International Journal of Multicultural and Multireligious Understanding

http://ijmmu.com
editor@ijmmu.com
ISSN 2364-5369
Volume 8, Issue 6
June, 2021
Pages: 413-426

Development and Validation of Public Speaking Performance Instruments based on Rasch Model

Tisya Nindi Nugraha Ramandha; Wardani Rahayu; Gaguk Margono, Erwin Sulaeman

Department of Educational Research and Evaluation, Universitas Negeri Jakarta, Indonesia

http://dx.doi.org/10.18415/ijmmu.v8i6.2771

Abstract

This study aims to see the ability of students' public speaking performance, so that the development and validation of the instrument is carried out for the evaluation of the learning. This type of research is quantitative through a survey of 976 students of communication study programs at universities in the DKI Jakarta area. The results show that the instrument developed from the modification of 43 items to 33 items has met the psychometric validity requirements with Rasch modeling. The monotonic is fulfilled by each item on the public speaking instrument through the Andrich Threshold criteria, so that the instrument with the developed response scale does not provide answers that are confusing to respond.

Keywords: *Development; Validation; Public Speaking; Rasch Model*

Introduction

The prospect of speaking in public evokes fear and anxiety for many individuals, and speech anxiety often poses a serious problem (Pfister & Robinson, 2010). Numerous occupations require people to speak publicly, at least on occasion, and for many individuals the fear and anxiety that it evokes can greatly impair performance (Slater et al., 2006). While anxiety tends to impair performance on difficult cognitive tasks generally, it poses especially difficult problems for public speaking (Noor et al., 2021).

Although feeling anxious while speaking in public is a normal occurrence (Glassman et al., 2016), 33% of the population experience severe and incapacitating anxiety in these situations (De Oliveira et al., 2012). In addition to debilitating speech-related anxiety, there is evidence that individuals with public speaking anxiety (PSA) experience impaired speech performance, which affects their social, occupational, and educational functioning(Hindo & González-Prendes, 2011). Simply stated, specific objects, thoughts, surroundings, situations, or activities come to elicit anxiety and fear responses (Noor et al., 2021; Slater et al., 2006).

A barrier for everyone who must speak publicly is public speaking anxiety, describes the most heavily examined and researched communication constructs (1909 to present) by the discipline of communication studies (Jones et al., 2012). Estimates suggest as many as 80% of the population experience context-based communication apprehension with over 70% related to the specific context of

public speaking (Wadkins, 2021). (Poeschl, 2017; Thomson & Rucker, 2002) estimated that 1.3 million students enroll each year in public speaking courses, translating into as many as 910,000 students experiencing anxiety. A major challenge for students involves learning to manage the anxiety associated with the speaking context. Most classes recommend students practice a speech to reduce the anxiety levels.

In Indonesia, it is rare to find research on public speaking, but there are several studies from (Habiby, 2012) which discuss the effectiveness of public speaking training. For this reason, the evaluation of learning can be done from the student's point of view, so that a measurement instrument is needed on public speaking performance in order to determine the development of students' abilities (Cirik et al., 2015).

Many questionnaires have been designed and used successfully in many countries, such as research related to measuring anxiety in speaking English using instruments related to the Public Speaking Class Anxiety Scale (PSCAS) (Ismail et al., 2019), Instrument Self-Statement During Public Speaking Scale (Hofmann & DiBartolo, 2000), Instrument Personal Report of Public Speaking Anxiety (PRPSA) (Nicolini & Cole, 2019), instrument Speaking Competencies from the Public Speaking Competency Instrument (Schreiber et al., 2012; Thomson & Rucker, 2002), Instrument Public Speaking Competency (Schonert-Reichl et al., 2009), and instrument Trait anxiety, experience, and the public speaking state responses of Finnish university students (PSSR) (Pörhölä, 1997).

Some of these instruments will be modified, namely the PSCAS instrument developed by (Noor et al., 2021) to measure and identify the specific causes of public speaking anxiety. The PSCAS instrument consists of four dimensions, namely Communication Apprehension, Fear of Negative Evaluation, Test Anxiety, dan General Anxiety of Language. Each dimension consists of five items with options always, often, sometimes, rarely and never (Fenty & Anderson, 2014; Goetz et al., 2003). The PRPSA instrument consists of 34 items. Instruments related to Public Speaking Competency Instrument (PSCS) consists of measuring seven dimensions, namely speech introduction, organization, supporting material, speech conclusion, verbal delivery, nonverbal delivery, and general competence (Schreiber et al., 2012). PSSR instrument using three measuring dimensions, namely: State Anxiety-Enthusiasm, Anxiety-Enthusiasm Behavior, and Reticence-Willingness to Communicate.

These instruments (PSCAS, PRPSA, PSCS, and PSSR) allow them to be adapted into Indonesian version of performance public speaking instruments for the needs of students in the communication study program. The instrument will be the latest in this research, by prioritizing the characteristics in the Indonesian version. Investigate the adapted dimensions and attempt to validate the questionnaire based on its psychometric characteristics.

Method

This type of research is quantitative through a survey adopted from the post-postivism paradigm with a questionnaire method in the form of a non-test instrument.

Participant

The respondents in this study were all students who took the communication study program in the campus area in DKI Jakarta. Number of respondents collected is 976 students to conduct a survey on public speaking instruments. The determination of respondent criteria was the same as done in the study (Skriner et al., 2017), the selection criteria were based on previous findings regarding the demographic variables differentiated respondent perception of instrument.

Instrument

Evaluation is carried out at each stage to minimize the level of instrument errors or deficiencies. The stages of expanding this instrument have reached the development stage and validation. The result of the concept study obtained in the conceptual definition of instrument public speaking Indonesian version in this study is the implementation of instrument Public Speaking Class Anxiety Scale (PSCAS), *Personal Report of Public Speaking Anxiety* (PRPSA), Public Speaking Competency Instrument (PSCS), and Public Speaking State Responses (PSSR). The modified into seven dimensions of measurement, namely: *speech introduction, organization, supporting material, speech conclusion, verbal delivery, nonverbal delivery, and general competence.* The measurement scale of the instrument consists of 4 options namely: never, rarely, often, and always.

