Original Research **Open Access**

Comparison of Rasch Model Logits and Likert Mean Score in Testing the Normality Assumption

Mohd Effendi @ Ewan Mohd Matore[*]
Department of Education Leadership and Policy, Faculty of Education, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

Ahmad Zamri Khairani
School of Educational Studies, Universiti Sains Malaysia, 11800 Penang, Malaysia

Abstract
The lack of empirical evidence about Rasch model logits as a potential score in normality testing boost the motivations on how this type of data can replace the Likert mean score. The research also would like to see how these logits will be variant by fixing the diversity of zones selected for Adversity Quotient (AQ) measurement. The AQ was tested using IKBAR (*Instrumen Kecerdasan Menghadapi Cabaran*) with CORE model (Control, Ownership, Reach and Endurance) on 1845 polytechnic students from five different zones in Malaysia using a proportionate stratified multistage cluster sampling. The findings revealed that the skewness and kurtosis pattern for all zones showed that Rasch model logits are greater than the Likert mean score. The gaps pattern showed that Borneo dominated the largest gap for Control and AQ for both skewness and kurtosis. In terms of skewness, North region dominated the Endurance gap and the West dominated the Reach gaps. The East dominated the kurtosis gaps for Reach and Endurance. The Rasch model logits for all zones generally had a probability of failing to meet the normality acceptance range. A paired-samples t-test revealed that there was a significant difference for both scores. There was a significant average difference between both score. Rasch model logits were higher than Likert mean score. Based on the results, it is clear that the patterns of the Rasch model logits applicability exposed a great potential and alternative to be used by the researchers as normality data testing besides the Likert mean score. Further studies are proposed to be carried out to explore the potential of using logits for different settings and to ensure applicability to the other normality testing for educational research.

Keywords: Pattern; Normality; Skewness; Kurtosis; Likert mean score; Rasch model logits.

1. Introduction
Normality assumption should be taken into account when using parametric statistical tests (Ghasemi and Zahediasl, 2012). The requirement in most statistical techniques such as *t* tests, correlation, regression, and ANOVA needs the data to be normal and that is vital (Chua Yan Piaw, 2006; Drezner *et al.*, 2009; Hair *et al.*, 2013; Mohd Rafi Yaacob, 2013). Wrong interpretation of normality testing will affect the final results of the findings. If the normality assumption was interjected, the interpretation of data is likely to be invalid or unreliable (Nornadiah and Wah, 2011). (Lay and Khoo, 2008) emphasized this by stating the importance of examining the distribution of the data before any decision can be made in choosing the appropriate statistical test for data analysis or else the interpretations may not be reliable or valid (Park, 2008). One of the new instruments developed was *Instrumen Kecerdasan Menghadapi Cabaran* (IKBAR) using the Rasch model. It is used to measure Adversity Quotient (AQ) in Malaysia using samples of polytechnic students (Mohd and Ahmad, 2015), (Mohd and Ahmad, 2016). This instrument used logits in conducting the normality analysis. Based on the assumptions and measurement capabilities of Classical Test Theory (CTT) on rating scale, several limitations have been identified (Smith et al., 2002) Rating scale such as the Likert scale or ordinal data is often mistaken as an interval data and misused in parametric statistical procedures (Bode and Wright, 1999). In this study, IKBAR uses a four-point Likert scale (strongly agree, agree, disagree, and strongly disagree). Nowadays, most of the researchers prefer to use Likert mean score even though they realized that the data were ordinal. In Rasch, logits are known as interval data and they have good potential in fulfilling the requirement of normality analysis. Hence, this research would like to investigate the pattern of normality of Rasch logits and Likert mean score applied in different settings (polytechnic zones). This paper also investigates how far the applicability of Rasch logits in normality testing and also about the gaps of the Rasch model logits and Likert mean score in terms of the values produced. It has been conducted researches of Rasch based, but very lack information regarding the integration of normality and rasch (Mohd *et al.*, 2018b; Mohd *et al.*, 2018d; Mohd *et al.*, 2018a; Mohd *et al.*, 2018c; Muhamad *et al.*, 2018).

The other issue was the misunderstanding in normality testing. The distribution of larger samples tends to spread better and form a bell-shape curve when the characteristics of the population are plotted based on the frequencies. Confusion occurs when there is a sample from a normal population that does not show normal distribution; instead, only a few samples showed a seemingly normal distribution, especially with small sample size.

