Using Rasch Model to Assess the Foreign Language Speaking Anxiety Scale (FLSAS) Among University Students in Salatiga, Indonesia

Rizki Parahita Anandi[1], Bambang Sumintono[2], Muhammad Azhar Zailaini [3,] and Resa Syafitri[4]

[1]Universitas Islam Negeri Salatiga, Indonesia
[2] Universitas Islam International Indonesia, Indonesia
[3] Universiti Malaya, Malaysia
[4]Universiti Malaya, Malaysia
rizkiparahita91@uinsalatiga.ac.id

Abstract. This research was aimed to investigate the use of the instrument Foreign Language Speaking Anxiety Scale (FLSAS) with Rasch measurement model approach. About 46 Arabic Language Education students have participated in the study. Rasch model provided some analysis to examine the quality aspect of an instrument reliability and validity. The findings revealed that both value of person (0.90) and item (0.92) reliability are considered very high. There were 2 misfit items in the questionnaire based on the criteria of acceptable value in Outfit MNSQ, Outfit ZSTD and Point Measure Correlation, therefore, they were eliminated. The value of raw variance of the questionnaire was 49% and the unexplained variance index of it did not exceed 15%. This result indicated that the questionnaire was valid in terms of its internal validity and it can be used to measure students' anxiety in speaking Arabic language. Regard to rating scale analysis, it is revealed that one of the five-point rating scale need to be collapsed. The Wright map showed one item which was placed on the top of the map that represented the situation where students did not feel anxious when speaking Arabic language. On the other hand, the map also showed an item at the bottom of the map which means that most of students get anxious when speaking Arabic language in that situation. Overall, FLSAS instrument is appropriate to be used by the foreign language teachers to measure students' speaking anxiety.

Keywords: Arabic language, Foreign Language Speaking Anxiety Scale (FLSAS), speaking anxiety, Rasch Model

1 Introduction

Anxiety is found to be one of many problems faced by both foreign language teachers and learners during the language learning process. Some people believe that anxiety is only a minor inconvenience for language learners, while some others believe that it has a strong effect on them [1]. Horwitz and Young (as cited in [1]) explained that around half of those who attended language courses experience the kind of language anxiety which negatively affected them. For instance, those who experienced language anxiety

Q. Zhang (ed.), *Proceedings of the Pacific-Rim Objective Measurement Symposium (PROMS 2023)*, Atlantis Highlights in Social Sciences, Education and Humanities 23,
https://doi.org/10.2991/978-94-6463-494-5_6

said that their mind goes blank when their teacher called their names during the foreign language class, whereas they were considered good learners in other subjects and situation, also highly motivated as well [2]. Doqaruni [3] also found some students who experienced anxiety would keep silent and do not actively involve in speaking class.

Anxiety itself can be defined as a blend of both subjective and physiological responses. The subjective responses are linked with the anxiety that comprise feelings of unease, stress, worry, and nervousness [4]. It also can be viewed as a state of apprehension of fear triggered by the expectation of something threatening [5]. Students expressed that they struggle with mental block in learning foreign language [2]. When they cannot express themselves in the target language, they are not sure of their performance and they feel threatened. Therefore, researchers are encouraged to find appropriate way how to investigate more about language anxiety both qualitatively or quantitatively.

Some researchers have established instruments to measure students' anxiety, such as the Foreign Language Classroom Anxiety Scale by Horwitz, Horwitz and Cope [2], but only a few of them focused on measuring anxiety in speaking. The Foreign Language Speaking Anxiety Scale (FLSAS) developed by Balemir [6] from Huang is one of the questionnaires which is aimed to measure students' speaking anxiety. The items in the FLSAS represented various situations in the speaking classroom such as speaking in front of classmates, discussion in a group, teacher's manner in speaking class and many more. The questionnaire itself is using a five-point Likert-type scale choice of answer which comprised of 28 items. Balemir found that the students encountered a moderate level of anxiety.