Table 1:Guidelines of Instrument

Dimension	No. Item	Total
speech introduction	1,2,3,4,5,6,7,8,9	9
organization	10,11,12,13,14,15,16	7
supporting material	17,18,19,20,21,22,23	7
speech conclusion	24,25,26	3
verbal delivery	27,28,29,30,31,32,33,34	8
nonverbal delivery	35,36,37,38	4
general competence	39,40,41,42,43	5
Total		43

Prosedure

Preparation of guidelines for translating and modifying psychological instruments from (Gudmundsson, 2009) based on standards for psychological education and testing (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education) can be seen in Figure 1 below:

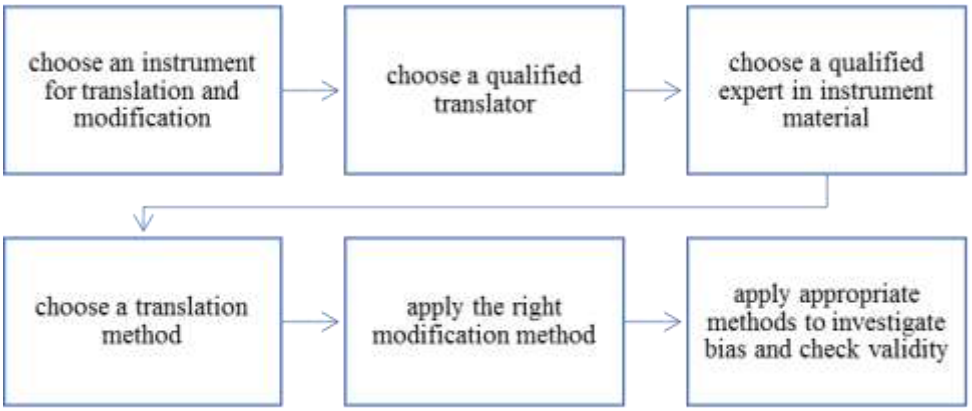


Figure 1: Guidelines for translating and adapting

The validation process was carried out by providing an instrument that had been prepared with 43 items, conceptual and operational definitions, and instrument guidelines to each respondent. The results of the instrument repair were inseparable from expert member checking until the arrangement of instruments according to dimensions, indicators, and items were stated as appropriate. The next step was that the Instrument, which had been finalized, was brought back to the experts for item evaluation using a measurement scale with the following options:

Table 2: Measurement Scale

Category Options			
Never	Rarely	Often	Always
1	2	4	5

Based on Rasch modeling, the validity analysis is generated using Winsteps software. The information provided is in the form of information on the suitability of items with criteria, namely: according to (Cheng et al., 2011; Yan & Mok, 2012) explain that when the data match the model's expectations, the infit and outfit mean square (MNSQ) in the Rasch analysis. Items with MNSQ values ranging from 0.50 to 1.50 to indicate a fairly good fit of the data model.

Another opinion of (Lundgren-Nilsson et al., 2014) stated that standard match residues between -2.5 and +2.5 (99% confidence interval) indicated adequate fit of each individual and item residue. Reliability is viewed as an index of the accuracy of the measures generated by the Rasch model. A value with a minimum of 0.7 is included in the sufficient category while the correlation value is moderate to strong between 0.52-0.90.

Finding

Rasch modeling requirements analysis testing is important to note, namely: unidimensional, monotonization, and fit item:

Unidimensional

Unidimensional calculations for the four response category scales are presented in Table 3 below:

Table 3: Unidimensionality on a scale of four response categories

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = ITEM information units

	Eigenvalue	Observed		Expected
Total raw variance in observations =	65.4540	100.0%		100.0%
Raw variance explained by measures =	28.4540	43.5%		43.1%
Raw variance explained by persons =	14.2432	21.8%		21.6%
Raw Variance explained by items =	14.2108	21.7%		21.5%
Raw unexplained variance (total) =	37.0000	56.5%	100.0%	56.9%
Unexplned variance in 1st contrast =	2.6960	4.1%	7.3%	
Unexplned variance in 2nd contrast =	2.3729	3.6%	6.4%	
Unexplned variance in 3rd contrast =	1.9289	2.9%	5.2%	
Unexplned variance in 4th contrast =	1.7761	2.7%	4.8%	
Unexplned variance in 5th contrast =	1.6282	2.5%	4.4%	

The results obtained in Table 3 are shown by "Raw variance explained by measure" located in the "observed" column with a value of 43.5% greater than 20%, so that the instrument meets the unidimensional requirements (Yu, 2015). Eigenvalue units, namely: 2.7, 2.4, 1.9, 1.8, and 1.6, unexplained variances, namely: 4.1%, 3.6%, 2.9%, 2.7%, and 2.5%, are in the 3-5% category with very strong criteria (Ozgul et al., 2018). The percentage variance does not exceed 15% (Sinnema et al., 2016). Thus empirically the instrument is unidimensional and builds construct validity.

Monotonization

Instruments with a scale of four response categories of frequency scale types from the rating scale are shown in Table 4 below:

Table 4: Frequency scale of four response categories

Alternative Answer Options			
1	2	3	4
Never	Rarely	Often	Always

The four-category response scale in Table 4 negates the middle value function, so that student responses are wiser, produce a more precise ranking (Cronbach, 1950; Fordham, 1981; Green, 2010). It can be seen in the "Observed Average" column which shows the measurements in each response category:

Table 5: Monotonic properties through the Andrich threshold

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE	
1	1	1143	3	-.83	-.77	.97	.98	NONE	(-3.62)	1
2	2	10411	29	.17	.20	.96	.97	-2.48	-1.11	2
3	3	14546	40	1.21	1.16	1.00	1.05	.33	1.26	3
4	4	10012	28	2.43	2.46	1.03	1.03	2.14	(3.35)	4

The results of the analysis on the rating scale show that there is an increase in the value in the Observed Average column from negative to positive (Andrich, 2011). The logit value starts from -0.83 for category one choice (never) to 2.43 for category four choice (always). There is a monotonic increase in logit value, indicates that the respondent is able to discriminate between category choices and verify the level of response that agrees.

