International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 06, 2020
ISSN: 1475-7192

Items' Difficulty in the Lower Secondary Bahasa Melayu System Diagnostic Test (LSBMSD) Instruments

[1]Mohd Zaharuddin Omar,*[2]Zamri Mahamod

ABSTRACT--- *The main purpose of this research is to assess the suitability of the questions' level and the items' difficulty in the instruments used in the Lower Secondary Bahasa Melayu System Diagnostic Test (LSBMDT). The study is quantitative. Items' detection is carried using Winstepts application (Rasch Model) to determine the overall ranking of 289 items via the analysis of items map. This analysis could provide a clear picture of the right side items' rankings which were calibrated with the respondents on the left side of the logits scale to determine the items' difficulty level. This analysis was validated by seven experts. This study was participated by 935 form one students in Pahang and it involved 10 schools in 10 districts. Alpha Cronbach reliability index for all items was 0.98, whereby the reliability index for each construct of the instruments was between 0.71 and 0.95. It was at the `very good' level. The only exception was the vocabulary construct with 0.71 index (fisher, 2007). The result of the study shows the existence of item free area at the top part with the interval of 1.24 logits and at the bottom part of the variable map with the interval of 0.82 logits. The items classified at 'knowledge', `understanding' and `application' levels were at the right levels. No item was too difficult or too easy. To sum up, UDSBMR instrument contains valid and suitable items as a diagnostic tool to determine any weakness of the language system used in Form 1 Bahasa Melayu.*

Keywords--- *Question Level, Item Difficulty, Reliability Index, Item-Free Area, Items Map Analysis.*

I. INTRODUCTION

The diagnostic test is an achievement test in the cognitive domain (Neukkrug & Fewcett, 2014) which has the purpose to identify the specific strengths and weaknesses of learning as well as to determine the causes of the weaknesses (Abu Bakar & Bhasah, 2008; Noll et al., 1979). In education, Malaysia Examination Syndicate (2006) defines diagnosis as problems or obstacles in terms of difficulty, deficiency and mistakes faced by students in their learning. Teachers are expected to realize which important elements in their teaching that their students had difficulties to master. Consequently, teachers should be able to make remedial in their teaching based on the information gained from the test. This diagnosis could give suggestions to teachers to perform specific remedial programme if they know their students' weaknesses (Siti Rahayah, 2008). According to Bachman and Palmer (1996), Brown (2004) and Yin (2011), the diagnostic test result helps teachers to choose the problematic or difficult subject contents and henceforth plan a more meaningful teaching.

[1] *Education, IPG Campus Dato' Razali Ismail, Terengganu, Malaysia.*
[2]*, *National University of Malaysia, Bangi, Selangor, Malaysia d-zam@ukm.edu.my.*

Received: 22 Sep 2019 | Revised: 13 Oct 2019 | Accepted: 15 Jan 2020

In this study, Bahasa Melayu *Menengah Rendah* diagnostic test was one set of instruments designed to identify or determine the students' difficulties and weaknesses in the aspect of language system for Bahasa Melayu subject at the secondary level and more specifically the form one students. This test consists of five components of language system based on Curriculum and Assessment Standard Document Primary School Bahasa Melayu (Level 2) and Curriculum and Assessment Standard Document Form 1 Bahasa Melayu (Lower Secondary). These two documents cover the morphology, word formation, syntax, spelling, vocabulary and proverbs.

II. METHODOLOGY

This study intends to:

1. Detecting the very difficult and very easy items in the LSBMDT instrument using items map analysis

Ho1: There are no detecting the very difficult and very easy items in the LSBMDT instrument using items map analysis.

2. Detecting the difficulty index values LSBMDT instruments.

Ho2: There are no item had the difficulty index value at the levels of `too difficult' and `too easy'.

3. Detecting the discrimination index values of LSBMDT instruments.

Ho3: There are no item had `weak' and `too weak' discrimination index value.

The main objective of this study is to develop one high quality instrument for Lower Secondary Bahasa Melayu System Diagnostic Test (UDSBMR) which can help students and teachers to identify the existing weaknesses before the students start learning Bahasa Melayu in Form one. The research analysis was to detect the suitability of questions' level, items' difficulty, discrimination index (ID) and difficulty index (IK). In this study, the researcher used *Rasch* model (*Winsteps* application) to analyse the suitability of the questions' level and the items' difficulty. *Excel* programme was used to analyse ID and IK. This study fulfilled the purpose to identify `too difficult' and `too easy' items based on the analysis of items map and *Excel*.