In Indonesia, especially in the teaching of Arabic language, researchers still found various problems during teaching and learning. For instance, Fajri [7], who conducted interviews with Arabic language university students, found that students were shy to speak in front of others in the *al Kalam 3* or Arabic language speaking class because they were afraid of making mistakes. Afraid of making mistake has been identified as one of reason why students feel anxiety. While Hidayati [8] conducted research to measure students' anxiety. She employed FLCAS by Horwitz to measure the students' anxiety. But her study was not focused on measuring the specific skill, like speaking skill.

To sum up, there are not many Arabic language researchers who measure students' anxiety in speaking using a questionnaire survey. As Apple [9] explained that a foreign language teacher needs to measure students' level of anxiety before dealing effectively with classroom anxiety. Therefore, the first author has translated the FLSAS into Indonesian language and made some modifications for the purpose of identifying students' anxiety in speaking Arabic language. The FLSAS was chosen because it is focused on measuring students' speaking anxiety.

Then to check the quality of an instrument is crucial because the data and finding's quality would rely on it [10]. The Rasch analysis is chosen because it will provide the empirical proof to the instrument's quality both at instrument and items level, using individual centered statistics approach. The quality of the FLSAS will be examined over several stages of analysis.

2 Method

2.1 Research Objectives

This research attempted to examine the quality of the Foreign Language Speaking Anxiety Scale (FLSAS) by using the Rasch measurement model approach in particular is Rating Scale Model (RSM). By examining the quality of this instrument, it would give further information for the next researchers or foreign language instructors who intended to use it for the purpose of measuring students' anxiety in speaking Arabic language.

2.2 Sample

This study was conducted at Central Java, Indonesia. About 46 university students majoring in the Arabic Language Education (ALE) have participated in this study. They were selected using simple random sampling technique for those who were attending the *muhadatsah* or Arabic speaking class. They were at their 3rd year of bachelor degree.

After checking Person Fit statistics, there were 8 students who were their responses did not fit to the model (oufit MnSq more than 1.5) or can be called as misfit person, therefore, those students were excluded from the analysis. The total students included in the analysis were 38 students, which can be considered as pilot testing of the instrument.

2.3 Instrument

There were many established instruments developed to examine foreign language students' anxiety, for instance the well-known scale of Foreign Language Classroom Anxiety Scale (FLCAS) developed by Horwitz et al., [2]. But among those instruments, there were only a few which emphasized on measuring foreign language speaking anxiety. Foreign Language Speaking Anxiety Scale (FLSAS) by Balemir [6] was one of many established instruments focusing on measuring students' anxiety in speaking foreign language.

The FLSAS comprised 28 items with a five-point rating scale, varied from "strongly disagree" to "strongly agree" (score 1 to 5). Some items were written in favorable statements while others were unfavorable. This questionnaire was initially developed in English and was used by the English language learners. To fulfill the need of Indonesian Arabic language learners, it was modified and translated into Indonesian language by the first writer. Some specific word such as "English" was changed to "Arabic" since the questionnaire was aimed at Arabic language learners. The translation process was carried out by an expert in both English and Indonesian language to ensure that the translation does not change the meaning of the statements. The instrument was translated using the back-translation method, involving a forward translation into the target language followed by a reverse translation back into the original language to ensure accuracy and consistency Then, several experts in both English and Indonesian language were asked to check whether or not the translated FLSAS questionnaire could

be understood by the respondents and some modifications were made based on the comments from the experts.

Beside those 28 items in the instrument asking the students' opinion about their anxiety, there are also eleven other questions in the first section which asked them about their demographic information like gender, their latest speaking score, their latest CGPA, the length of time they have spent in learning Arabic language, the previous school they went to and whether or not they ever attend speaking class in their previous schools and the last part in the first section is about their willingness to practice speaking outside the classroom, the place they live in, their parents educational background and the reasons which provoke their anxiety while speaking Arabic language in the classroom.

