

Differences in Attitude and Behaviour towards Learning English Monolingually and Bilingually: A Rasch Analysis with Banda Aceh Data

Khairiah Syahabuddin¹

¹ *Department of English Language Education, Faculty of Education and Teacher Training, Ar-Raniry State Islamic University (UIN Ar-Raniry), Banda Aceh – Indonesia And Researcher of Centre for Area Studies (CfAS) Ar-Raniry State Islamic University (UIN Ar-Raniry), Banda Aceh – Indonesia*

Keywords: Behaviour in EFL learning, Gender, Bilingual and Monolingual Learning, Rasch Analysis.

Abstract: Many educational variables like motivation, self-concept, self-regulation, learning and so on, have been measured using attitudes and their corresponding behaviours with the same items on a linear scale in the past two decades using Rasch analysis. Supported by the Rasch measures, the theory showed that each attitude item was easier than its corresponding behaviour item. The present study was participated by 780 male and female first-year middle school students (12-13 years old) in 2011, consisting of 394 students taught in bilingual schools and 386 students taught in monolingual schools, which were selected from a number of schools with bilingual and monolingual teaching programs in Aceh. The result of the study showed that attitudes were not always easier than their corresponding behaviours, apparently because students taught monolingually have different attitudes to learning English than bilingually taught students, although both groups of students have similar behaviours towards learning English, as determined by a Rasch measure. This is an interesting finding not previously reported and suggests that Rasch measures of student attitudes by monolingual and bilingual teaching should be further investigated.

1 INTRODUCTION

Aceh province in Indonesia experienced a huge Tsunami in 2004 that involved a massive loss of life (over 220,000 were killed) and almost total destruction of infrastructure and housing. A lot of help and support were offered by Indonesia and the international community, including both financial and material support. One aspect of international supports was the introduction of bilingual education to Aceh province which was part of the newly-implemented curriculum 2004 launched by the Indonesian Ministry of Education in Jakarta. The establishment of formerly standard schools into schools with bilingual programs ran well. Some internationally funded organisations from Turkey, for example, successfully established new private bilingual dormitory schools, with Turkish and English as the main languages used. Public schools (*Sekolah Negeri*) with bilingual programs started to gain popularity as well. Both private bilingual

dormitory schools and public schools with bilingual programs became favourites with the local population (anecdotal evidence from the author). Most of the Aceh parents tried to enroll their students in the bilingual schools and felt unhappy if their children were rejected.

However, there was no a scholarly evidence that students who studied in schools with bilingual programs in Aceh performed better than their counterparts studying at the standard monolingual schools. There was anecdotal folk-rumour that students studying at public schools with bilingual programs out-performed students studying at public schools with monolingual programs, yet there was no published evidence for this conclusion.

This study was part of larger study which investigated English language achievement of first-year Middle School students in Banda Aceh, Indonesia, regarding their ability in English Reading Comprehension, English Writing, and their Behaviour towards Learning English regarding their experience in learning English as a foreign language,

in context of difference in type of schools and Gender. The aim of this study was to investigate students' attitude and behaviour towards learning English bilingually and monolingually in Banda Aceh. This paper reports on the measurement of students' English behaviour using Rasch analysis with the RUMM2030 computer program (Andrich et al. 2010).

English used in bilingual classes is around 50% with another 50% for Bahasa Indonesia. English used in monolingual classes is around 20% with the rest 80% for Bahasa Indonesia. For the purpose of this study, a bilingual school program is defined as the teaching of English by using a combination of English and Bahasa Indonesia with equal percentage of time, and the monolingual school program is defined as the teaching of English by using a majority of Bahasa Indonesia and a small amount of English by time. Except for the medium of language instruction, bilingual and monolingual groups maintained all other conditions and situations the same. Both groups shared the same curriculum/syllabi, the same length of learning time, the same textbooks, the same teacher's qualification, the same kind of materials for homework, and the same English activities at school.

Teachers in both types of schools were counselled and monitored during the experimental study in order to ensure that there were no changes in conditions and situations and that all schools were working the same way throughout the two months of study. The two months were chosen due to two considerations. First is the study or English content issue. This is related to students' exposure to outside classroom or outside school activities that could confound the experiment and call the conclusions into question. Should any types of students learn extra English content which differentiated them from the other group, the conclusions of the study could be called into question. So, outside school English learning (such as homework) was closely monitored. Second is teacher willingness to help with the monitoring and checking. The study encouraged the school English teachers to participate in the study voluntarily. Beside their participation in the study, they had their own professional and personal commitments and two months was the agreed time between pre-tests and post-tests.