1) Sample

The total of samples was 935. The study was carried out in Pahang, Malaysia. 450 samples were in the urban category whereas the rural had 485 respondents. Ten districts were chosen. Cameron Highlands was combined with Kuala Lipis since Cameron Highlands had only three schools and their students had almost the same characteristics as those in Kuala Lipis. The schools were randomly chosen in each district. The study samples included low and average achievers in Bahasa Melayu UPSR. According to Wright and Stone (2004), Rasch model needs a total of 200 samples for each study group. Schulz (1990) believed that Rasch Model was highly reliable when the sample group had more than 300 people. Linacre (2010) said that the comparison among groups with almost the same number of samples was not necessary.

Pahang was chosen as the study location because this state had an average achievement record in UPSR examination compared to other states in Bahasa Melayu Comprehension and Writing subjects in the last five years. Pahang also had students with different mother tongues like Bahasa Melayu, Chinese, Tamil and other native languages. The three main dialects used were Pahang, Terengganu and Kelantan dialects.

2) Tool

LSBMDT instruments was the main tool used in this study. LSBMDT instruments had 289 items divided into five main constructs which were morphology (111 items), syntax (50 items), vocabulary (15 items), proverbs (28 items), and spelling (35 items). The items' construction covers the important aspect of language system in Bahasa Melayu based on the primary school Bahasa Melayu Curriculum and Assessment Standard Document. The items constructed in the forms of matching, filling in the blanks, giving short answers and structuring sentences.

The procedure of the data analysis was to validate items' reliability based on the suitability of questions' level and items' difficulty in the UDSBMR instrument using Rasch model. According to Grunlund (1993) and Siti Rahayah Ariffin (2008), the item analysis procedure depicts three essential information which concerns (1) item difficulty index, (2) item discrimination index, and (3) the effectiveness of response questions. Herman (1984) and Mohd Najib (2011) believe that difficulty index and discrimination index would determine whether an item can be used, corrected or rejected. The difficulty and discrimination indexes used *Excell* programme whereby Winstept (Rasch Model) was used to analyse the response and items' difficulty levels.

III. RESULTS AND DISCUSSION

Detecting the very difficult and very easy items in the LSBMDT instrument using items map. Winsteps application (Rasch Model) was used to determine the overall picture of all items using the analysis of variable map. This analysis managed to give the overall picture of the items' positions on the right side which was calibrated with the respondents on the left side at the logit scale. Figure 1 shows the LSBMDT items map on 289 items and 935 respondents. The item map analysis shows the max person is at +2.75 logits scale, μ person is at the +0.55 logits scale and the min person is at the -1.56 logits scale. Meanwhile, the max item is at the +1.51 logits scale and the min item is at the -0.74 logits scale. The person separation index value is 5.94 and the item separation index is at 6.14 (standard deviation 0.46). The item map analysis shows the existence of item-free area at the bottom part with the interval of 0.82 logits (max person -1.56 logits – max item -0.74 logits). This is good because no item was too easy that all respondents were able to answer it causing the unsuccessfulness in determining the weak, average and excellent students. The μ person scale shows higher logits value (+0.55 logits) compared to the μ item (0.00 logits). The result of the study shows all items are generally easy since the μ person's logits value was on the μ item's logits value. This is good for UDSBMR instruments and it is considered acceptable as a diagnostic test since more respondents managed to answer the questions compared to those who did not answer the questions.

The positions of the items group's difficulty levels ranging from low to high show the items work well as `knowledge', `comprehension' and `application' items. To summarise, no item is too difficult and no item is too easy. This is due to the existence of item-free area at the upper and lower parts of items' map. All items are considered easy because the μ person logits value (+0.55) is on top of the μ item logits value (0.00) with the interval of 0.55 logits. This result fits the diagnostic test concept which is supposed to be easier. This is because the main objective here is to identify students who could not master certain sub topics, not to make comparison among the students' overall performance. `Knowledge', `comprehension' and `application' level items work well and are in the correct group levels.