The questionnaire was constructed based on two domains of anxiety, namely facilitating and debilitating domain of anxiety. The debilitating domain was consisted of four main triggers of anxiety, they are personal reason, teacher's manner in classroom, teaching procedures [11] and testing situation [6]. Besides, there were some items related to their general feeling of Arabic language speaking anxiety.

Here are the items list of the questionnaire as it is adapted and modified to meet the purpose of this study. The item number 1, 24 and 25 are the item for general statement asking about their anxiety. So, it can be concluded that the domains of the modified and translated FLSAS are:

Table 1. Items of the Foreign Language Speaking Anxiety Scale (FLSAS)

No	Construct/Domain	Quantity	Item Number
1	General statement	3	1, 24, 25
2	Facilitating conditions	6	2, 3, 13, 15, 20, 21
3	Personal reason	8	4,6,8,11,14,18,19,28
4	Teachers' manner	3	9,22,23
5	Teaching procedures	5	7,10,16,17,26
6	Testing situation	3	5,12,27

2.4 Scoring Model

The FLSAS is a five-point Likert-type scale, which ranged from "strongly disagree" to "strongly agree". The participants' response from the negative items would be coded using 1 for strongly disagree, 2 for disagree, 3 for neutral, 4 for agree and 5 for strongly agree. Some positive items would be reverse coded using 5 for strongly disagree to 1 for strongly agree.

2.5 Data Analysis

The data underwent analysis using Rasch model approach, employing the Winstep software version 3.73. The following analysis would be carried out to examine the FLSAS using the software output, which include:

1. Summary Statistics,

2. Item Fit,
3. Principal Component Analysis,
4. Rating Scale Analysis
5. Wright Map

3 Method

The data gathered from 38 participants were computed and analyzed using Winstep from Rasch measurement model. Rasch measurement model provided various analysis which can be used to examine the quality of an instrument. Below are the results from the undergone analysis:

3.1 Summary Statistics

Summary statistics of the FLSAS instrument would give the further information about the reliability of both respondents of this study and items of the FLSAS. Below is the table of summary statistics of 26 items in the FLSAS:

Table 2. Summary Statistics of Person and Item (N=38)

	Mean	**Std Deviation**	**Strata**	**Reliability**	**Cronbach Alpha**
Person	-0.32	0.90	4.56	0.91	0.92
Item	0.00	0.80	5.08	0.93	

The table of summary statistics above describes the overall quality of the participants' responses and the instrument. The Cronbach alpha value of the instrument 0.92 indicated that the FLSAS is high internal consistency, can be called as a reliable instrument [12]. The summary statistics also revealed the person reliability as well as the item reliability. Person reliability 0.91 indicated good [13], means that respondents' consistency in answering the FLSAS is good, while item reliability 0.93 indicated as very good, meaning that the quality of items is very good and highly reliable.

The other result revealed in the figure above is both person and item strata. The bigger the person and item strata implied the better the quality of the instrument, meaning that the it is adequate to measure different groups of respondents and items [13]. Here is the formula to calculate the person and item strata:

$$H = \frac{[(4 \text{ x separation}) + 1]}{3}$$

The person strata from this study was 4.56 indicated that the respondents of the study can be divided up to 4 different groups, while the item separation is 5.08. This result of the person and item strata are indicated as very good spread [13].

3.2 Item Fit

The table below shows the measurement of each item on the questionnaire. Sumintono and Widhiarso [13] explained that item fit table from Rasch Model described whether the items fit the ideal model (i.e. the Rasch model). If the item does not fit the model, or can be called as misfit item, indicating that there is misconception from some of the respondents towards the misfit item.

The acceptable value of outfit mean-square, outfit z-standard and point measure correlation should be checked to determine the item fit:

a. Outfit mean-square value (MNSQ) must be between 0.5 < MNSQ < 1.5
b. Outfit Z-standard value (ZSTD) must be in the range of -2.0 < ZSTD < +2.0
c. Point Measure Correlation value (Pt Measure Corr) must be between 0.4 < Pt Measure Corr < 0.85 [13]

The value of Pt Measure Correlation which falls between 0.28 to 0.85 is also acceptable. Researchers are suggested to check the items with negative value of Pt Mean Correlation, or items which the value of outfit MNSQ and ZSTD falls outside the accepted range, whether it is need to be eliminated or revised [14].