1.1 Previous Studies on ESL Students and Rasch

There were no reported studies conducted, similar to the current study, from Indonesia or Aceh where Middle School Students' behaviour in learning English as a foreign language was measured, or in Southeast Asia. The following are previous studies on ESL students. Smith (2009) examined the behaviours of EFL Haitian students. The examination included engagement in lessons with students' dissimilar levels of English proficiency (beginning, intermediate, and advanced) during a Sheltered English Immersion (SEI) English Language Arts (ELA) literacy block. The study resulted in two inferences: First, support is needed for English language learners in their primary language during English instruction, and second, specific structural grouping preferences should be recognised and supported, along with exposure to all structural groupings. Tillemma et al. (2011) investigated whether students' response on offline questionnaire can predict their online meta-cognitive processing during their writing tasks. The results showed that the online sequential disseminations of reading the task and preparation are unlike for dissimilar degrees of informed writing styles. Brown & Sachdev (2009) looked at bilingual identity, vitality, behaviour, and attitudes of 95 Japanese speakers in United Kingdom. The findings confirmed the dominance of Japanese in proficiency and identity. It suggested some methodical variances in use and attitudes in relation to context; integrating the three factors related to identities and vitalities to predict English use and attitudes; and the use and attitudes of Japanese to deal with the students' social contact. Du-Babcock (2006) analysed topic management strategies and turn-taking behaviours in the Hong Kong bilingual environment. The result suggested that different communication behaviours occurred in tasks assigned for both Cantonese and English meetings, indicating that second-language proficiency is likely a contributing factor that affects the topic management of Chinese bilinguals when participating in Cantonese. Farghal & Haggan (2006) examined compliments behaviour in bilingual Kuwait college students. With the 632 compliment responses, the result indicated that the influence of Arabic was very robust over English, implying the linguistic and culture foundations rousing the responses.

Research on attitude and behaviour by using Rasch measurement have been widely conducted during past decades. Some were reported here. Leung & Waugh (2010) conducted a study on attitude towards career counselling of secondary level students in Hong Kong. The study was participated by 182 students. The study revealed that the majority of the students have positive attitude towards career counselling at the schools. Merrell (2005) carried out a study on hyperactive and impulsive behaviour in young children which was linked to their academic performances by using The Attention Deficit Hyperactivity Disorder (ADHD) criteria. The data was taken from 1821 students from 70 schools and was analysed by Rasch analysis. The study supported previous works and theoretical points which lead to an issue whether the criteria was appropriate to use for all ages or not.

2 RESEARCH METHOD

2.1 Data Collection, Procedure, and Participants

Three hundred and eighty year-7-students from thirteen public Middle Schools in Banda Aceh participated in the study. Three hundred and ninety four students were selected from schools with bilingual instruction and 386 students were selected from schools with monolingual instruction. Similar characteristics were shared between the two groups. Their English competence was limited, which was due to the fact that they just started learning English in their year 7 because English became one of the compulsory subjects in that year. They spoke both Bahasa Indonesia and Acehnese inside and outside the classroom. In the classroom English was not spoken, but learnt, especially through reading texts and grammatical drills. Only a few spoke English, read English, and watched movies as their hobby. At schools, both of the students were taught using the same English syllabi and curriculum; similar English teaching approaches, methods and strategies, and similar tasks.

The amount of time spent on English lessons at schools became the only difference between the two groups. The amount spent on English lessons a week for bilingual-taught schools were around 55 hours while the amount spent on English lessons were 38 hours for monolingually-taught schools for a week.

So, the amount of time spent was the major difference in spite of their similarity in their English scarcity.

2.2 Instrumentation

To investigate students' attitude and behaviour towards their learning English in both schools type in Banda Aceh's context, a Questionnaire test was designed. The questionnaire which consisted of questions for both attitude and behaviour, comprising a total of 21 statements, was divided into four categories: (1) tasks for listening; (2) tasks for speaking; (3) tasks for reading; and (4) tasks for writing. Each statement was responded in three response categories, 'never or rarely', 'some of the time' and 'most of the time'. 'Never or rarely' was scored 0; 'some of the time' was scored 1, and 'most or all the time' was scored 2.

2.3 Measurement

This study uses the computer program RUMM2030 (Andrich et al. 2010). It provides statistics and graphs to support the assertion that a linear unidimensional measure was made. There are number of steps to be considered in using this analysis (see for example Waugh 2003, 2005, 2010b, 2010a). These included, for example but not limited to, checking the theoretical ordering of the items against the measured order; checking the Standardized Fit Residuals; Item-Trait Interaction (for respondent agreement on the difficulties of the items); Person Separation Reliability Index, Individual Item Fit, Response Category Curves, differential item functioning and appropriate targeting (by gender and type of English program).

3 RESULT AND DISCUSSION

3.1 Initial Analysis

The original questionnaire involved attitude and behaviour responses to 21 items, making an effective scale with 42 items. The items were ordered theoretically from easy-to-medium-to-hard on a continuum with the attitude items expected to be easier than their corresponding behaviour items.