*Corresponding Author

The confusion might be due to graph techniques or eyeball test and formal normality tests that may show inconsistent results for the same data. Some tests can only be applied to certain situations or assumptions. Furthermore, differences in normality tests often produce different results, for example when some tests rejected and other tests fail to reject the null hypothesis of a hypothesis (Nornadiah and Wah, 2011). These different findings have been an issue and have been misleading researchers and normality test selection must be given priority (Nornadiah and Wah, 2011). To overcome this problem, skewness and kurtosis test methods are proposed where it is relatively accurate for both small and large (Kim, 2013). For this research, the skewness and kurtosis were applied for large samples.

2. Rasch Model and Logits

The Rasch Model (RM) helps explains the response on observable traits to estimate unobservable traits of a particular construct. RM involves modelling two important parameters in testing, namely, (1) item difficulty (observable trait), and (2) respondents' ability (unobservable trait). The Equation (1) for dichotomous and Equation (2) for polytomous relationship between the parameters [20]

$$P_{ni}(x = 1, D_i) = \frac{e}{1+e^{(B_n - D_i)]}} \qquad (1)$$

$$P_{ni1}(x = \frac{1}{B_n}, D_i, F_1) = \frac{e^{(B_n - [D_i + F_1])}}{1 + e^{(B_n - [D_i + F_1])}} \qquad (2)$$

Where the probability of getting a correct answer for Item i, P_i, person ability, B_n and item difficulty, D_i. This probability is equal to the constant e, or natural log function (2.7183) and then divided by one and the same value. Equation (2) shows the F_1 that was the difficulty of the first threshold and this difficulty calibration is estimated only once for this threshold across the entire set of items in the rating scale. For Logits, the construction of a logit scale is in the same way that we construct an Amp scale. We deduced a theory that produces equal interval, linear measures and derive a method for applying that theory. In the case of qualitative ordered observations (right/wrong, present/absent, none/some/all), the necessary and sufficient theory is the Rasch model, and the method of application is numerous administrations of similar agents (test items) to relevant objects (persons). The theory is mentioned in Equation (3).

$$[\![log(\frac{Probability\ of\ success}{Probability\ of\ failure}) = Ability - Difficulty \qquad (3)$$

This is a "linear" model because all elements can be represented as fixed positions along one straight line. In games of chance, the (Probability of Success) / (Probability of Failure) is called the "odds of success". "Log$_e$[(Probability of Success) / (Probability of Failure)]" is called log-odds. The units of measurement constructed by this theory are called "log-odds units" or "logits". (Wright, 1993) This logits help to turn the type of data from ordinal to the interval for this normality testing.

3. Methodology

This study used a quantitative approach and survey research design with a set of questionnaire from the *Instrumen Kecerdasan Menghadapi Cabaran* (IKBAR) by Mohd and Ahmad (2015) . The questionnaire contains 66 items of four-point Likert scale; representing four domains of Adversity Quotient (AQ) namely CORE or Control, Ownership, Reach and Endurance. The sample of this study consists of 1,892 respondents from 18,828 polytechnic students based on the clustered multistage stratified proportional sampling technique The respondents consists 994 males (53.9%) and 851 females (46.1%) which involving five polytechnics in Malaysia according to the five zones namely Western zones (456, 24.7%); Northern zones (393, 21.3%); Southern zones (375, 20.3%); Eastern zones (363, 19.7%) and Borneo zones (258, 14%). Data were analysed using SPSS for Windows Version 23. Then the data was collected and analysed based on polythomous Rasch model data by using computer applications of WINSTEPS 3.71.0.1. The return rate of the questionnaire was 97.52% (1,845 students) and considered as acceptable (Loewenthal, 2001);(Christensen *et al.*, 2011). The normality test of Skewness and Kurtosis was chosen because of the different layers of acceptance range of normality and the suitability for large sample sizes.

4. Results

Objective 1: To investigate the normality pattern of skewness and kurtosis by using CORE model based on different zones from Rasch model logits and Likert mean score perspectives.