The Table 3 below shows the item number 2 which represents the facilitating dimension of anxiety ("*Kegelisahan saya berkurang ketika berbicara bahasa Arab di depan orang yang saya kenal*" or I feel less anxious when speaking Arabic in front of people I know) did not fulfil the acceptable value of Outfit MNSQ (1.83) and Outfit ZSTD (3.2), while the value of Pt Measure Corr (0.31) was accepted [14]. While item number 13, which also represent the facilitating dimension of anxiety ("*Saya merasa tenang untuk berbicara bahasa Arab apabila muridnya sedikit* or I feel comfortable speaking Arabic when there are few students"), did not met the requirement of the three criteria of fit statistics. The value of its Outfit MNSQ was 1.76. Outfit ZSTD 2.9 and Pt Measure Corr 0.05. Therefore, item number 2 and 13 will be eliminated because both items did not fit to be used to measure students' speaking anxiety.

Table 3. Item Fit of the Foreign Language Speaking Anxiety Scale (FLSAS)

Item	Measure	Standard Error	Outfit MNSQ	Outfit ZSTD	Pt Measure Corr
2	0.42	0.20	*1.83*	*3.15*	0.31
13	0.58	0.20	*1.76*	*2.90*	*0.05*
22	-0.54	0.19	1.34	1.52	0.38
16	0.88	0.21	1.28	1.21	*0.20*
14	-1.63	0.21	1.21	0.92	0.28
21	1.02	0.22	1.15	0.72	0.53
28	-0.31	0.19	1.17	0.84	0.44
3	1.17	0.22	1.13	0.60	0.41
6	-1.12	0.20	1.02	0.15	0.67
23	0.80	0.21	1.04	0.24	*0.26*
15	0.84	0.21	1.04	0.23	0.42
25	-1.00	0.20	1.02	0.16	0.58
9	0.98	0.22	0.94	-0.17	0.69
5	-0.69	0.20	0.98	-0.04	0.59
18	0.58	0.20	0.96	-0.12	0.53

24	-0.24	0.19	0.93	-0.26	0.68
8	-0.58	0.19	0.92	-0.29	0.84
20	0.98	0.22	0.92	-0.28	0.57
4	-0.61	0.20	0.87	-0.54	0.62
11	-0.69	0.20	0.88	-0.49	0.77
1	-0.65	0.20	0.86	-0.61	0.75
7	0.18	0.20	0.80	-0.94	0.65
17	-0.12	0.19	0.81	-0.86	0.61
12	-0.39	0.19	0.75	-1.21	0.82
10	0.22	0.20	0.64	-1.84	0.64
27	-0.46	0.19	0.65	-1.79	0.73
26	0.38	0.20	0.64	-1.86	0.42
19	0.03	0.20	0.59	*-2.19*	0.64

The rest of items have met the criteria of Outfit MNSQ, Outfit ZSTD and Pt Measure Corr except for item 16 and 23 which value of the Pt Measure Corr is less than 0.28. But both items will be kept because the value of Outfit MNSQ and Outfit ZSTD have fulfilled the requirement. To sum up, out of 28, there are 26 items which have met the requirements. All these 26 items are used to measure the students' anxiety in speaking Arabic language.

3.3 Principal Component Analysis (PCA)

Rasch analysis provided us the Principal Component Analysis to check the unidimensionality of the test [15] and decide whether or not the items in the FLSAS are measuring the same variable, which is the students' anxiety in speaking Arabic language.