Received: 22 Sep 2019 | Revised: 13 Oct 2019 | Accepted: 15 Jan 2020

IV. RESULTS AND DISCUSSION

Table 1: Item Difficulty Index Value of LSBMDT (Ebel & Frisbie, 1986; 1991)

Index	Interpretation	Item / Construct	Total
0.80 – 1.00	Too Easy (corrected or dropped)	No Item/ Construct	0
0.60 – 0.79	Easy (acceptable)	Morphology (62); syntax (37); Sintaksis (33); vocabulary (12); proverbs (19); and spelling (23)	186 64.35%
0.30 – 0.59	Moderate (acceptable)	Morphology (49); syntax (13); Sintaksis (17); vocabulary (3); proverbs (9); and spelling (12)	103 36.7%
0.01 – 0.29	Too Difficult (corrected or dropped)	No Item/ Construct	0

Received: 22 Sep 2019 | Revised: 13 Oct 2019 | Accepted: 15 Jan 2020

Table 2: Item Discrimination Index Value of LSBMDT (Ebel & Frisbie, 1986; 1991)

Index	Interpretation	Item / Construct	Total
0.40 – 1.00	Very good (acceptable)	Morphology (64); syntax (77); Sintaksis (33); proverbs (24); vocabulary (11); spelling (25)	234 80.96%
0.30 – 0.39	Good (acceptable)	Morphology (36); syntax (13); Sintaksis (17); proverbs (4); vocabulary (1); spelling (10)	51 17.64%
0.10 – 0.29	Average (can corrected)	Morphology (1); vocabulary (3)	4 1.38%
0.01 – 0.09	Weak (corrected or dropped)	No Item	0
0.00 and negative (-)	Too Weak (dropped)	No Item	0

From Table 1 and 2 shown that discrimination and difficulty index values of LSBMDT instrument. The analysis of difficulty index and discrimination index is very important to identify the `too difficult' and `too easy' items. The researcher carried out ID and IK to substantiate the result of the study. `Too difficult' and `too easy' items can affect the items' ability to differentiate between the capable and weak students and this might cause some items are not functioning well. The difficulty index is based on the value between 0.0 and 1.0. For example, 0.5 index is equal to 50%. The best index value is 0.5 which is `not easy' and `not difficult'. The determination of items' difficulty index is based on Ebel and Frisbie's suggestion (1991) which is from 0.01 to 0.29 indexes for `difficult' items, 0.30 to 0.59 for `moderate' items and from 0.60 to 1.00 for `easy' items. However, `too easy' items (from 0.80 to 1.00) are still paid attention to. The result of the difficulty index analysis found that no item was too difficult (0.80-1.00) or too easy (0.01 – 0.29). This proves that LSBMDT has very good and acceptable difficulty indexes. The discrimination index (ID) works to differentiate between the good and weak students. Higher index values show the items are better and serve to differentiate between the capable and weak students. The best index value was 1.0 whereby all respondents in top group answered correctly and all respondents in the bottom group failed to answer correctly. The researcher used the index values suggested by Ebal and Frishie (1986, 1991) which were >0.40 (very good), 0.3 to 0.39 (good), 0.20 to 0.29 (average, the items can be corrected) and <0.19 (weak, the items should be corrected or dropped). The analysis of discrimination index on 935 respondents used Excel Programme based on 27% of the top group and 27% of the bottom group in order to get better quality items (Abu Bakar Nordin, 1986). The respondent with the highest score was listed the highest whereby the weakest respondent was listed the lowest. The 27% was 252 top and bottom groups who answered each item correctly and they would be used to calculate the discrimination index.

The analysis of difficulty and discrimination indexes using excel programme was shown in Table 1 and 2. There were 234 items (80.96%) which were at `very good' level, 51 items (17.64%) at `good' level, and 4 items (1.38%) at `average' level. The four items were item A32 (IK 0.57, ID 0.28); item D11 (IK 0.72, ID 0.26); item D13 (IK 0.68, ID 0.28); and item D14 (IK 0.53, ID 0.26) found to have moderate discrimination index and moderate and easy difficulty index value. The items have very good difficulty index value to be accepted which is between 0.53-0.72 as well as moderate discrimination index value (0.26 – 0.28). No item had the difficulty index value at the levels of `too difficult' and `too easy'. Also, no item had `weak' and `very weak' discrimination index value which required the researcher to correct or get rid of the item. Therefore, the researcher decided to accept the items without any correction after checking and analyzing again the items with the agreement of the expert

.