Table 1 revealed that the skewness and kurtosis pattern for all zones showed that Rasch model logits are greater than Likert mean score. It was apparently shown in Figure 1 and 2 that it was occurred across the constructs and zones. The sequential mean range between logits for skewness are 1.179 (North), 1.159 (West), 1.144 (Borneo), 0.875 (East) and 0.864 (South). The sequential mean range between logits for kurtosis are 2.947 (Borneo), 2.174 (West), 2.154 (North), 1.585 (East) and 1.412 (South). Table 2 emphasized that the range for skewness was between 0.864 to 1.179 and for kurtosis was 1.412 to 2.947.

Table-1. Comparison of Likert and logits based on zone, CORE and testing (N= 1845)

Constructs	Testing	Borneo		East		South		North		West	
		Likert	Logits	Likert	Logits	Likert	Logits	Likert	Logits	Likert	Logits
Control	Skewness	-0.256	0.701	-0.354	0.266	-0.264	0.279	-0.169	0.592	-0.486	0.458
	Kurtosis	0.154	1.321	-0.485	0.043	-0.146	-0.057	-0.079	0.667	0.577	0.822
Ownership	Skewness	0.009	1.182	-0.139	0.752	-0.08	1.071	-0.245	0.983	0.136	1.179
	Kurtosis	0.09	3.244	-0.309	1.89	-0.062	3.463	0.509	2.659	0.103	2.404
Reach	Skewness	-0.024	0.89	-0.242	0.877	-0.214	0.659	-0.162	1.069	-0.100	1.186
	Kurtosis	-0.188	1.81	0.066	2.587	-0.015	1.646	0.323	2.58	0.467	2.933
Endurance	Skewness	-0.403	0.754	-0.114	0.949	-0.423	0.624	-0.196	1.006	-0.073	1.005
	Kurtosis	0.301	1.517	-0.377	1.604	0.426	1.193	0.328	1.806	-0.088	1.201
AQ	Skewness	-0.045	1.473	-0.263	0.419	-0.211	0.493	-0.252	1.219	-0.098	1.346
	Kurtosis	-0.074	7.126	-0.066	0.632	0.017	1.033	0.784	4.938	0.434	5.004

Figure-1. Normality pattern of skewness and kurtosis by using CORE model based on different zones from Likert mean score perspectives.
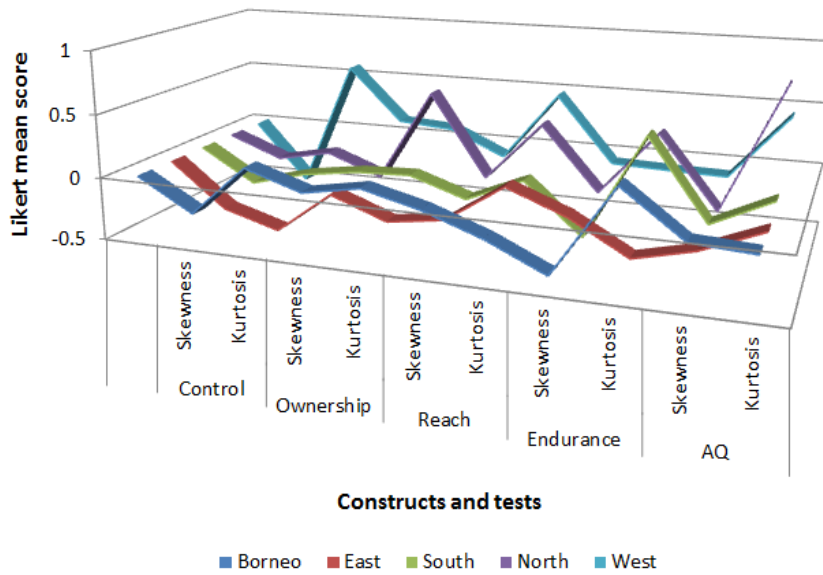