Table 4. Principal Component Analysis of FLSAS

	Empirical		**Modeled**
Total raw variance in observation	51	100%	100%
Raw variance explained by measure	25	49%	48.2%
Raw unexplained variance (total)	26	51%	100%
Unexplned variance in 1st contrast	3.6	7%	13.7%

The Table 4 of PCA above describes the unidimensionality of FLSAS. The first value we need to check is the raw variance explained by measure that must be above 40%. The value of 49% has exceeded the minimum requirement of it. Meaning that the instrument is adequate to be employed which can cover range of anxiety of respondents. The next value needs to be checked is noise or the unexplained variance in the 1st contrast that must not exceed 15% [14]. The table of PCA above depicts that the value is only 7 % which does not exceed 15%. It means that the noise found in the instrument is still acceptable and did not disturb the measurement. Whereas the Eigen value of the instrument in unexplained variance is 3.6, indicating there are one item that come from another dimension.

3.4 Rating Scale Analysis

Ensuring the quality of the instrument prior to conducting the actual study is imperative, as the reliability of the data hinges upon it. Test or instrument developers should carefully choose the appropriate type of response option to use depending on the purposes of their measurement. According to Bond and Fox [16], some people will employ a simple "yes/no", or frequency scale like "never/sometimes/often/always" and others may choose a Likert type scale from "strongly agree to strongly disagree". However, the empirical evidence should be provided in order to ensure that the participants could tell the differences between the options given in the instrument.

Linacre [17] emphasized the necessity of conducting an initial phase before advancing with further analysis by exploring the functionality of rating scale categories. Therefore, the Rasch approach is used to provide empirical insights about its quality.

Table 5 depicts the analysis of rating scale in the rubric that put four rating scales in every item. It shows that each rating has more than ten responses, the least one is 'strongly agree' (53 times), whereas 'disagree' the highest choice (334 times).

Table 5. Rating Scale Analysis

Score & Rating	Observed count (%)	Observed Average	Outfit MNSQ	Andrich Threshold
1 = strongly disagree	105 (11%)	-1.66	1.07	None
2 = disagree	334 (34%)	-0.90	1.09	-2.49
3 = neutral	260 (26%)	-0.32	1.09	-0.29
4 = agree	236 (24%)	0.71	0.82	0.34
5 = strongly agree	53 (5%)	1.33	0.93	2.44

Winstep helps us investigate whether or not the five rating scales categories in the instrument should be combined or separated, also if these categories can be understood by the participants. It also tells us about the value of the observed average that rose steadily from negative to positive value which shows normal response from the respondents.

Fit statistics also provides another criterion to examine the quality of rating scale. The outfit mean-square value that is bigger than 2 implies that the specific category is showing more noise that may affect the measurement process [16]. Table 5 shows that all the Outfit MNSQ values were less than 2, meaning that the quality of five rating scales of the FLSAS is good.

The Andrich Threshold value which is displayed in the last column should be considered as well. The *s* values, which imply the gap between two categories, must fall within $1.4 < s < 5.0$ [14]. The distance from the first rating (strongly disagree) to next rating (disagree) is still in the acceptable range. The gap from the second rating (disagree) to the third (neutral) is also still in the acceptable range. Next is the gap between the third rating (neutral) to the fourth (agree) is less than 1.4. While the gap from the fourth rating (agree) to the fifth (strongly agree) falls in the acceptable range. One of the rating scales need to be collapsed because the students seem cannot clearly

differentiate it, where the possibility of four rating without using 'neutral' choice will be more appropriate. Overall, the suitable rating scale categories to be employed is four-rating scale.

