

Psychometric evidence of statistical self-efficacy instrument based on postgraduate



Siti Mistima Maat*, Mohd Effendi Ewan Mohd Matore, Haryanti Mohd Affandi, Suziyani Mohamed

Faculty of Education, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

ARTICLE INFO

Article history:

Received 4 June 2018

Received in revised form

11 September 2018

Accepted 17 September 2018

Keywords:

Self-efficacy

Statistics

Rasch model

ABSTRACT

Self-efficacy determines students' capabilities in designing their performance and personal well – being. How their feelings and motivation shape the success whenever tasks are given to them. This study aimed at assessing the psychometric properties of statistical self-efficacy among 241 postgraduate students that been selected using purposive sampling. Survey research design applied and respondents were given a set of the questionnaire of Statistical Self-Efficacy (SSE) instrument which originally has 22 items. The data was analyzed using Rasch Model which produced a good item and person reliability. Likewise, the item fit, unidimensionality of the instrument were tested to identify the psychometric properties of SSE were fulfilled. The data from Rasch analysis had shown that SSE instrument had achieved the central assumptions such as item fit and unidimensionality. The analysis also covered item polarity, Wright Map, reliability and separation index. Although, past researchers have studied self-efficacy, yet some limitations including the psychometric properties have been overcome in this research. Thus, the Rasch Model analysis provides empirical evidence for SSE for future studies, particularly in the psychometrical aspect.

© 2018 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Self-efficacy is vital in psychology in defining beliefs of one's behavior, which lead to motivation, action and social environment. How students think and motivate themselves are related to their self-efficacy (Goulão, 2014). The source of efficacy is mainly from the individual experience, vicarious experience, social encouragement and the physical and emotional conditions (Bandura and Locke, 2003). Likewise, self-efficacy is highly required in making any outcome a success through behavior and approaches that are shaped by their belief of knowledge in subject matter. Besides, self-efficacy boost students' motivation in learning statistics through their awareness in controlling the motivation. An access to the development of knowledge and skills, self-efficacy has an impact to students' participation in class which leads to their academic achievement (Richardson et al., 2012; Ferla et al., 2009).

Therefore, the statistical self-efficacy concerning this context of the study, influence the students'

motivation and resilience in learning statistics. It may affect their cognitive or the affective domain of the learning process. Even in tertiary level, students' grade point average is correlated with their self-efficacy as well as learning strategies (Bartimote-Aufflick et al., 2015). Looking specifically at the self-efficacy in the academic issue; it is appropriate to measure the self-efficacy particularly in the subject matter, the students' outcome and consideration on demographic factors like gender should be explored (Nielsen et al., 2017). Differences in the analysis can be either from the items or the positive attribute of self-efficacy (Scherer and Siddiq, 2015).

Furthermore, to have a meaningful generalization the findings, statistical and psychometrical evidence for the responses are preferably in the study. Relevant information of the respondents' self-efficacy scales should be sufficient in providing the required proof of the study (Kreiner, 2013).

The justification for delivering the psychometric evidence is due to problems in replicating the scales based on past studies (Gaudiano and Herbert, 2003; Smith and Betz, 2000). In addition, self-efficacy has been measured widely yet in the more general context like General Self-Efficacy (GSE) (Schwarzer and Jerusalem, 1995) and Scale of Perceived Social Self-Efficacy (Smith and Betz, 2000). Hence, the importance of measuring self-efficacy is required through the use of Rasch Model analysis. Although

* Corresponding Author.

Email Address: sitimistima@ukm.edu.my (S. M. Maat)

<https://doi.org/10.21833/ijaas.2018.11.014>

2313-626X/© 2018 The Authors. Published by IASE.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

past research has shown efforts in producing the psychometric properties of the statistical self-efficacy, the differences in research setting the respondents particularly have gained insight into the said matter. Nevertheless, vigorous studies have been done yet there exist the missing part in validation aspect (Ward et al., 2002).

The Rasch Model is commonly used in providing extensive information on individuals and items of an assessment or instrument (Matore et al., 2018a; Maat et al., 2016). The robustness criteria of an instrument can be analyzed using the Rasch Model. Abd-el-fattah (2015) stated the limitations of measuring self-efficacy regarding too extensive or not related to a particular domain. Some issues of self-efficacy in academic have been debated among researchers (Nielsen et al., 2017). Therefore, this study was conducted to provide the psychometric evidence of self-efficacy instrument using Rasch Model. A quantitative approach methodology has been applied in gathering the data.