Figure-2. Normality pattern of skewness and kurtosis by using CORE model based on different zones from Rasch model logits perspectives.
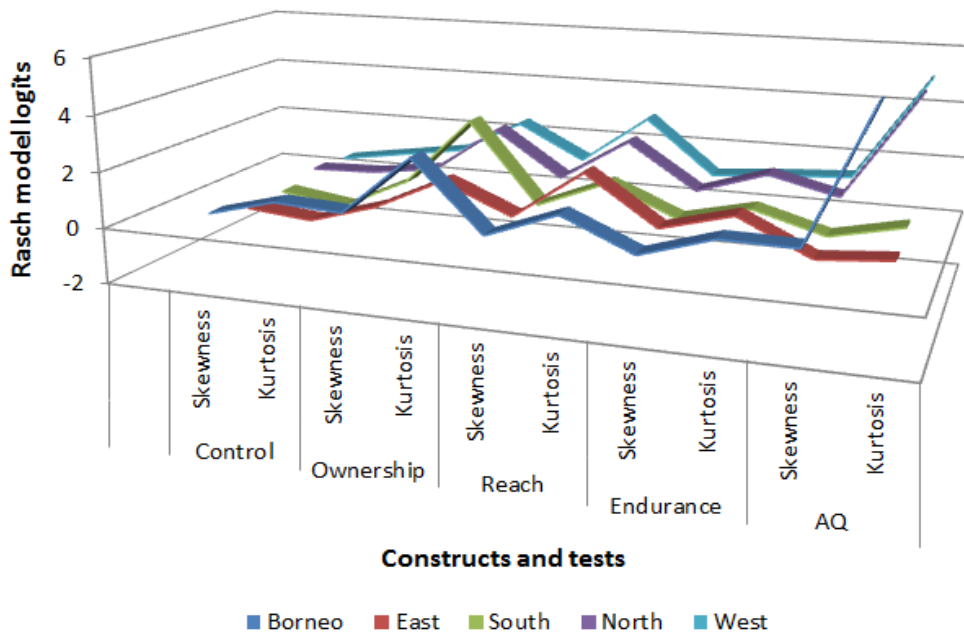
Table-2. Range of Logit-Likert based on zone, CORE and type of testing

	Borneo	East	South	North	West		Borneo	East	South	North	West
Mean Skewness Logit	-1.000	0.653	0.625	0.974	1.035	Mean - Kurtosis Logit	3.004	1.351	1.456	2.530	2.473
Mean Skewness Likert	-0.144	-0.222	-0.238	-0.205	-0.124	Mean - Kurtosis Likert	0.056	-0.234	0.044	0.373	0.298
Range	1.144	0.875	0.864	1.179	1.159	Range	2.947	1.585	1.412	2.157	2.174

Objective 2: To examine the gaps of skewness and kurtosis pattern between Rasch model logits and Likert mean score based on different zones.

This objective emphasized on examining the gaps between both scores (logits and Likert score) according to zones. Table 3 and Figure 3 showed that Borneo dominated the largest gaps for Control and AQ for both skewness and kurtosis. In terms of skewness, the North region dominated the Endurance gap and the West dominated the Reach gap. For kurtosis in Figure 4 showed that East dominated the gaps for Reach and Endurance.

Table-3. Gaps of Logit-Likert based on zone, CORE and type of testing

Testing	Construct	Borneo	East	South	North	West
Skewness	Control	0.957	0.620	0.543	0.761	0.944
	Ownership	1.173	0.891	1.151	1.228	1.043
	Reach	0.914	1.119	0.873	1.231	1.285
	Endurance	1.157	1.063	1.047	1.202	1.078
	AQ	1.518	0.682	0.704	1.472	1.443
Kurtosis	Control	1.1675	0.5276	0.0891	0.7453	0.2447
	Ownership	3.1536	2.1986	3.5242	2.1500	2.3007
	Reach	1.9982	2.5216	1.6611	2.2573	2.4659
	Endurance	1.2162	1.9814	0.7665	1.4777	1.2889
	AQ	7.1998	0.6970	1.0165	4.1537	4.5702