After the analyses, 20 of the 21 attitude items (item 7 was the exception) were deleted because of misfit to the measurement model. This was not consistent with the model used to develop the questionnaire that was based on many previous studies where attitude and behaviour were measured together (see Waugh 2003, 2005, 2010a, 2010b). The RUMM program does not tell the researcher why an item doesn't fit the Rasch Measurement

Model; just that it doesn't fit. It was difficult to see why the attitude items didn't fit the measurement model, but the students were in their first year of middle school and, because they were not strong in English reading, the misfit may have been primarily due to their low command of reading and understanding of English, related to the different classroom culture in the two groups, in a test situation. The two groups of students (bilingually-taught and monolingually taught) did not have agreement about the difficulties of the items because of their differences in their command of English, combined with some differences in culture (living in an Indonesian culture, and learning English culture through English lessons that was taught mostly in Bahasa Indonesia), and that was the substantial cause of the misfit. Analysing the 21 attitude items alone also did not produce a linear scale and so the analysis was continued after deleting the 20 non-fitting attitude items.

A final analysis showed ten behaviour items (items 2, 8, 10, 12, 16, 18, 26, 28, 36 and 38) and one attitude item (item 7), produced a good fit to the measurement model. Deletion of the attitude item 7 (*"I say new words several times in English"*) and a re-analysis with the ten behaviour items produced a worse fit to the measurement model and so the attitude item 7 was re-instated. The following material shows the output from the RUMM program when a good, unidimensional, linear scale of Behaviour with respect to Learning English was created with 11 items for these Acehese students.

3.2 Output from Final Analysis

3.2.1 Standardised Fit Residuals

To match the measurement model, the Fit Residual for both items and students should be near 0 and the standard deviation should near 1. For this study, the mean for the Fit Residual had 0.195 for items and -0.306 for persons and the Standard Deviation was 1.112 for items and 1.525 for persons. It means the data fit the model in which it had a reasonable reliability of item-student response pattern. Table 1 shows item-person fit to the measurement model for the Behaviour measure.

Table 1: Overall Fit Statistics for the Behaviour Measure (N=779, I=11)

ITEM-PERSON INTERACTION				
	ITEMS		PERSONS	
	Location	Fit Residual	Location	Fit Residual
Mean	0.000	0.195	0.458	-0.306

Standard Deviation	0.217	1.112	0.972	1.525
--------------------	-------	-------	-------	-------

Notes:

1. The mean of the item difficulties is constrained to zero by the measurement model.
2. The fit residuals will approximate a distribution with a mean near zero and a standard deviation near one, when the data fit the measurement model.

3.2.2 Dimensionality and Item-Trait Interaction

In order to determine whether a unidimensionality trait has been measured, an item-trait interaction chi-square is needed. The item trait interaction chi-square for this study was 103.82, df=99, and p=0.35 (see Table 2). This indicated that all along the scale, the agreement amongst the students about the item difficulties was good. It shows that the students agreed as to which items were the hardest, which were of medium difficulty, and which were the easiest. This, sequentially, means that, prediction to each student's response to each item can use the person measure (a single parameter for each student) and the item measure (a single parameter for each item). The first eigenvalue which was shown from a major component examination of the residuals was 1.45 which was acceptable, indicating that the data produced a unidimensional measure.

Table 2: Item-Trait Interaction for Behaviour Scale

Total Item Chi-Square	103.82
Separation Index	0.71
Total Degree of Freedom	99.00
Total Chi-Square Probability	0.35
Cronbach Alpha	0.73

Notes:

1. The Index of Person Separation was good.
2. The item-trait interaction test specified that there was good agreement among the students on the item difficulties.
3. All numbers are given to two decimal points because the errors are only up to two decimal points.

3.2.3 Person Separation Index

A Person Separation Index is an indicator that the student measures are well separated along the scale in comparison to the students' measurement errors. To have a good measure, the Person Separation Index should be 0.75 or larger. The Person Separation Index for this study was 0.71 (see Table 2 above) showing a rational separation of measures with regard to the errors. Unlike Cronbach Alpha

which is calculated on the data raw scores, Person Separation Index is calculated on the parameters that are created by Rasch. For this study, the Cronbach Alpha was 0.73, which once more shows that the data had a reasonable consistency.

3.2.4 Individual Item Fit

For the Attitude and Behaviour measure, eleven items fit the measurement model (see Table 9.3). In addition, all the standardized residuals fall within the range -1.2 to +1.1, supporting a reasonable fit to the measurement model (which usually has to be within plus or minus two SDs).

Table 3: Locations, Standard Errors, Residuals and Chi-Squares for Behaviour Items

Item No	Location	SE	Residual	DF	Chi-Square	df	Probability
2	-0.270	0.069	0.471	706.27	5.065	9	0.83
7	-0.173	0.056	0.946	706.27	8.522	9	0.48
8	0.133	0.060	-0.098	706.27	4.401	9	0.88
10	0.129	0.059	0.699	706.27	15.060	9	0.01
12	0.382	0.059	0.030	706.27	7.224	9	0.61
16	-0.281	0.059	-1.133	706.27	14.555	9	0.10
18	0.101	0.058	0.314	706.27	5.077	9	0.83
26	0.065	0.059	0.229	706.27	6.843	9	0.65
28	0.185	0.057	0.563	706.27	4.026	9	0.91
36	-0.481	0.059	-0.225	706.27	11.377	9	0.25
38	0.211	0.057	1.072	706.27	9.999	9	0.35

Notes:

1. Location refers to the item difficulty in logits (the log odds of answering the response categories positively). SE is the standard error in logits.
2. Residual is the difference between the observed and expected responses.
3. df means degrees of freedom. Probability is based on the chi-square fit to the measurement model.