2. Methodology

A survey research design was implemented in order to get 255 postgraduate students as the respondent of the study. The characteristics of the respondents should be diversely across the targeted population (Ayob and Yassin, 2017). However, after the cleaning and screening data process, only 241 respondents were considered to participate in measuring the SSE. Using a purposive sampling technique, the respondents were postgraduate students, who enrolled in various master programs such as mathematics education, psychology, science education, measurement and evaluation, leadership, counseling and many more.

As part of the graduation requirement, they were required to enroll in one statistical course which was conducted in the second semester. The SSE consists of 14 items which were adapted from Schneider (2011) which has been tested twice and produced an acceptable rate of Cronbach alpha of 0.902 for the first administration and the reliability value increased by 0.935 during the second part of the study (Schneider, 2011). The psychometric properties of SSE involving the assumptions for Item Response Theory (IRT) such as the reliability (Apple, 2013), the item fit and unidimensionality (Effendi and Zamri, 2015) were also revealed. The justification for determining the item fit is to ensure items sustainability. Rasch Model is used as a probabilistic model which can predict the success of an event using maximum likelihood estimation using the transformation of ordinal data to logits data (Matore et al., 2018b).

3. Results and discussion

The results covers the analyses based on Rasch Model analysis, which include Item Fit, Unidimensionality and Wright Map. Each analysis would cover the cut off point for every assumption.

3.1. Item fit

Table 1 shows the item measure order, which displays the information on the logit measurement for the SSE items. The important points that can be considered in measuring the item fit include Mean Square (MNSQ) and Z-std. These criteria can be used to detect Item Outlier or Misfit.

Table 1: Item statistics: Measure order

Entry Number	Total Score	Count	Measure	Mode S.E.	Infit		Outfit		Pt-Measure		Exact Match		Item
					MNSQ	ZSTD	MNSQ	Zstd	Corr.	Exp.	OBS%	EXP%	
6	761	246	-91	-10	-98	-1	1.00	0	72	-72	64.3	59.1	D6
4	803	246	-66	-20	-81	-2.1	-82	-2.0	-78	-73	65.1	56.2	D4
5	805	246	-53	-10	-82	-2.1	-84	-1.9	-77	-72	64.3	56.9	D5
9	810	246	-52	-10	-83	-2.0	-83	-1.9	-77	-73	64.7	54.0	D9
7	817	246	-39	-10	-84	-1.8	-87	-1.5	-76	-72	66.4	55.5	D7
8	821	246	-36	-09	1.11	1.2	1.10	1.2	71	-73	52.7	52.1	D8
10	844	246	-27	-10	-90	-1.1	-90	-1.0	-75	-72	58.1	55.9	D10
11	839	246	-22	-10	-85	-1.7	-85	-1.7	-76	-72	64.7	56.5	D11
1	875	246	-1.3	-10	1.18	1.9	1.15	1.6	-67	-71	58.9	57.5	D1
2	890	246	-1.3	-10	-95	-.5	.97	-.3	-72	-71	61.8	56.0	D2
3	891	246	-.27	-10	-93	-7.7	-91	-1.0	-74	-72	63.9	55.8	D3
14	952	246	-.81	-10	1.05	-.5	1.05	-.6	.69	-.70	61.4	57.2	D14
12	993	246	-1.07	-10	1.47	4.3	1.43	4.0	.56	.69	43.2	59.4	D12
13	999	246	-1.44	-10	1.22	2.3	1.15	1.5	-.64	-.69	49.8	58.7	D13
Mean	864.3	246.0	-.00	-10	1.00	-.1	.99	-.2			60.0	56.5	
S.D.	70.7	.1	.67	.00	.18	1.9	.17	1.7			6.6	1.9	

In order to obtain the item validity for SSE, then Infit-Outfit Mean Square Analysis and Point Measure Correlation (PTMEA Corr) are required to be analysed. Fisher (2007) suggested that the Infit-Outfit Mean Square Analysis (MNSQ) which their values should be between 0.77 logits to 1.30 logits. Any items that beyond this range are suggested to be

modified or removed in order to ensure the psychometric properties are fulfilled.

Furthermore, the Zstd value that represents a normal unit is also used to test item fit with the model. However, the Zstd value can be neglected if the MNSQ value is acceptable. Moreover, the PTMEA Corr values indicate the direction of the item in measuring the construct (Hanafi et al., 2014)

As shown by Table 2, all Infit-Outfit values for SSE items are within the suggested range of 0.77 to 1.30 except item D12 (“Distinguish between a population parameter and a sample statistics”) of having 1.47 and its Zstd value is also more than 0. The Zstd value for item D12 is greater than 0 which reflects that the item is not able to predict.