Figure-3. Gaps of skewness pattern between Rasch model logits and Likert mean score based on different zones.
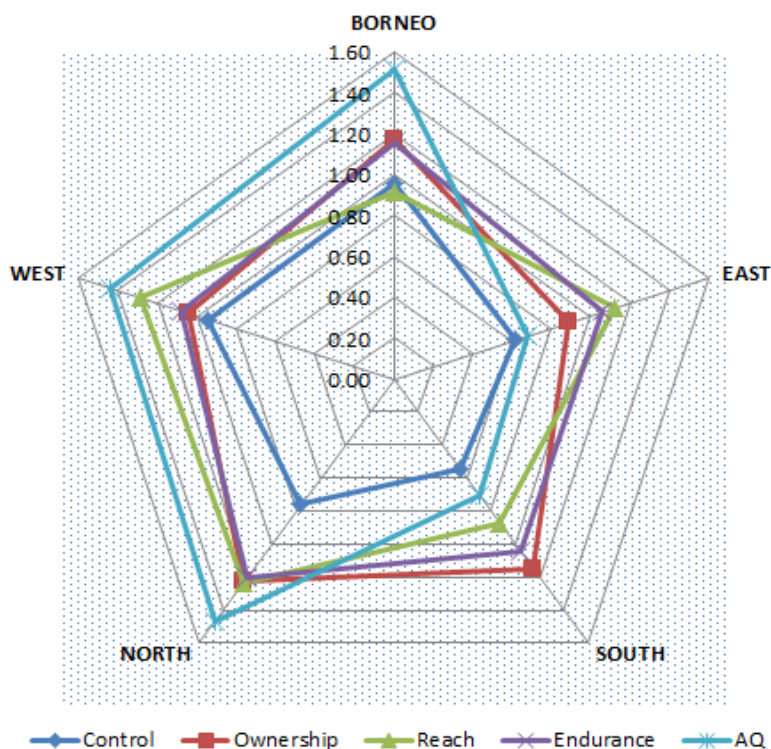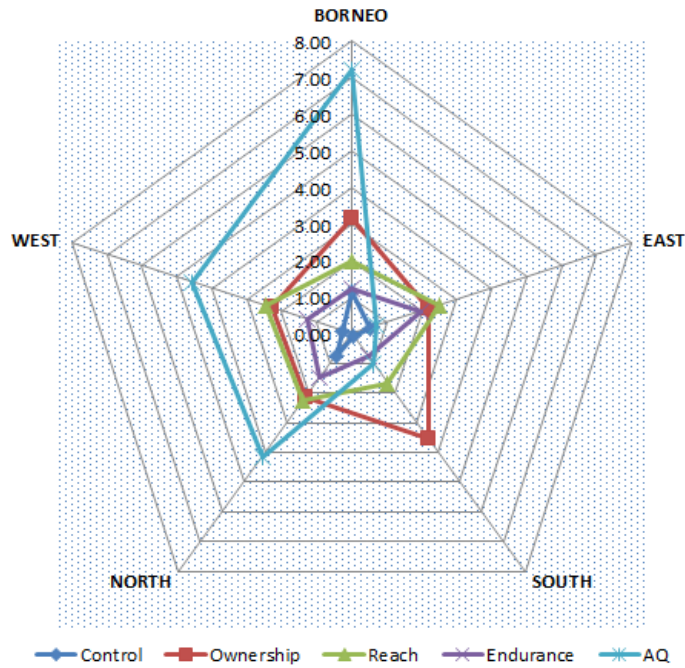
Figure-4. Gaps of kurtosis pattern between Rasch model logits and Likert mean score based on different zones.

Objective 3: To identify the capability of Rasch model logits and Likert mean score to meet the normality acceptance range by using CORE model based on different zones.

Figure 5 to 9 also revealed that the Rasch logits tend to deviate from the normality range compared to Likert mean score. If the data referred to ± 1.0 (Leech *et al.*, 2015), ± 2.0 (Garson, 2012);(Lomax and Hahs-Vaughn, 2012) or ± 3.0 (Peat and Barton, 2005), some of the constructs and testing also failed to meet the normality acceptance range. So, it was proven that Rasch logits for all zones mostly failed fulfil the requirement of skewness and kurtosis analysis.

Figure-5. Comparison of logits and mean based on skewness and kurtosis by using CORE model (Borneo zone).
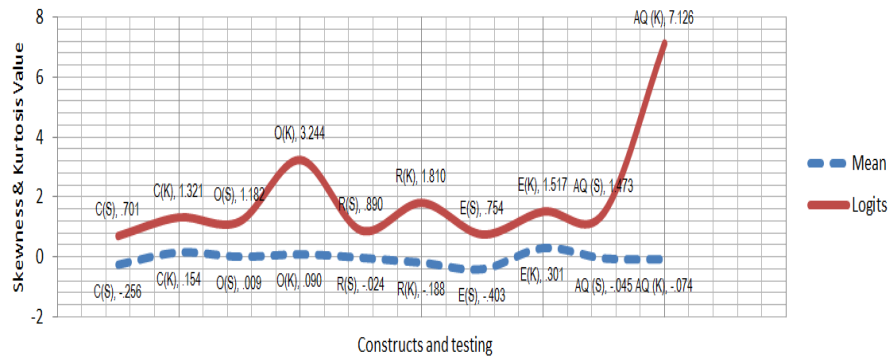