The Andrich threshold value based on Table 5 moves monotonically from NONE to the negative logit direction (-2.48), followed by (0.33) logit, and leads to a positive logit (2.14). This monotonic movement illustrates that the items have conformity with the choice of response category for measurement. Other information can be found in the "options response functions" through the Category Characteristic Curves (CCC) analysis in Figure 2 below:

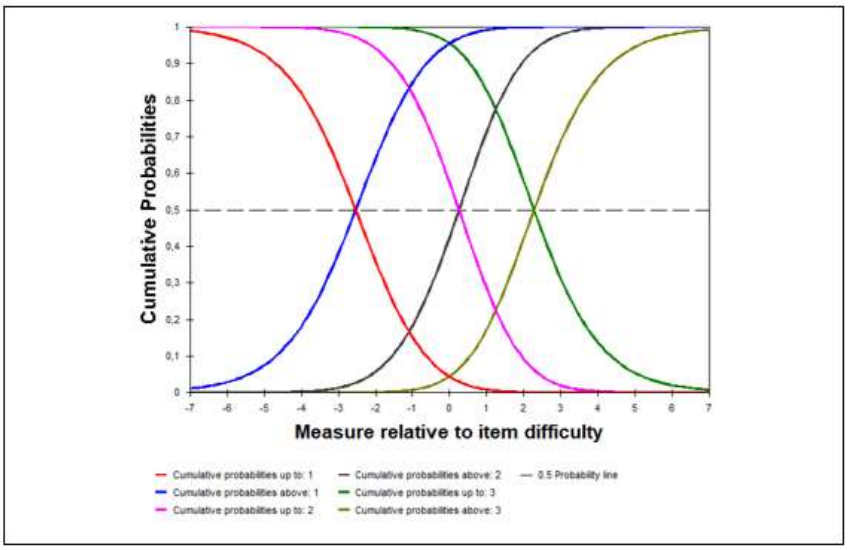


Figure 2: Category Characteristic Curves scale of four response categories

Figure 2 provides information on the boundaries between the specified response categories. Threshold item values are monotonically determined from low to high and presented in the form of the curve. The probability of the first curve (1) intersects the probability of the second curve (2), and finally by the third curve (3). This can be interpreted that the value of the specified threshold item is in the good category.

Fit Item

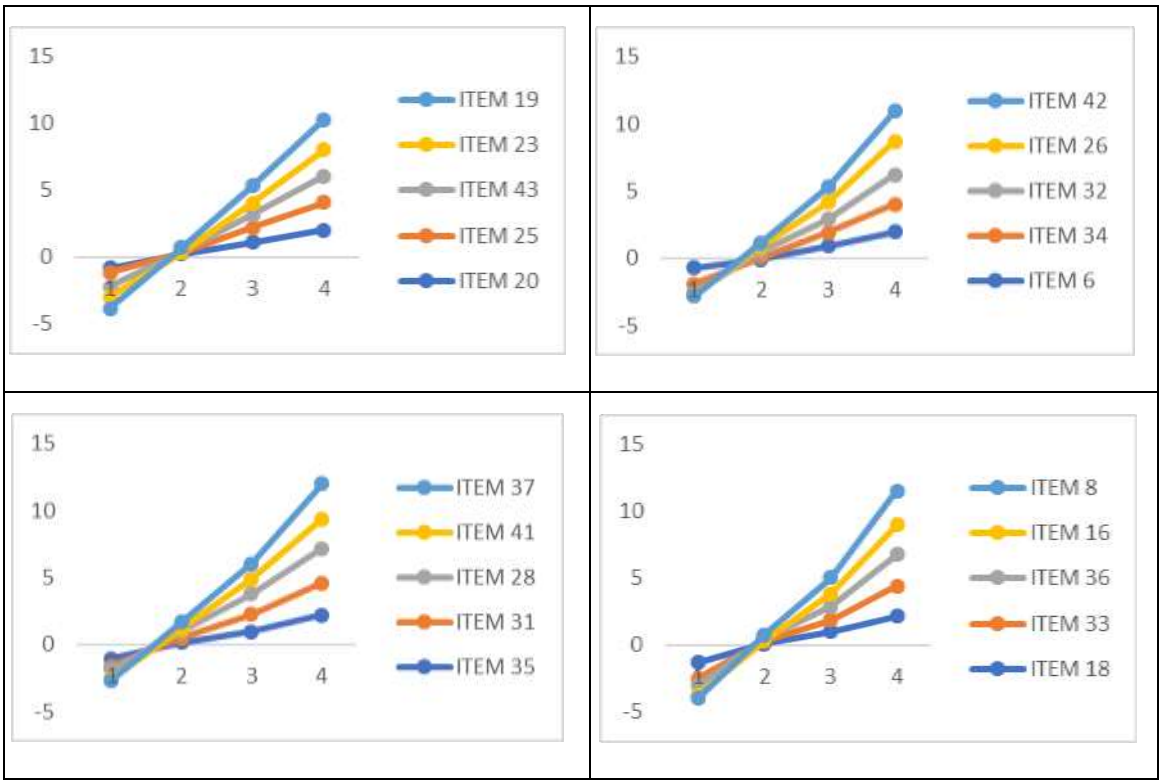
The instrument tested 43 items, 37 items were found to be fit and six items were misfit. Further information can be seen in Table 6 below:

Table 6: Items fit and items do not fit on a scale of four response categories

PT-Measure Correlation (Fit item)	***MisFit***	***Fit item***
0,84 logit to 0,68 logit	B9, B10, B11, B13, B15, B27	B1, B2, B3, B4, B5, B6, B7, B8, B12, B14, B16, B17, B18, B19, B20, B21, B22, B23, B24, B25, B26, B28, B29, B30, B31, B32, B33, B34, B35, B36, B37, B38, B39, B40,B41, B42, B43
Total	**6**	**37**

Table 6 provides information about the fit of the item with the model. Items that match the model on a scale of four response categories are 37 items, the remaining six items do not fit the model, namely: B9, B10, B11, B13, B15, and B27. The MNSQ values obtained are sorted from 0.76 logit to 1.33 logit and the PT-Measure Correlation value is in the range 0.84 logit to 0.68 logit. Items that don't fit, no repairs were made but it was decided not to use or drop out.