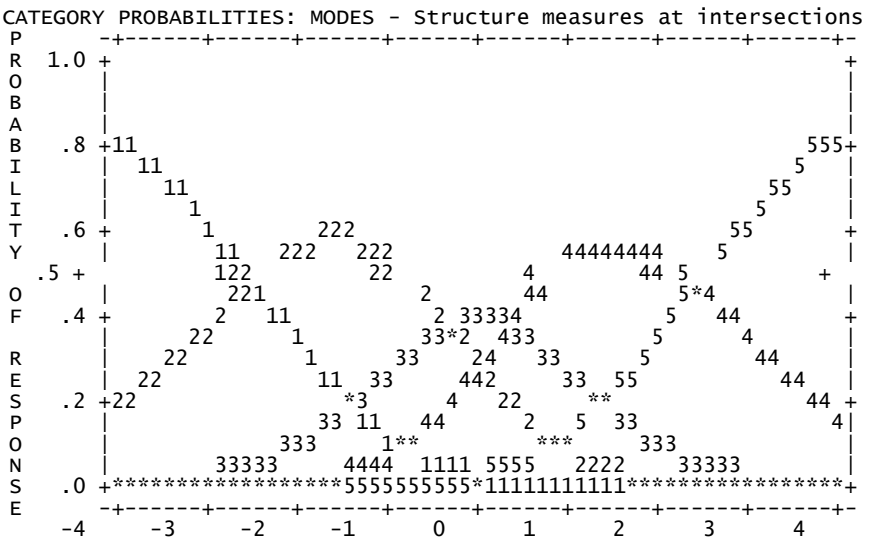
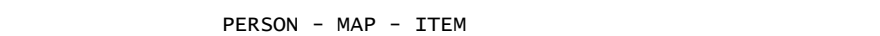


Fig. 1. The rating scale of item response probability for the instrument.

The figure illustrates that all the rating scales peak can be clearly seen except for the rating scale 3 (neutral), showing that respondents cannot clearly differentiate it with the other choices. Based on the result above, the rating scale categories that will be used in the questionnaire are four choices, start from strongly disagree, disagree, agree to strongly agree.

3.5 Wright Map

Wright Map analysis describes the spread of both items and person on the same scale. Both items and persons are placed along the map based on their ability and difficulty calculations, respectively [15]. The spread of the items is shown at the right side, while the spread of person is illustrated at the left side of the map. The more difficult item to agree with, the higher it will be located on the map, and the easier items will be located at the bottom of the map. Thus, the item considered as difficult if it is located at the top of the map which means the students did not find it as anxiety-provoking situation. On the contrary, items at the bottom of the maps are considered as easy item, meaning the respondents often find it as the anxiety-provoking situation.



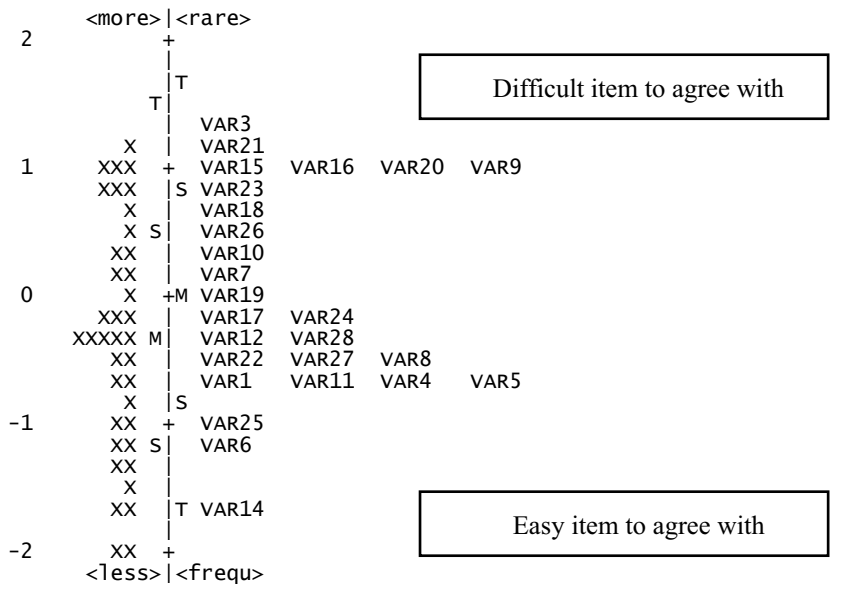


Fig 2. Wight Map of the FLSAS.