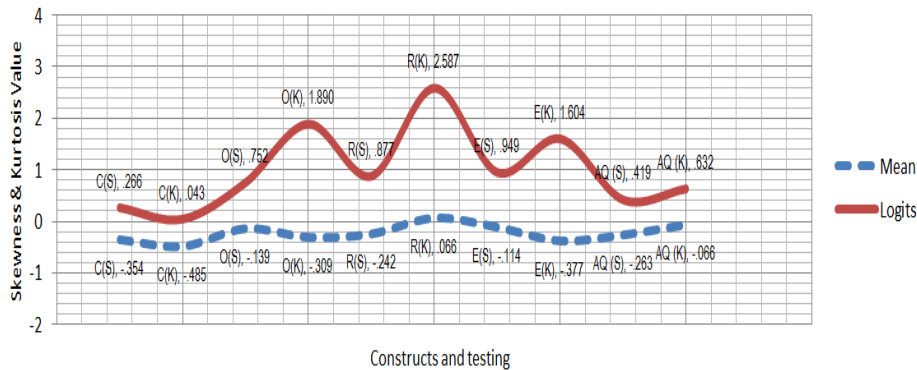**Figure-6.** Comparison of logits and mean based on skewness and kurtosis by using CORE model (East zone).

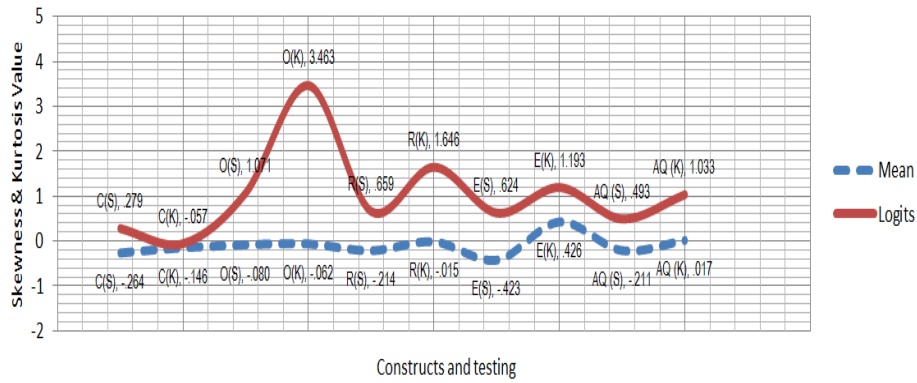**Figure 7:** Comparison of logits and mean based on skewness and kurtosis by using CORE model (South zone).

Figure -8. Comparison of logits and mean based on skewness and kurtosis by using CORE model (North zone).
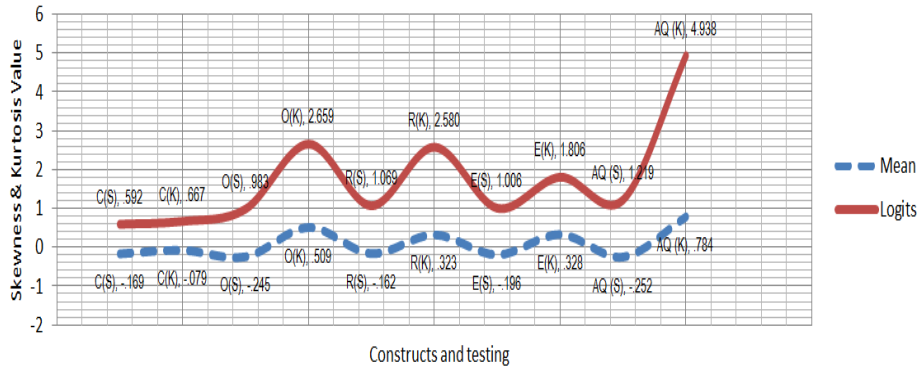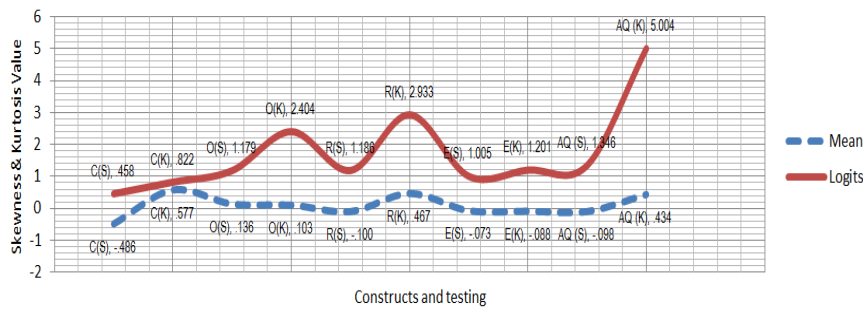

Figure-9. Comparison of logits and mean based on skewness and kurtosis by using CORE model (West zone).