It is important to know the monotonic movement of each measured item, pay attention to Figure 3 below:



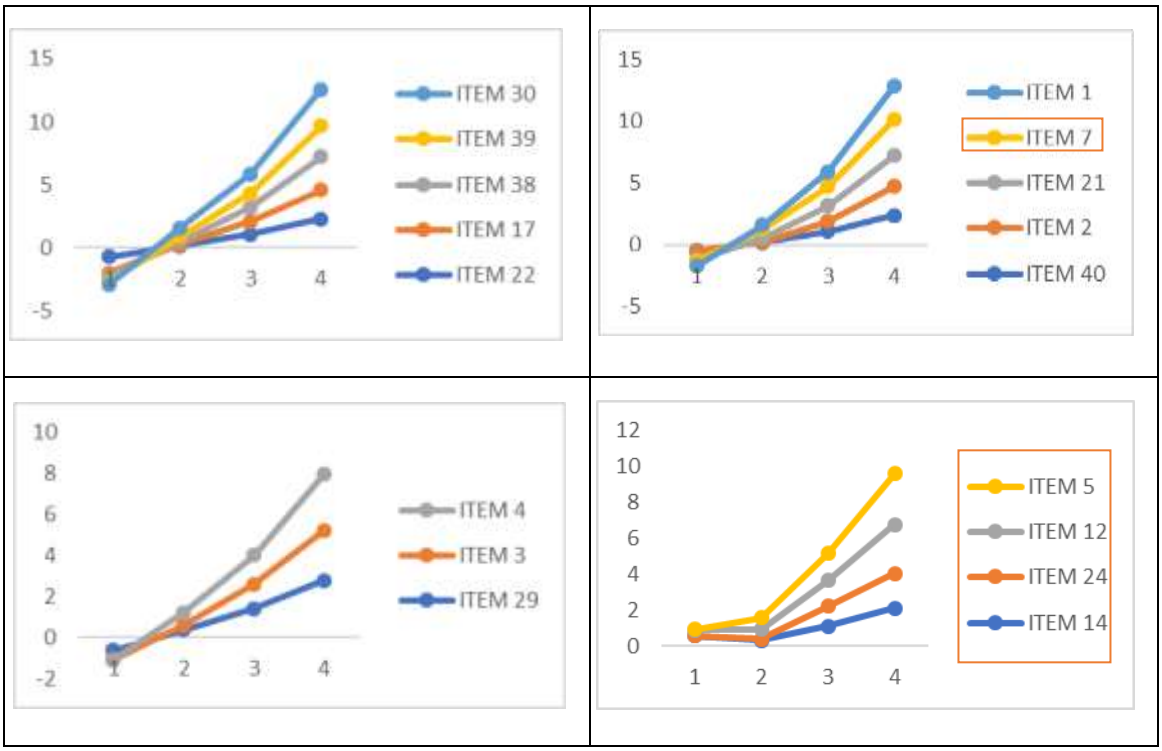


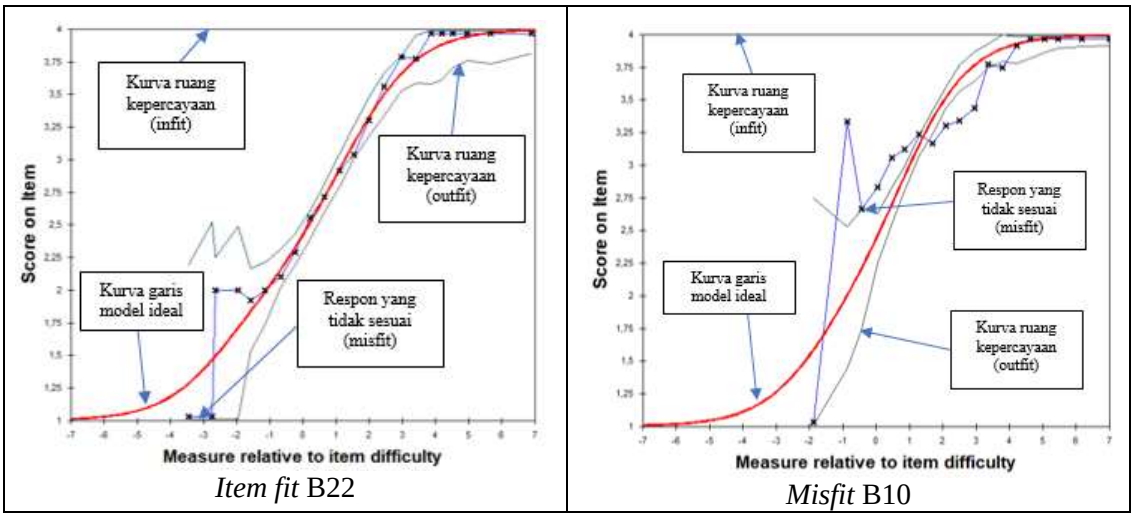
Figure 3: Monotonic Andrich threshold on a scale of four response categories

Based on Figure 3 it shows 37 items on a scale of four response categories, Item analysis is described in detail to show that a four-category response scale that excludes the middle function is more likely to be selected. Description of the items of student responses in distinguishing between the choices 'never', 'rarely', 'often', and 'always'. The findings are based on Figure 3 the student's response style is clearly observed. The line graph for each item is monotonically increasing which shows that the average is the highest between the two response categories, namely: 'often' and 'always'.

The highest average score fell on item B7 on the “speech introduction” dimension with the statement "I am not satisfied with my performance". The four-category response scale stands out on the items with the care dimension. This shows that according to (Kupana, 2015; Lapoint & Butty, 2009) student response choices based on perspectives on individual and group differences. Student responses on a scale of four response categories were consistent in reflecting higher agreement (Moors, 2008).

Of the 37 items, there are four items that do not meet the Andrich threshold requirements, namely: items B5, B12, B14, B24. The four items have a positive increase in movement, but do not meet the Andrich threshold. The four non-conforming items were isolated as an ordinal scale response category (Andrich, 2011), while the other 33 items show evidence of using a four-category response scale that is ordinal.

Other evidence can be explained through the item characteristic curves representing fit and unfit items, which examines the relationship between items and latent traits, and student responses. Figure 4 presents a grammatical and systematic relationship between these quantities as follows:



Item fit B22 *Misfit* B10

Figure 4: Item fit B22 and item not fit B10 on a scale of four response categories

An illustration of item mismatch can be displayed through Figure 4, where the curve shows one of the very good items, namely: item B22 lies in the "supporting material" dimension with the statement "My listeners seemed to be interested in the topic of my presentation", because it meets MNSQ, ZSTD, and Pt Measure Corr. Item match is indicated by the red line curve, so that it can be indicated that the item fits the model and can be used. The curve for the unfit item is represented by one of the items, namely: item B10 is in the "organization" dimension with the statement "I found it uncomfortable to be the center of attention". MNSQ, ZSTD, and Pt. Measure Corr. standards are not met in item B9. It was concluded that item B9 was declared unsuitable for the model and the item could not be used or discarded.