V. CONCLUSION

The result of the study shows many students belonged to higher respondents' min logits group compared to items' min logits. This proves items' difficulty level fulfilled the diagnostic test concept which was supposed to be easier. This is because the main objective of the study was to identify students who could not master certain sub topics not to make comparison of the students' overall performance. The `knowledge', `comprehension' and `application' items worked well and were in their respective group levels: the `knowledge' group was at the lowest level, followed by the `comprehension' and at the highest level was the `application'. LSBMDT instrument items also had an easy and moderate difficulty index levels while the discrimination index value was `very good', `good' and `moderate'. All items were accepted without any correction. Four items had discrimination index value

between 0.26 and 0.29 but their difficulty index value was `very good'. This proves that UDSBMR items managed to differentiate weak, average and excellent students as well as support the item map analysis which shows that no item was too difficult and too easy.

LSBMDT instrument building makes it the first language system diagnostic test in Malaysia for the lower secondary Bahasa Melayu subject. Therefore, this instrument contributes towards Bahasa Melayu subject in secondary school as a standard test instrument to diagnose the language system weaknesses in this particular subject. This study also benefits lower secondary school students and their teachers. The problems of time constraint and lack of skills faced by teachers in preparing high quality test tool could be resolved. The information gathered from this study is useful for planning the teaching approach, method and strategies and it also helps to make teaching to be more focused on students' weaknesses. Teachers can decide the right remedial programmes for their students. Choosing the right curriculum standard for the students can be done correctly. In a way UDSBMR helped the implementation of KSSM.

To summarise, this study is important and helpful in education especially in Bahasa Melayu subject. Mastery Bahasa Melayu could increase the unity of the different races which could lead to the development of the country. KPM with the FPK has the vision to develop individuals who could contribute towards the harmony and prosperity of a nation

REFERENCES

1. Abdul Ghafar, M. N. (2011). Pembinaan dan Analisis Ujian Bilik Darjah. Johor: Universiti Teknologi Malaysia Press
2. Ariffin, S. R. (2008). Inovasi dalam Pengukuran dan Penilaian Pendidikan. Selangor: Universiti Kebangsaan Malaysia.
3. Bachman, L. F., Lyle, F., & Palmer, A. S. (1996). Language testing in practice: Designing and developing useful language tests. England: Oxford University Press.
4. Bond, T. G., & Fox, C. M. (2007). Applying the Rasch model: Fundamental measurement in the human sciences. East Sussex: Psychology Press.
5. Brown, H. D., & Abeywickrama, P. (2010). Principles of language assessment. In Language Assessment: Principles and Classroom Practices. H. D. Brown (Ed.), Boston: Allyn & Bacon, pp. 25-51.
6. Ebel, R. L., & Frisbie, D. A. (1991). Essentials of Educational Measurement. New Jersey: Prentice Hall.
7. Gronlund, N. E. (1993). How to make achievement tests and assessments. Boston: Allyn & Bacon.
8. Gronlund, N. E., & Linn, R. L. (1990). Measurement and Evaluation in Teaching. London: Macmillan Publishing Company.
9. Henning, G. (1987). A guide to language testing: Development, evaluation, research. California: CreateSpace Independent Publishing Platform.
10. Kementerian Pelajaran Malaysia. (2006). Manual Prosedur Pengendalian Ujian Diagnostik Pengajaran-Pembelajaran Sains dan Matematik dalam Bahasa Inggeris (PPSMI) 2006. Putrajaya: Lembaga Peperiksaan Malaysia.
11. Nordin, A. B. & Bakar, A. B. 2008. Pentaksiran dalam Bilik Darjah. Selangor: Longman.
12. Noll, V. H., Scannell, D. P., & Craig, R. C. (1979). Introduction to educational measurement. Massachusetts: Houghton Mifflin Harcourt.

13. Neukrug, E. S., & Fawcett, R. C. (2014). Essentials of testing and assessment: A practical guide for counselors, social workers, and psychologists. Scotland: Nelson Education.
14. Osterlind, S. J. (1998). What Is constructing test items? In: Constructing Test Items. Dordrecht: Springer, pp. 1-16.
15. Yin, R. K. (2011). Applications of case study research. California: Sage Publications.

Received: 22 Sep 2019 | Revised: 13 Oct 2019 | Accepted: 15 Jan 2020