3.2.5 Threshold Values

Items thresholds are positions on the scale between adjacent response categories where the odds are 1:1 that students will respond to a particular item, in

either category. It is expected that the students would use the thresholds in the way that they were intended by the researchers and so the thresholds should be ordered in line with the conceptual ordering of the response and scoring categories. The thresholds were ordered in line with the conceptual ordering of the scoring categories and thus were in agreement with the intention of the researchers for this measure (see Table 4).

Table 4: Item Thresholds Uncentralised (Item=11, Number=779) for Behaviour Measure

Item	Item Location	Thresholds	
		1	2
2	-.222	-1.805	1.361
7	-.132	-.804	.539
8	.157	-1.188	1.504
10	.124	-1.038	1.287
12	.352	-.623	1.327
16	-.142	-1.144	.858
18	.061	-.935	1.058
26	.092	-.975	1.160
28	.030	-.831	.893
36	-.439	-1.109	.230
38	.117	-.831	1.066

Note: The thresholds are ordered in line with the scoring categories.

3.2.6 Scoring Category Curve

The RUMM2030 program produces curves of the scoring categories for each item. The Scoring Category Curves show the relationship between the probabilities of scoring in each category. Each item has three response categories: 'Never or rarely' (scored 0); 'Some of the time' (scored 1); and 'Most or all the time' (scored 2). The Scoring Category Curves should show a consistent relationship between the probability of scoring and the measure from low to high indicating that the scoring was done consistently and logically. A Scoring Category Curve for Item 2 is shown in Figure 1.

Figure 1 (Item 2: "I pay attention to someone speaking English") showed that the scoring was done logically and consistently. When students have low measures on item 2, then they have a high probability of obtaining a zero score (the lowest response); and when they have a medium measure, they have a medium probability of scoring 1 (the moderate response); and when they have a high measure, they have a high probability of scoring 2 (the highest response). The Scoring Category Curves for the other items were checked and they showed logical and consistent scoring as well.

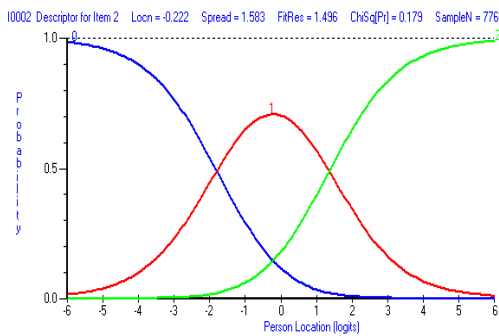


Figure 1: Item Category Curve for Item 2

Note: The blue curve is for a score of 0, the red curve for a score of 1 and the green curve for a score of 2.

3.2.7 Item Characteristic Curves

The Item Characteristic Curve provides information on item differentiation between persons and the item location. A group of students is considered to have performed well if their values (in the form of black dots) fit on the ogive curve. The ogive curve is the expected values for an item against the student measures (low to high). Take as an example Item 2 (Figure 2). With most of the dots close to the curve, it is considered a good fit to the measurement model and shows good discrimination. The Item Characteristic Curves for the other items were checked and found to be satisfactory.

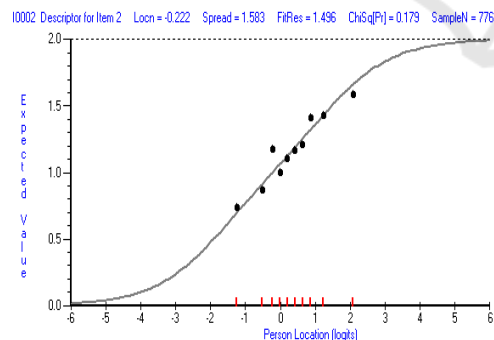


Figure 2: Item Characteristic Curve for Item 2

3.2.8 Person-Item Threshold Distribution (Targeting)

The RUMM2030 program produces item difficulty on student measure which is also known as targeting graph. On that graph, the student measures are placed on the same scale as the item difficulty in a standard unit. A well-targeted measure is achieved when the thresholds and student measures are at

about the same range, indicating that the items are within the capability of the students to answer. Figure 3 illustrates the targeting graph for the Behaviour measure. The targeting of the Behaviour measure items is not as satisfied as estimated because there were inadequate easy, medium, and hard items. This was because several original items misfit the measurement model and therefore were deleted after the initial analysis in which it showed that the two types of students disagreed on most item difficulties.