The figure above depicts the logit scale which is displayed in the middle of the map, as a joint scale of person ability and item difficulty [15]. Boone suggested to look at and compare the mean of items and persons which are illustrated along the logit scale [18]. The letters "M" on the right side of the logit scale is the mean for the items which is located higher than the "M" for persons in this map. This indicates that many of participants tend to answer "agree" or "strongly agree" to the statements. The item labelled as var14 which is placed at lower is considered as the situation that easily provoke students' anxiety, while var3 on the top of the map is perceived as the least situation that affect their anxiety.

4 Method

The findings provided the reliability and validity analysis of the instrument such as Summary statistics, Item Fit, Principal Component Analysis (PCA), Rating Scale Analysis and Wright Map that will be discussed further in this section. First is the Summary Statistics which showed us the reliability of the FLSAS. Reliability tells that an instrument can produce stable and consistent score even with small number of respondents [19]. Meaning that the instrument can always give the consistent result although it is repeatedly used many times. The Summary Statistics showed that Cronbach alpha value of FLSAS is highly reliable (0.92). While the item and person reliability both showed very good reliability.

Meanwhile the item fit gives further analysis on how the items fit the standard model. It can be decided by referring to the value of Outfit MNSQ, Outfit ZSTD and Point Measure Correlation [13]. The result from item fit analysis shows that among 28

items of the FLSAS, 2 items were found to be misfit. Therefore, both are eliminated to improve the validity of the questionnaire. While the other 26 items have acceptable fit statistics indices, indicating that they can be used to measure students' anxiety in speaking Arabic.

The other analysis, Principal Component Analysis (PCA), is one among other important analysis to identify whether the instrument can really measure only one dimension, which in this context is speaking anxiety. The value of raw variance explained by measure indicated that all the 26 items in the Foreign Language Speaking Anxiety Scale by Balemir [6] are going to the same direction and are measuring the same dimension, except for one item. While the noise found in the instrument is less than 15% which means it does not disturb the measurement of speaking anxiety.

Researchers should check the quality of the response categories when a new instrument is developed or when an established one is applied to a new population [16]. As the FLSAS was used in the new population, the rating scale analysis was carried out. The analysis gave empirical evidence whether or not the five-point rating scale categories in the FLSAS are easily understood by the participants. The result showed that the respondents have difficulty to tell the difference between the "neutral" and "agree", therefore, the two categories were collapsed. Consequently, the rating scale category used in the questionnaire is four, namely strongly disagree, disagree, agree and strongly agree.

Furthermore, the Wright map analysis further explained the information about the hierarchy of both item and person. The map helps to identify which item is the easiest statement and the most difficult one to agree. In this study, the respondents considered that item labelled as var14 was the easiest item to agree which means they agree that the item represents the factor which easily provoke their anxiety. The item var14 was about personal reason which showed that they are afraid if they cannot convey their idea in a good way that their classmates can understand ("I feel anxious if I cannot convey my ideas well when speaking Arabic or *Saya merasa cemas apabila saya tidak bisa menyampaikan ide saya dengan baik ketika berbicara Bahasa Arab*").

While the most difficult item to agree was var3 which means that the item contain the factor that do not easily affect their anxiety. The var3 was related to facilitating condition of anxiety which showed that students prefer to know and study the material that is going to be learned in the speaking class ("I feel very relaxed about speaking in English class when I study the planned contents before the class" or "*Saya merasa lebih santai saat saya sudah mempelajari terlebih dahulu materi yang akan dipelajari di dalam kelas muhadatsah*"). This information will give insights to teachers to create the learning environment that will help students lower their anxiety in speaking Arabic language.

5 Conclusion

It can be concluded that the translated and modified edition of the FLSAS is adequate to measure students' anxiety in speaking Arabic. This conclusion is supported by an

analysis of summary statistics, item fit, principal component analysis, rating scale analysis, and Wright map. There are 2 items eliminated because they did not meet the requirement of the item fit. The high values of Cronbach's alpha, as well as item and person reliability, indicate strong reliability, affirming the suitability of the FLSAS for measuring speaking anxiety among students. Foreign language teachers may find it beneficial to utilize this FLSAS in their classes to identify the students' speaking anxiety to give them insight about to what extent do their students experience anxiety.