3.2. Unidimensionality

The unidimensionality characteristic can be used to identify the measurement alignment of the

construct. Aziz et al. (2013) proposed that at least 40% of the raw variance explained by measures should be achieved.

Based on Table 3, it can be shown that 55.9% of the raw variance explained by measures has become the evidence that the instrument is measuring in the right direction. The justification of not reaching 60% is due to the item disturbance or noise of 8.9% which can be shown by the unexplained variance in 1st contrast.

Table 2: Standardized residual variance in (eigenvalue units)

	Empirical		Modeled	
Total raw variance in observations	31.7	100.0		100.0%
Raw variance explained by measures	17.7	55.9		55.5%
Raw variance explained by persons	10.8	33.9		33.7%
Raw Variance explained by items	7.0	21.9		21.8%
Raw unexplained variance (total)	14.0	44.1	100.0%	44.5%
Unexplained variance in 1st contrast	2.8	8.9%	20.1%	
Unexplained variance in 2nd contrast	2.0	6.2%	14.0%	
Unexplained variance in 3rd contrast	1.5	4.7%	10.7%	
Unexplained variance in 4th contrast	1.3	4.2%	9.6%	
Unexplained variance in 5th contrast	1.0	3.2%	7.3%	

Table 3: Summary of 241 respondents

	Total Score	Count	Measure	Model Error	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	49.0	14.0	.96	.42	1.00	-.4	.99	-.4
S.D.	9.2	0.0	1.55	.04	.81	2.0	.80	1.9
MAX.	66.0	14.0	4.48	.63	5.69	6.1	5.67	6.1

Real RMSE: .48; True SD: 1.47; Separation: 3.06; Person Reliability: .90; Model RMSE: .42; True SD: 1.47; Separation: 3.53; Person Reliability: .93; S.E. of Person Mean: .10

By using Rasch model, the logit values were obtained as the log-odds which are based on Natural logarithm unit. These logits are located along the variable line, which separates the distribution of item as well as the respondents as shown in Fig.1. The left side of the vertical ruler represents the distribution of the respondents. Each symbol of “#” represents 2 while the symbol “.” represent one respondent respectively. All 154 items of SSE are located at the right side of the ruler. The distribution of the respondents’ ability is not well mapped according to the items of SSE. The person mean value can be represented by the “M” on the left side of the ruler is higher than the item mean. Four respondents at the maximum location show a logit value of +7.45 and one respondent with a logit value of +1.45 was the minimum location. The location of these respondents on the ruler is related to their agreement on the respective items of D6 (“Identify the factors that influence power”) and D13 (“Identify when the mean, median and mode should be used as a measure of central tendency”). Item D6 is considered the most difficult item for the

respondents to agree yet item D13 is the easiest among all the 14 items.

3.3. Reliability and separation index

Table 4 shows the statistical summary of 241 respondents, which indicates excellent reliability (Fisher, 2007) value of 0.90. This suggests that the SSE instrument is consistent if given to another set of samples that have the same characteristics features (Arasinah et al., 2015). The person separation index shows 3.06, which exceed the 2 as the suggested value by Jones and Fox (1998) and Linacre (2002) which reflects the number of groups that were categorized according to their ability in responding towards their SSE. In addition, the person mean is +0.96 that indicates these respondents are aware with the importance of their self-efficacy in statistics.

Likewise, the item reliability of SSE has shown an outstanding value of 0.98 as displayed in Table 4. The item separation index of 6.34 exceeds the cut of point of 3 (Linacre, 2002) and indicating 6 strata of respondents’ ability in SSE.

Table 4: Summary of 14 SSE items

	Total Score	Count	Measure	Model Error	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
Mean	864.3	246.0	.00	.10	1.00	-.1	.99	-.2
S.D.	70.7	.1	.67	.00	.18	1.9	.17	1.7
Max.	999.0	246.0	.91	.10	1.47	4.3	1.43	4.0
Min.	761.0	246.0	-1.44	.09	.81	-2.1	.82	-2.0

Real RMSE: .10; True Sd: .66; Separation: 6.34; Person Reliability: .98; Model RMSE: .10; True Sd: .66; Separation: 6.59; Person Reliability: .98