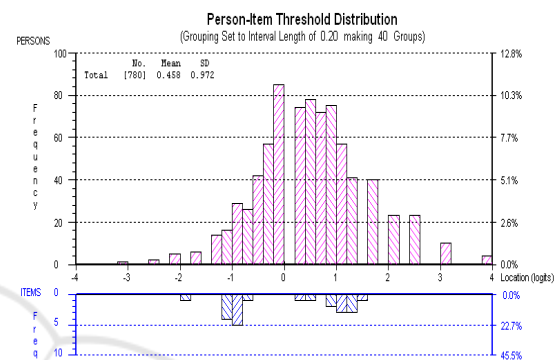


Figure 3: Person-Item Threshold Distribution for Behaviour

Figure 4 shows the distribution of item thresholds and the students on a 'map' and this also shows the restricted range of item thresholds. Somehow, this targeting problem would have to be rectified in any future use of the scale for these students.

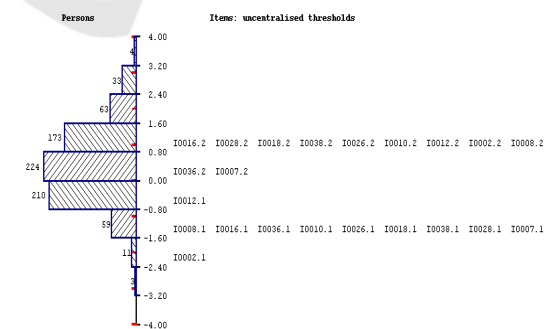


Figure 4: Behaviour Map

Note: I0016.2 means threshold 2 for item 16, I0002.1 means threshold 1 for item 2, and so on.

3.2.9 Differential Item Functioning (DIF) by Gender

Each of the 11 items of the Behaviour measure showed no statistically significant Differential Item Functioning (DIF) by gender (see Figures 5, 6 and 7, for examples).

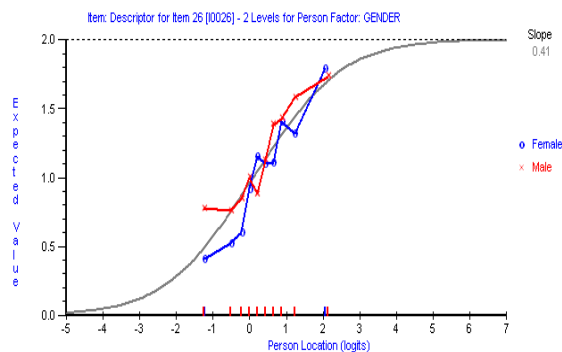


Figure 5: Item Characteristic Curves by Gender for Behaviour Item 26

Note: No statistically significant main effect by gender, $F=5.02$, $df=19,1$, $p=0.025$.

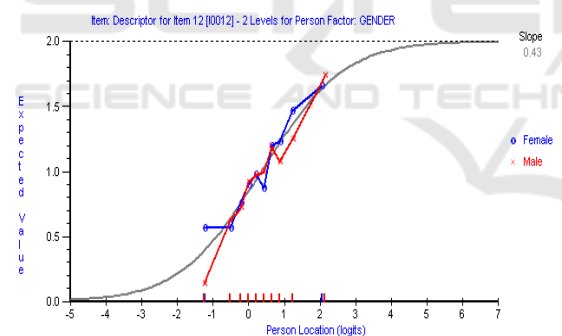


Figure 6: Item Characteristic Curves by Gender for Behaviour Item 12

Note: No Statistically significant main effect by gender, $F=2.96$, $df=19,1$, $p=0.09$.

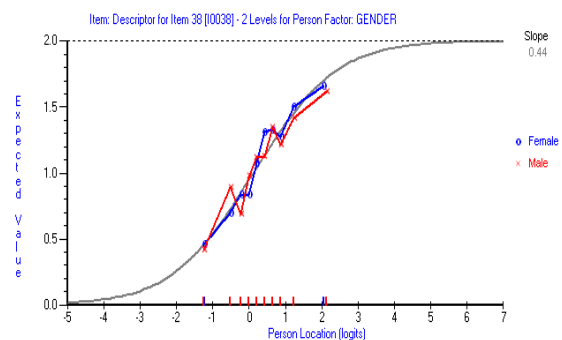


Figure 7: Item Characteristic Curves by Gender for Behaviour Item 38

Note: No statistically significant interaction effect by gender, $F=0.19$, $df=19,1$, $p=0.66$

All the Item Characteristic Curves for the other items were checked and showed no statistically significant difference by gender.

3.2.10 DIF such by Type of Language Instruction

Only one item (Item 7) showed DIF by type where bilinguals had improved results on the Behaviour measure compared to monolinguals ($F=23.81$, $df=19,1$, $p=0.00000$, see Figure 8). Figures 9 and 10 show that there was no DIF for item 2 and item 38. However, over the Rasch measures for all the 11 items together, bilinguals had a statistically significantly higher Behaviour measure than monolinguals ($F=20.56$, $df=1,778$, $p=0.0000$, see Figure 12).