References

1 P. MacIntyre, "Language Anxiety: A Review of the Research for Language Teachers," in *Affect in Foreign Language and Second Language Learning: A Practical Guide to Creating a Low-Anxiety Classroom Atmosphere*, D. J. Young, Ed. McGraw-Hill Education, pp. 24–45. (1999)
2 E. K. Horwitz, M. B. Horwitz, and J. Cope, "Foreign Language Classroom Anxiety," *Mod. Lang. J.*, vol. 70, no. 2, pp. 125–132 (1986)
3 V. R. Doqaruni, "A Quantitative Action Research on Promoting Confidence in a Foreign Language Classroom : Implications for Second Language Teachers Quantitative Action Research on Promoting Confidence in a Foreign Language Classroom," *i.e. Inq. Educ.*, vol. 5, no. 1 (2014)
4 L. Kristal, *A. B. C. of Psychology*. Michael Joseph Ltd, (1981)
5 M. J. Riasati, "Language learning anxiety from EFL perspective," *Middle-East J. Sci. Res.*, vol. 7, no. 6, pp. 907–914 (2011)
6 S. H. Balemir, "The Sources of Foreign Language Speaking Anxiety and the Relationship Between Proficiency Level and Degree of Foreign Language Speaking Anxiety," Bilkent University. (2009)
7 R. . Fajri, "Problematika Kemampuan Berbicara Bahasa Arab Mahasiswa Jurusan Pendidikan Bahasa Arab Fakultas Ilmu Tarbiyyah Dan Keguruan UIN Sunan Kalijaga Yogyakarta Tahun Akademik 2014/2015," Universitas Islam Negeri Yogyakarta. (2015)
8 T. Hidayati, "Student Language Anxiety in Learning English: Examining non-English Major Students in Rural Area Hidayati, Student Language Anxiety in Learning English," *IJELTAL (Indonesian J. English Lang. Teach. Appl. Linguist.*, vol. 2, no. 2, p. 95 (2018)
9 M. T. Apple, "A Critique to 'Using Rasch Analysis to Create and Evaluate a Measurement Instrument for Foreign Language Classroom Speaking Anxiety,'" *JALT J.*, vol. 35, no. 2, pp. 205–209 (2013)
10 K. Arasinah, A. R. Bakar, H. Ramlah, A. Soaib, and H. Zaliza, "Using Rasch Model and Confirmatory Factor Analysis to Assess Instrument for Clothing Fashion Design Competency," *Int. J. Soc. Sci. Humanit.*, vol. 5, no. 5, pp. 418–421 (2015)
11 B. Aydin, "A Study of Sources of Foreign Language Classroom Anxiety in Speaking and Writing Classes," Anadolu University. (1999)
12 L. Cohen, L. Manion, and K. Morrison, *Research Methods in Education*. London: Routledge. (2007)
13 B. Sumintono and W. Widhiarso, *Aplikasi Pemodelan Rasch Pada Assessment Pendidikan*. Cimahi: Trim Komunikata. (2015)
14 A. A. Abdul, M. S. Masodi, and A. Zaharim, *Asal Model Pengukuran Rasch*. Universiti Kebangsaan Malaysia. (2013)
15 T. G. Bond, Z. Yan, and M. Heene, Applying the Rasch Model: Fundamental Measurement in the Human Sciences, Fourth. Routledge. (2021)

16 T. Bond and C. Fox, Applying the Rasch Model: Fundamental Measurement in the Human Sciences, 3rd ed. New York: Routledge. (2015)
17 J. M. Linacre, "Understanding Rasch Measurement: Estimation Methods for Rasch Measures," *J. Outcome Meas.*, vol. 3, no. 4, pp. 381–405 (1999)
18 W. J. Boone, J. . Staver, and M. . Yale, *Rasch Analysis in the Human Sciences*. Dordrecht: Springer. (2014)
19 J. Creswell, Educational Research: Planning, Conducting and Evaluating Quantitative and Qualitative Research. Boston. (2012)