Psychometric validity of the instrument with a scale of four response categories

The validity of the instrument with a scale of four response categories from the data of 976 student responses on 43 items. The data obtained fit person as many as 976 student responses and 37 fit items, further information is presented in Table 7 below:

Table 7: Summary Statistics of the four response categories

SUMMARY OF 976 MEASURED (EXTREME AND NON-EXTREME) PERSON

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	108.2	37.0	1.19	.28				
P.SD	18.4	.0	1.33	.09				
S.SD	18.4	.0	1.33	.09				
MAX.	148.0	37.0	7.11	1.83				
MIN.	50.0	37.0	-3.23	.24				

REAL RMSE .30 TRUE SD 1.29 SEPARATION 4.24 PERSON RELIABILITY .95
MODEL RMSE .29 TRUE SD 1.29 SEPARATION 4.48 PERSON RELIABILITY .95
S.E. OF PERSON MEAN = .04

PERSON RAW SCORE-TO-MEASURE CORRELATION = .98
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .96 SEM = 3.86

SUMMARY OF 37 MEASURED (NON-EXTREME) ITEM

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	2855.4	976.0	.00	.05	.99	-.3	1.01	.1
P.SD	212.1	.0	.55	.00	.12	2.8	.13	2.7
S.SD	215.0	.0	.56	.00	.12	2.8	.13	2.8
MAX.	3180.0	976.0	1.20	.05	1.31	6.4	1.28	5.7
MIN.	2391.0	976.0	-.86	.05	.77	-5.7	.78	-4.6

REAL RMSE .05 TRUE SD .55 SEPARATION 10.45 ITEM RELIABILITY .99
MODEL RMSE .05 TRUE SD .55 SEPARATION 10.67 ITEM RELIABILITY .99
S.E. OF ITEM MEAN = .09

Summary Statistic is the output table of Winsteps version 4.0.1 for analysis of the Rasch model. Table 7 provides information on the results of internal reliability. Fit statistics or information from the reliability index in measuring logit which describes the quality of the instrument:

Person Realiability **dan** ***Item Realiability*** in the summary statistics table, person reliability index with a value of 0.95 and item reliability of 0.99. This is in line with the opinion (Harachi, 2012) that the ideal level of reliability is greater than 0.90. The Cronbach Alpha value is obtained at 0.96. This indicates that the quality of the instrument has ideal criteria kriteria (Harachi, 2012; Perera et al., 2018).

Person **dan** ***Item Seperation Index*** the person separation index is 4.24 and the item separation index is 10.45 which provides information that the level of public speaking is in the range of student responses. This is in line with the opinion (Mez et al., 2012; Perera et al., 2018) The greater the person separation index and the item separation index, it means that the possibility of students responding to the item is likely to be appropriate and how wide the spread of items from easy to difficult items.

Precision of measurement the estimated value of the item is generated in the column "S.E. Model." measure the standard error for each item estimate. A good standard error in an instrument should be less than 0,5 (< 0,5) (Perera et al., 2018). The standard error value from data analysis is 0.05 logit, This means that the precision of measurement can be indicated on a reliable fit item.

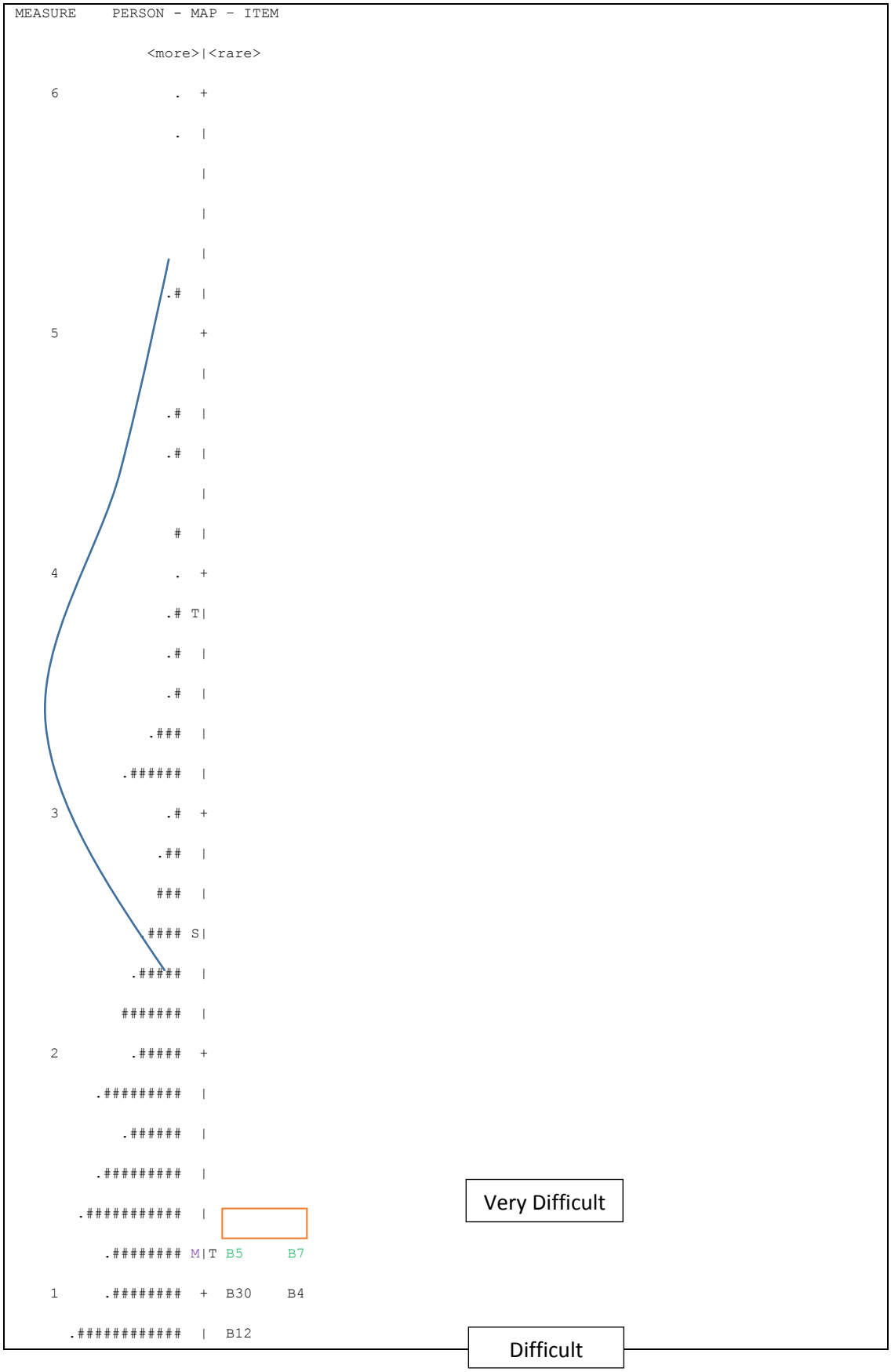
Item Calibration inform about the mean score on the item of 0.00 logit and the standard deviation of 0.55 logit. Item calibration shows the presence or absence of balance across the scale regarding difficult items (above log 0) and easy items (below log 0) (Mez et al., 2012). Item range information can be seen in Table 8 below:

Table 8: Range of item scales with four response categories

Category	***Measure***	**Range *Item***
Mean	0,00 logit	+1,05 logit to -0,96 logit
P.SD	0,55 logit	
S.SD	0,56 logit	
Total 2SD = 0,55 logit + 0,56 logit = 1,11 logit		

In Table 8 the mean score on the items is 0.00 logit and Person Standard Deviation (P.SD) is 0.55 logit. Statistics Standard Deviation (S.SD) is 0.56 logit, the item scale ranges from +1.05 logit to -0.96 logit obtained from the interpretation of Figure 4.7 item-person variable map. This means the calibration item is in the range of two SD, while the sum of the two SD = 0.55 logit + 0.56 logit = 1.11 logit. The 1.15 logit range is not in the +1.05 logit item range. It can be concluded that the instrument with a scale of four response categories indicated the presence of a misfit item.

Item-Person Variable Map shows an overview of student responses and items through a variable map plot on the same logit interval scale (W J Boone & Staver, 2014; William J Boone & Noltemeyer, 2017). Checking on item difficulty, so that effective instruments can be produced to use (William J Boone & Noltemeyer, 2017; Yacob et al., 2014). A clearer understanding can be seen in Figure 5 below:


MEASURE PERSON - MAP - ITEM
<more>|<rare>
6
5
4
3
2
1
M|T B5 B7
B30 B4
B12
Very Difficult
Difficult

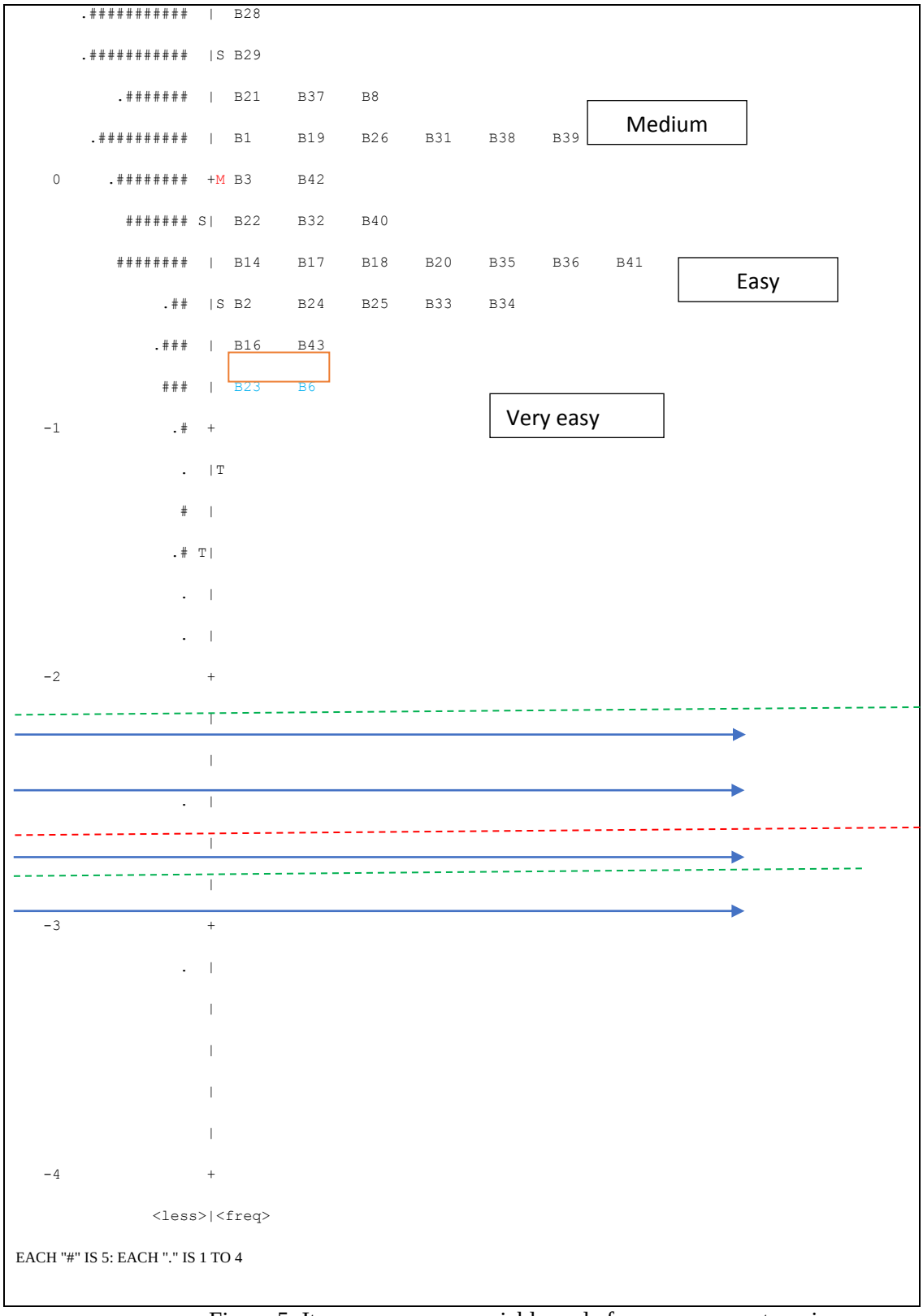


Figure 5: Item-person map variable scale four response categories

Based on Figure 5 above, where the location of "M" as the mean item lies at zero logit (W J Boone & Staver, 2014; Perera et al., 2018), in Figure 5 presents a test of 37 items, the mapping results