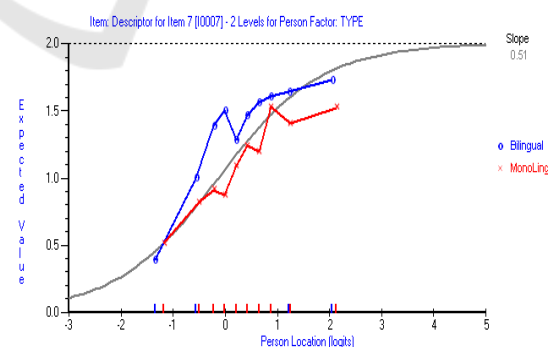


Figure 8 Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Behaviour Item 7

Note: There is a statistically significant main effect by type, $F=23.81$, $df=19,1$, $p=0.00000$

Bilinguals have a statistically significantly higher Behaviour measure.

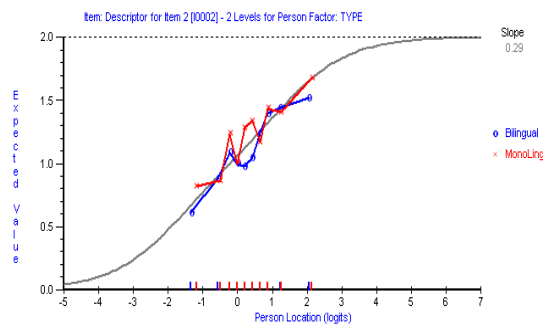


Figure 9: Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Behaviour Item 2

Note: Not statistically significant by type of teaching ($F=5.65$, $df=19,1$, $p=0.02$)

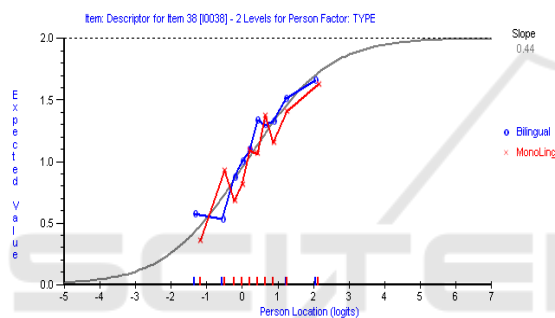


Figure 10: Item Characteristic Curves by Type of Teaching Methods (Bilingual v. Monolingual) for Behaviour Item 38

Note: Not statistically significant by type of teaching $F=4.66$, $df=19,1$, $p=0.03$

3.2.11 Targeting by Gender and Type of Language Instruction

For targeting by gender and type of language instructions (bilingual and monolingual), girls revealed a statistically significantly improved Behaviour measure than boys ($F=11.52$, $df=1,778$, $p=0.0007$, see Figure 11) and bilingually-taught students have a statistically significantly better Behaviour measure than monolingually-taught students ($F=20.56$, $df=1,778$, $p=0.00002$, see Figure 12).

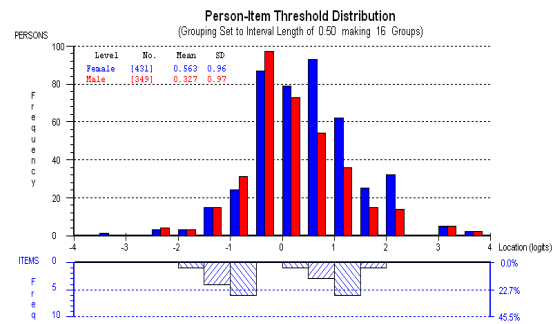


Figure 11: Targeting of Behaviour by Gender

Notes:

1. The person measures are on the upper-side of the graph from low (LHS) to high (RHS).
2. The item difficulties are on the lower-side side from easy (LHS) to hard (RHS). $F=11.52$, $df=1,778$, $p=0.0007$, which is a statistically significant difference.

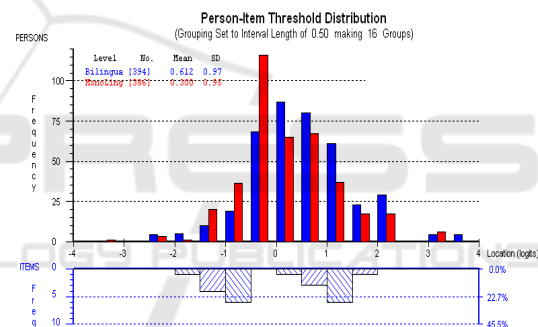


Figure 12: Targeting of Behaviour by Type of Language Instruction

Note: $F=20.56$, $df=1,778$, $p=0.00002$ which is statistically significant for bilingual teaching.

3.2.12 Scale of Item Difficulties

Table 5 shows the item wording for Behaviour. The items have been ordered by difficulty from the easiest to the most difficult on the linear Rasch-created scale. The ordering of the items is consistent with the initial predicted conceptualized order, supporting the construct validity of the scale.

Table 5: Order of Difficulty of Items on the Linear Scale

Item No.	Item Location	Item Statements on the Behaviour
36 (easy)	-0.44	I actually like the way my teacher teaches English writing.
2	-0.22	I actually pay attention to someone

		speaking English.
16	-0.14	I actually read carefully words in English.
7	-0.31	I wish to say new English words several times.
28	0.01	I actually can understand English better when I do activities with friends
18	0.06	I actually can read English at home on my own.
26	0.09	I actually learn more when I study English in groups.
38	0.12	I actually like English because we use it in the classroom.
10	0.13	I actually practice English with other students.
8	0.16	I actually say new English words several times.
12 (hard)	0.35	I actually start conversation in English with my friends.