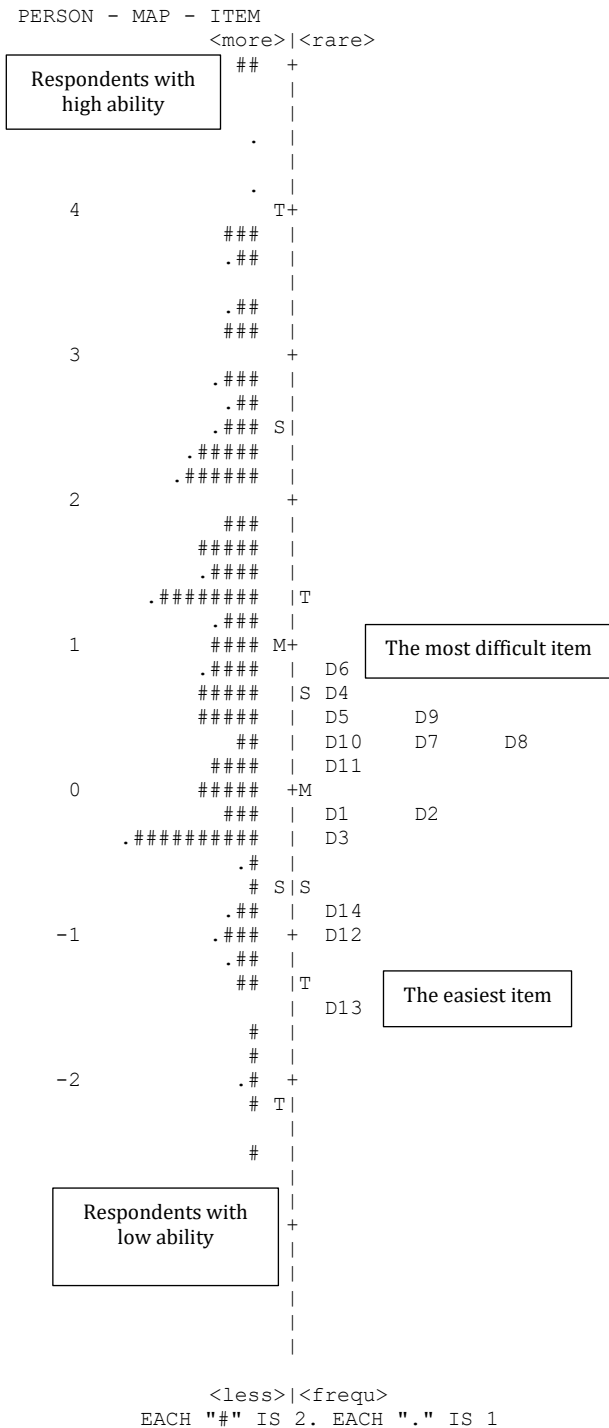


Fig. 1: Person -item map distribution

4. Conclusion

This study provides the psychometrical evidence of statistical self-efficacy instrument particularly in terms of postgraduate students. The psychometrical evidence of self-efficacy has become the endorsement in improve the SSE instrument. Every characteristic of Rasch model analysis indicates that the self-efficacy instrument can be used in the similar study of the same research context. Construct of self-efficacy can be clarified in terms of flaws and the responses. The hierarchy of the item difficulty and the ability of the respondents involved are measured on an equal continuum so that self-efficacy

is well assessed in terms of Item Response Theory (IRT). Some limitations do occur in this study that include uncertainty response from the respondents such as choosing a uniform Likert scale point for all items and the distribution of the questionnaire using online version has to be monitored. To overcome those challenges than, the involvement of the research during the distribution if the questionnaire is highly required. By monitoring them when choosing the item would able to avoid the uniform selection of Likert Scale point. While the restriction of using online version can be solved if the researcher fixes the time and venue during the data collection process.

Whatever the challenges are, a wide area of exploring self-efficacy is recommended for future studies. For instance, Differential Item Functioning is suggested in order to identify biases or unfairness concerning individuals. Another future research that can be done is by using Confirmatory Factor Analysis with regard to items and construct validations. Therefore, more extensive research on self-efficacy is highly recommended for the betterment in terms of the instrument as well as the relation with other important area.

Acknowledgement

The authors would like to acknowledge Universiti Kebangsaan Malaysia for supporting this research through GGPM2015-031.

References

Abd-El-Fattah SM (2015). Rasch rating scale analysis of the Arabic version of the physical activity self-efficacy scale for adolescents: A social cognitive perspective. *Psychology*, 6(16): 2161-2180.

Apple MT (2013). Using Rasch analysis to create and evaluate a measurement instrument for foreign language classroom speaking anxiety. *JALT journal*, 35(1): 5-28.

Arasinah K, Bakar AR, Ramlah H, Soaib A, and Zaliza H (2015). Using Rasch model and confirmatory factor analysis to assess instrument for clothing fashion design competency. *International Journal of Social Science and Humanity*, 5(5): 418-421.

