



**Ahmad Zamri bin Khairani** was born in Parit, Perak on 2<sup>nd</sup> of April 1971. He received his undergraduate and Master in Education from the Universiti Sains Malaysia, Penang in 1995 and 2001 respectively. In 2010 he completed his PhD from the International Islamic University Malaysia Kuala Lumpur. The area of interest is psychometric and educational testing.

He is a former mathematics teacher and now holds a position of senior lecturer at the School of Educational Studies in Universiti Sains Malaysia, Penang. He has published several articles in international journal, mainly in Rasch Model analysis as well as articles related to assessment program. He is currently involves in several research grants ranging from development of calibrated item bank to teacher education program. Dr Ahmad Zamri bin Khairani is a member of Psychometric Society and International Economics Development and Research Center



**Nordin Abd Razak** is a lecturer at the School of Educational Studies, Universiti Sains Malaysia. He finished his undergraduate study in Bachelor in Education (1991) and Master in Education (1997) from Universiti Sains Malaysia, Pulau Pinang Malaysia and at the doctoral level in 2008 from the Flinders University, Adelaide, South Australia. His area of interest is in the Quantitative Research Methodology, Measurement and Applied Statistical Data Analysis, Cross-cultural study as well as investigating organizational behaviour from the socio-psychological perspective.



**Hasni Shamsuddin** is a postgraduate student at the School of Educational Studies, Universiti Sains Malaysia. Her field of interest is psychometric and educational testing. Her work revolves around linking test items for the development of mathematics item bank. Mrs Hasni is a teacher by profession, and has been teaching Mathematics for the past 16 years.