Note: Item difficulties (locations) are measured in logits, the log odds of answering successfully.

The items were ordered from easy to hard on a linear scale (see Table 5) so that it can be seen which items are easy and which are hard. The easiest item involved the students' preference for the way their English teachers teach in the classroom, as expected (item 36 difficulty = -0.44 logits). The hardest item involved oral skill in ability to initiate speaking in English with their friends, as expected (item 12 difficulty = +0.35 logits).

4 CONCLUSIONS

This study presents the results of the data analysis for the process of students' behaviour in learning English as a second language in Aceh. Eleven items from the original 42 items produced a linear, unidimensional measure (31 items were deleted: 20 attitude items and 11 behaviour items). The Fit Residual data showed that there was a good consistency for the item-person response pattern. The Item-Trait Interaction (dimensionality) indicated that there was good agreement about the item difficulties along the scale. The Person Separation Index indicated that there was good separation of measures in comparison to errors. All items fitted the Rasch measurement model. The threshold values and the Scoring Category Curves showed that the scoring categories were used consistently and logically. The Item Characteristic Curves showed reasonable discrimination. All these data support the view that a linear, unidimensional measure of Behaviour was created so that valid inferences could be made.

However, 31 items, which were consisted of 20 attitude items and 11 behaviour items and which were initially considered to be conceptually valid, had to be deleted because of misfit to the measurement model, apparently because the bilinguals and the monolinguals did not agree on the item difficulties.

On the attitude items (see Appendix A for unfit Attitude items), the disagreement occurred because bilinguals and monolinguals think differently about learning English, consisting tasks on listening, tasks for speaking, tasks for reading, tasks for writing, student/student relationship, student/teacher relationship, and common views. This agreement was related to different views bilingual and monolinguals had towards their learning English as a foreign language, especially regarding the attitude items that previously mentioned. For tasks for listening, both groups did not reach agreement on what they thought on paying attention to someone speaking English, on asking others to speak slowly or repeat words in English, and on listening to English songs. Based on observation conducted on the time of data collection, it showed that monolingual students did not pay an adequate attention on listening tasks. It was strongly associated with possession of inadequate vocabulary which was considered crucial in order students to be able to listen to English conversations well. Added to it was that this lack of vocabulary contributed to their un-readiness for English listening activities that took a place in a monolingual English classroom in Aceh province. Therefore, it was unlikely that the monolingual students asked others to speak slowly or repeat words in English. On the other hand, bilingual students seemed to have better achievement in all listening tasks that monolinguals students were lack from. The bilingual students showed that they thought they liked English listening tasks. They were aware that they paid good attention to someone who was speaking English, for example, their English teacher or other students with good ability in English speaking. In line with that, they also believed that they asked others to speak slowly or repeat words in English in order that they learned listening better, or on how to pronounce or how to grasp a meaning of an utterance. The bilingual students also thought that they liked listening to English vocabulary, especially for meaning of words or expressions, or on spellings or intonations.

Similarly, for other tasks, bilingual students had better thought towards tasks for speaking, tasks for readings, and tasks for writing. They showed better

motivation and dedication on their learning English, in this case, for English speaking, English reading and English Writing, due to the fact that they were considered to possess sufficient English vocabulary to start with the tasks. Sufficient English vocabulary was very crucial because students could create spoken utterances, understand reading passages, or write something with. With adequate vocabulary, bilingual students wished to practice English with other students and to start conversation in English with their friends. Having sufficient vocabulary encouraged them to guess the meaning of the English words in the text, to read carefully words in English and to read English at home on their own. In addition to that, having adequate vocabulary motivated them to look for similar words in their own language, and to divide English words into parts that they understand as well as to write to write feelings in a diary in English.

On the contrary, monolingual students thought differently due to the limited vocabulary that they possessed regarding the above tasks. They did not have enough vocabulary to be engaged in most of English tasks. Most of them could only know basic English vocabulary which was insufficient to trigger English conversation, English reading, and English writing.

Different agreement also occurred in the way both groups of students view their relationship with their classmates, with their English teacher and on common views. Bilingual students wished to learn more when they studied English in groups, to understand English better when doing activities with friends as well as to lead other friends in doing English activities. They also wished to learn a lot from their English teachers, thought that they like the way their teachers taught English Reading Comprehension, and English writing. They liked English because they used it in the classroom, because it helped them in higher study, and because it helped them go abroad. On the other hand, monolingual students seemed to believe these differently.