Ayob A and Yassin RM (2017). A confirmatory factor analysis of the attitude towards mathematics scale using multiply imputed datasets. *International Journal of Advanced and Applied Sciences*, 4(3): 7-12.

Aziz AA, Masodi MS, and Zaharim A (2013). Asas model pengukuran Rasch: Pembentukan skala dan struktur pengukuran. Penerbit Universiti Kebangsaan Malaysia, Bangi, Malaysia.

Bandura A and Locke EA (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology*, 88(1): 87-99.

Bartimote-Aufflick K, Bridgeman A, Walker R, Sharma M, and Smith L (2015). The study, evaluation, and improvement of university student self-efficacy. *Studies in Higher Education*, 41(11): 1918-1942.

Effendi M and Zamri KA (2015). Psychometric assessment on Adversity Quotient instrument (IKBAR) among polytechnic students using Rasch model. In the International Conference on Education and Educational Technologies (EET), Institute for Natural Sciences and Engineering, Barcelona, Spain: 52-57.

- Ferla J, Valcke M, and Cai Y (2009). Academic self-efficacy and academic self-concept: Reconsidering structural relationships. *Learning and Individual Differences*, 19(4): 499-505.
- Fisher WP (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions*, 21(1): 1095.
- Fox CM and Jones JA (1998). Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology*, 45(1): 30-45.
- Gaudiano BA and Herbert JD (2003). Preliminary psychometric evaluation of a new self-efficacy scale and its relationship to treatment outcome in social anxiety disorder. *Cognitive Therapy and Research*, 27(5): 537-555.
- Goulão FM (2014). The relationship between self-efficacy and academic achievement in adults' learners. *Athens Journal of Education*, 1(3): 237-246.
- Hanafi NM, Rahman A, Mukhtar M, Ahmad J, and Warman S (2014). Validity and reliability of competency assessment implementation (cai) instrument using Rasch model. *International Journal of Social, Education, Economics and Management Engineering*, 8(1): 162-167.
- Kreiner S (2013). The Rasch model for dichotomous items. In: Christensen KB, Kreiner S, and Mesbah M (Eds.), *Rasch models in health*: 5-26. Wiley, New Jersey, USA.
- Linacre JM (2012). What do infit and outfit, mean-square and standardized mean. *Rasch Measurement Transactions*, 16(2): 878.
- Maat SM, Zakaria E, and Rosli R (2016). Validating mathematics teachers teaching practices questionnaire using Rasch measurement model. *Social Sciences (Pakistan)*, 11(18): 4419-4422.
- Matore MEEM, Khairani AZ, Maat SM, Nor Adila A and Effa Rina MM (2018a). The influence of intellectual quotient (iq), emotional quotient (eq) and spiritual quotient (sq) against adversity quotient (aq) on polytechnic students in Malaysia. *Journal of Engineering Science and Technology*, 13(Special Issue): 83-91.
- Matore MEEM, Maat SM, Affandi HM, Mohamad S, and Khairani AZ (2018b). Assessment of psychometric properties for Raven Advanced Progressive Matrices in measuring intellectual quotient (IQ) using Rasch model. *Asian Journal of Scientific Research*, 11(3): 393-400.
- Nielsen T, Dammeyer J, Vang ML, and Makransky G (2017). Gender fairness in self-efficacy? A Rasch-based validity study of the general academic self-efficacy scale (GASE). *Scandinavian Journal of Educational Research*, 62(5): 664-681.
- Richardson M, Abraham C, and Bond R (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2): 353-387.
- Scherer R and Siddiq F (2015). Revisiting teachers' computer self-efficacy: A differentiated view on gender differences. *Computers in Human Behavior*, 53: 48-57.
- Schneider WR (2011). The relationship between statistics self-efficacy, statistics anxiety, and performance in an introductory graduate statistics course. PhD Dissertation, University of South Florida, Tampa, Florida, USA.
- Schwarzer R and Jerusalem M (1995). Self-efficacy measurement: Generalized Self-Efficacy Scale (GSES). In: Weinman SWJ and Johnston M (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs*: 35-37. NFER-NELSON, Windsor, UK.
- Smith HM and Betz NE (2000). Development and validation of a scale of perceived social self-efficacy. *Journal of Career Assessment*, 8(3): 283-301.
- Ward M, Gruppen L, and Regehr G (2002). Measuring self-assessment: Current state of the art. *Advances in Health Sciences Education*, 7(1), 63-80.