from the variable map show that the analysis of the statement items in the item measure contains information on the mean value and Standard Deviation (SD), when the two are combined then the level of statement items can be grouped. It is known that the acquisition of mean values of 0.0 logit and SD 0.45 logit on items using a scale of four response categories. Mean value 0.0 logit + SD 0.45 logit = +0.45 logit (medium item group limit); Item difficulty level +0.45 logit to +0.90 logit (difficult item group limit); item difficulty level > +0.90 logit (item group limit is very difficult); then if the mean is 0.0 logit - SD 0.45 logit = -0.45 logit (medium item group limit); item difficulty level from -0.45 logit to -0.90 logit (easy item group limit); item difficulty level < -0.90 logit (item group limit is very easy).

Item B5 and item B7 are very difficult items for student responses in determining the choice of response categories. Item B5 is on the "speech introduction" dimension with the statement "I was enthusiastic about the task in advance". Item B7 is in the "speech introduction" dimension with the statement "I am not satisfied with my performance". The position of item B23 and item B6 is an item that is very easy to respond to students in determining answer choices. Item B23 is on the "supporting material" dimension with the statement "During the presentation my heart beat faster than usual", while item B6 is in the "speech introduction" dimension with the statement "I still feel anxious due to my presentation".

Conclusion

The characteristics of the modified public speaking instrument are generally related to the ability of students to interact in the learning environment, psychologically motivated. The instrument is designed using a scale of four response categories, have a different number of items as the final result of a meaningful analysis. The use of different scales with different number of items on standardized instruments is grouped based on the validity of the Andrich Threshold and statistical information analysis results using the Rasch modeling approach.

Student responses using a scale of four response categories on the Indonesian version of the public speaking instrument, shows the most dominant dimension of "speech introduction". This means that students' responses to the condition of public speaking abilities are based on the perspective of individual and group differences that are consistent in reflecting higher agreement.

Therefore, for the psychological scale compilers are expected to compile statement items that are easy to understand and build good interactions with respondents, so that the respondent feels not intervened. The diversity of data also increases and an overview of public speaking skills in student responses is more effective. Detailed information and proof of estimation, of course, there is still an opportunity to carry out the analysis process with other statistical approaches, so that furthermore can provide complete and more in depth information.

References

Andrich, D. (2011). Rating scales and Rasch measurement. *Educational and Psychological Measurement. 11*(5), 571–585.

Boone, W J, & Staver, J. R. (2014). *Book reviews 545*. *7*, 545–548. https://doi.org/10.1177/0265532214567642

Boone, William J, & Noltemeyer, A. (2017). Rasch analysis : A primer for school psychology researchers and practitioners. *Cogent Education*, *25*, 1–13. https://doi.org/10.1080/2331186X.2017.1416898

Cheng, Y. Y., Chen, L. M., Liu, K. S., & Chen, Y. L. (2011). Development and psychometric evaluation of the school bullying scales: A Rasch measurement approach. *Educational and Psychological Measurement*, *71*(1), 200–216. https://doi.org/10.1177/0013164410387387

Cirik, I., Çolak, E., & Kaya, D. (2015). Constructivist learning: The teachers and students perspectives. *International Journal on New Trends in Education and Their Implications*, *6*(2), 30–45.

Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement*, *10*(1), 3–31. https://doi.org/10.1177/001316445001000101

De Oliveira, D. C. G., Zuardi, A. W., Graeff, F. G., Queiroz, R. H. C., & Crippa, J. A. S. (2012). Anxiolytic-like effect of oxytocin in the simulated public speaking test. *Journal of Psychopharmacology*, *26*(4), 497–505. https://doi.org/10.1177/0269881111400642

Fenty, N. S., & Anderson, E. M. K. (2014). Examining Educators' Knowledge, Beliefs, and Practices About Using Technology With Young Children. *Journal of Early Childhood Teacher Education*, *35*(2), 114–134. https://doi.org/10.1080/10901027.2014.905808

Fordham1981_rating scale8.pdf. (n.d.).

Glassman, L. H., Forman, E. M., Herbert, J. D., Bradley, L. E., Foster, E. E., Izzetoglu, M., & Ruocco, A. C. (2016). The Effects of a Brief Acceptance-Based Behavioral Treatment Versus Traditional Cognitive-Behavioral Treatment for Public Speaking Anxiety: An Exploratory Trial Examining Differential Effects on Performance and Neurophysiology. *Behavior Modification*, *40*(5), 748–776. https://doi.org/10.1177/0145445516629939

Goetz, T., Zirogibl, A., Pekrun, R., & Hall, N. (2003). *Emotions, Learning and Achievement from an Educational, Psychological Perspective Thomas Goetz, Anne Zirogibl, Reinhard Pekrun & Nathan Hall Outline*. *1998*.

Green, P. E. (2010). Information and to Categories. *Response*, *34*(3), 33–39.

Gudmundsson, E. (2009). Guidelines for translating and adapting psychological instruments. *Nordic Psychology*, *61*(2), 29–45. https://doi.org/10.1027/1901-2276.61.2.29

Habiby, W. N. (2012). Efektivitas Pelatihan Motivasi. *Efektivitas Pelatihan Motivasi Untuk Menurunkan Tingkat Stres Siswa Kelas VI*, *1*(2). https://journal.uny.ac.id/index.php/jpe/article/view/2637.

Harachi, T. W. (2012). *NIH Public Access*. *7*(1), 16–38.