Regarding the behaviour items (see Appendix B for unfit Behaviour items), the disagreement was because the bilinguals and monolinguals behaved differently about learning English, consisting with tasks for listening, tasks for reading, on student/student relationship, and on common views. For tasks for listening, monolinguals students did not ask others to speak slowly or repeat words in English, nor listen to English songs the way bilingual students did. Further disagreement was also shown on tasks for reading, on student/student

relationship, student/teacher relationship, and common views. For tasks for reading, they guessed the meaning of the English words in the text and look for similar words in their own language. For tasks of writing, they wrote feelings in a diary in English and divided English words into parts that they understood. For student/student relationships, student/teacher relationships, and common views, they lead other friends in doing English activities, learned a lot from their English teacher, liked the way their teacher taught English Reading Comprehension, and liked English because it helped them in higher study and because it helped them go abroad. However, monolingual students did them differently with the bilinguals.

The reason for this is similar to the way they responded to the attitude items. Monolingual students seemed not to have as strong motivation as bilingual students and to have lack dedication to learn English, which could be related to their un-readiness to be engaged in learning English where sufficient vocabulary was involved. Compared to monolinguals, bilinguals had English-based activities at schools. These activities were believed to increase the students' motivation to learn English. Some activities that had been witnessed were, speech competition, story-telling, and narrative writing. Students and teacher dedicated some time in a week to do practices. In order to do the practices well, some other activities such as book reading, was involved, especially for the speech competition. In doing so, the children were lent with some books or magazines or articles from the Internet to ensure their well-informed prior to the story telling. On the contrary, such activities were not conducted at monolingual schools. When being confirmed to the teachers, they told that there were some problems either with the students' lack participation or with teachers' decision by not taking parts on the competition. In classrooms, bilingual students were found active and motivated. Most of the question-answer sessions were alive. Students raised their hands to ask questions or clarified things. Teachers explained materials under questions thoroughly. Those atmosphere, on the other hand, were rarely seen on the monolingual classrooms. Students seemed reluctant to raise hand and ask questions or did clarifications. When questioned the reason why, some of them told that they did not know what to ask for clarification because they knew very little. It was hard for the monolingual students to get involved into any English activities with their limited vocabulary.

Due to this disagreement, the 20 attitude items and 11 behaviour items need to be deleted. The deletion of these items caused a targeting problem, where the final scale had insufficient items across the whole difficulty range of the student measures. Thus this problem needs to be investigated further so that there are sufficient easy, medium and hard items to cover the full range of student measures in any future use of this scale.

The Item Characteristic Curves showed that the large majority of the items had no statistically significant differential item functioning (DIF) by gender and by type of language of instruction (bilingually-taught and monolingually-taught). However, over the Rasch measure for all 11 items together, bilinguals had statistically significantly better Behaviour measure than monolinguals and females had a statistically significantly better Behaviour measure than males.

ACKNOWLEDGEMENTS

The author (Khairiah Syahabuddin) thanks the Australian Endeavour Scholarship (Department of Education, Employment and Workplace Relation, DEEWR) for funding her doctorate study at the Edith Cowan University, Perth, Australia, during which she conducted this research. Her thank also goes to Professor Russell Waugh, her supervisor, without whom she could not figure out herself with her emerging knowledge of Rasch Measurement. Her thank is also dedicated to her family who always supports and prays for her.

REFERENCES

- Smith, J., 1998. *The book*, The publishing company. London, 2nd edition.
- Andrich, D., Sheridan, B. & Luo, G., 2010. Rasch unidimensional measurement model (RUMM2030) computer program.
- Brown, I. & Sachdev, I., 2009. Journal of Multilingual and Bilingual behaviour, attitudes, identity and vitality: Some data from Japanese speakers in London, UK. , 30(July, 2009), pp.37–41.
- Du-Babcock, B., 2006. An Analysis of Topic Management Strategies and Turn-Taking Behavior in the Hong Kong Bilingual Environment: The Impact of Culture and Language Use. , 43, pp.21–42.
- Farghal, M. & Haggan, M., 2006. Compliment Behaviour in Bilingual Kuwaiti College Students. , 9(1), pp.94–118.
- Leung, Thomas Kin Man; Waugh, R., 2010. A Rasch Measure of Career Decision-Making Self-Efficacy. In *Applications of Rasch Measurement in Education*. pp. 143–154.
- Merrell, C., 2005. Rasch Analysis of Inattentive, Hyperactive and impulsive Behaviour in Young Children and the Link with Academic Achievement. *Journal of Applied Measurement*, 6(1), pp.1–18.
- Smith, M., 2009. *Academically engaged behaviors during literacy instruction for three Haitian English Language Learners in a Sheltered English Immersion First Grade Classroom*.
- Tillema, M. et al., 2011. Relating self reports of writing behaviour and online task execution using a temporal model. , 6, pp.229–253.
- Waugh, R.F., 2010a. *Applications of Rasch measurement in education*, New York: Nova Science Publishers.
- Waugh, R.F., 2005. *Frontiers in educational psychology*, New York: Nova Science Publishers.
- Waugh, R.F., 2003. *On the forefront of educational psychology*, New York: Nova Science Publishers.
- Waugh, R.F., 2010b. *Specialized Rasch measures applied at the forefront of education*, New York: Nova Science Publishers.