Hindo, C. S., & González-Prendes, A. A. (2011). One-session exposure treatment for social anxiety with specific fear of public speaking. *Research on Social Work Practice*, *21*(5), 528–538. https://doi.org/10.1177/1049731510393984

Hofmann, S. G., & DiBartolo, P. M. (2000). An instrument to assess self-statements during public speaking: Scale development and preliminary psychometric properties. *Behavior Therapy*, *31*(3), 499–515. https://doi.org/10.1016/S0005-7894(00)80027-1

Ismail, S. N., Muhammad, S., Kanesan, A. G., & Ali, R. M. (2019). The influence of teachers' perception and readiness towards the implementation of Critical Thinking Skills (CTS) practice in mathematics. *International Journal of Instruction*, *12*(2), 337–352. https://doi.org/10.29333/iji.2019.12222a

Jones, C. R., Fazio, R. H., & Vasey, M. W. (2012). Attentional Control Buffers the Effect of Public-Speaking Anxiety on Performance. *Social Psychological and Personality Science*, *3*(5), 556–561. https://doi.org/10.1177/1948550611430166

Kupana, N. (2015). Social Emotional Learning and Music Education. *SED Journal of Art Education*, *3*(1). https://doi.org/10.7816/sed-03-01-05

Lapoint, V., & Butty, J. M. (2009). Encyclopedia of Cross-Cultural School Psychology. In *Encyclopedia of Cross-Cultural School Psychology* (Issue c). https://doi.org/10.1007/978-0-387-71799-9

Lundgren-Nilsson, Å., Dencker, A., Jakobsson, S., Taft, C., & Tennant, A. (2014). Construct validity of the Swedish version of the revised piper fatigue scale in an oncology sample - A rasch analysis. *Value in Health*, *17*(4), 360–363. https://doi.org/10.1016/j.jval.2014.02.010

Mez, L. E. G. Ó., Arias, B., Verdugo, M. Á. N., & Navas, P. (2012). *Application of the Rasch rating scale model to the assessment of quality of life of persons with intellectual disability*. *37*(June), 141–150. https://doi.org/10.3109/13668250.2012.682647

Moors, G. (2008). *Exploring the effect of a middle response category*. 779–794. https://doi.org/10.1007/s11135-006-9067-x

Nicolini, K. M., & Cole, A. W. (2019). Measuring peer feedback in face-to-face and online public-speaking workshops. *Communication Teacher*, *33*(1), 80–93. https://doi.org/10.1080/17404622.2017.1400678

Noor, W., Wan, F., Lee, N. W., & Abdullah, A. N. (2021). *Public Speaking Anxiety or English Language*

Anxiety ? 3(2), 58–65.

Ozgul, F., Kangalgil, M., Diker, G., & Yamen, E. (2018). Evaluation of the constructivist learning environments of physical education teacher candidates *. *European Journal of Educational Research*, *7*(3), 653–658. https://doi.org/10.12973/eu-jer.7.3.653

Perera, C. J., Sumintono, B., & Jiang, N. (2018). The Psychometric Validation Of The Principal Practices Questionnaire Based On Item Response Theory. *International Online Journal of Educational Leadership*, *2*(1), 21–38. https://doi.org/10.22452/iojel.vol2no1.3

Pfister, T., & Robinson, P. (2010). Speech emotion classification and public speaking skill assessment. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6219 LNCS*, 151–162. https://doi.org/10.1007/978-3-642-14715-9_15

Poeschl, S. (2017). Virtual reality training for public speaking-A QUEST-VR framework validation. *Frontiers in ICT*, *4*(JUN), 1–13. https://doi.org/10.3389/fict.2017.00013

Pörhölä, M. (1997). Trait anxiety, experience, and the public speaking state responses of Finnish university students. *International Journal of Phytoremediation*, *14*(3), 367–384. https://doi.org/10.1080/08824099709388680

Schonert-Reichl, K. A., Lawlor, M. S., Oberle, E., & Thomson, K. (2009). Identifying indicators and tools for measuring social and emotional healthy living : Children ages 5-12 years. In *Pan-Canadian Joint Consortioum for School Health*. https://doi.org/10.1007/s10489-018-1145-y

Schreiber, L. M., Paul, G. D., & Shibley, L. R. (2012). The Development and Test of the Public Speaking Competence Rubric. *Communication Education*, *61*(3), 205–233. https://doi.org/10.1080/03634523.2012.670709

Sinnema, C., Meyer, F., & Aitken, G. (2016). *Capturing the Complex , Situated , and Active Nature of Teaching Through Inquiry-Oriented Standards for Teaching*. https://doi.org/10.1177/0022487116668017

Skriner, L. C., Wolk, C. B., Stewart, R. E., Adams, D. R., Rubin, R. M., Evans, A. C., & Beidas, R. S. (2017). *Therapist and Organizational Factors Associated with Participation in Evidence-Based Practice Initiatives in a Large Urban Publicly-Funded Mental Health System*. https://doi.org/10.1007/s11414-017-9552-0

Slater, M., Pertaub, D. P., Barker, C., & Clark, D. M. (2006). An experimental study on fear of public speaking using a virtual environment. *Cyberpsychology and Behavior*, *9*(5), 627–633. https://doi.org/10.1089/cpb.2006.9.627

Thomson, S., & Rucker, M. L. (2002). The development of a specialized public speaking competency scale: Test of reliability. *Communication Research Reports*, *19*(1), 18–28. https://doi.org/10.1080/08824090209384828

Wadkins, T. (2021). Vocal Fillers, Contagion Effects, and, um, Overlooked Pedagogical Opportunities in the, uh, Public Speaking Classroom. *Basic Communication Course Annual*, *33*(1), 14.

Yacob, A., Ali, N. H., Yusoff, M. H., Saman, M. Y. M., & Hamzah, W. M. A. F. W. (2014). *Personalized Learning : An Analysis Using Item Response Theory*. *8*(4), 1107–1113.

Yan, Z., & Mok, M. M. C. (2012). Validating the coping scale for Chinese athletes using multidimensional Rasch analysis. *Psychology of Sport and Exercise*, *13*(3), 271–279. https://doi.org/10.1016/j.psychsport.2011.11.013

Yu, Y. H. C. S. W. (2015). *Examining unidimensionality and improving reliability for the eight subscales of the SF-36 in opioid-dependent patients using Rasch analysis*. 279–285. https://doi.org/10.1007/s11136-014-0771-z