Objective 4: To examine the significant differences between Rasch model logits and Likert mean score.

The reason of examining the significant differences between Rasch model logits and Likert mean score is to check whether both data were different for all zones and testing (skewness and kurtosis). Paired sample t test was chosen because both of the data were related each other in contributing to normality. The criteria for running the test had been fulfilled such as random sampling, normality and type of data. On average, Rasch model logits were 1.549 points higher than Likert mean score (95% CI [-1.898, -1.201]) as shown in Table 4. Table 5 had proven that Rasch model logits and Likert mean score were moderately and positively correlated ($r = 0.518, p < 0.001$). A paired-samples t-test was conducted to examine the significant differences between Rasch model logits and Likert mean score. Table 6 revealed that there was a significant difference in the scores for Likert mean score (M=-.0396, SD=.285) and Rasch model logits (M=1.510, SD=1.351) conditions; t (49) = -8.927, p = .000. There was a significant average difference between Rasch logits and Likert mean score scores for ($t_{49} = -8.927, p < 0.001$).

Table-4. Paired samples statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	mean score	-.03956	50	.285035	.040310
	logits	1.50996	50	1.350702	.191018

Table-5. Paired samples correlations

	N	Correlation	Sig.
Pair 1 mean score & logits	50	.518	.000

Table-6. Paired samples test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	S.D	S.E. Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	mean score - logits	-1.5495	1.2274	.1736	-1.8984	-1.2007	-8.927	49	.000

4. Conclusions

This study has proven that the pattern is quite fluctuated by both skewness and kurtosis. The potential of logits was tested because of their strength as interval data and how logits can replace Likert mean score for normality testing. This pattern proved that the gaps between both values of skewness and kurtosis and CORE produced the significant differences. Researchers can consider using logit for their normality tests, but the values deviate from normal compared to the Likert mean score. The research implication is directly to the researchers in terms of how difficult the logits in fulfilling the requirement of normality acceptance range even the data was in interval. Some future studies suggested to be carried out by seeing how this Rasch model logits will react to the different type of research setting and also the other type of normality testing.

References

Bode, R. K. and Wright, B. D. (1999). *Rasch measurement in higher education," in Higher Education: Handbook of Theory and Research, , J. C. Smart and W. G. Tierney.* Dordrecht: Springer Netherlands. 14: 287 – 316.

Christensen, L. B., Johnson, R. B. and Turner, L. A. (2011). *Research methods, design and analysis.* 11th ed edn: Pearson Education: Boston.

Chua Yan Piaw (2006). *Asas statistik penyelidikan (Buku 2).* McGraw Hill: Bentong.

Drezner, Z., Turel, O. and Zerom, D. (2009). *A modified Kolmogorov-Smirnov test for normality.* Fullerton, CA,. 14385.

Garson, G. D. (2012). *Testing statistical assumptions.* Statistical Associates Publishing: Asheboro, NC.

Ghasemi, A. and Zahediasl, S. (2012). Normality tests for statistical analysis, a guide for non-statisticians. *Int. J. Endocrinol. Metab,* 10(2): 486–9.

Hair, J. F., Celsi, M. W., Oritinau, D. J. and Bush, R. P. (2013). *Essentials of marketing research.* 3rd edn: McGraw Hill: New York.

Kim, H. Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor. Dent. Endod,* 38(1): 52–4.

Lay, Y. F. and Khoo, C. H. (2008). *Pengenalan kepada analisis statistik dalam penyelidikan sains sosial: Siri 2.* Venton Publishing (M) Sdn. Bhd: Batu Caves.

Leech, N. L., Barrett, K. C. and Morgan, G. A. (2015). *SPSS for intermediate statistics: use and interpretation,.* 5th ed edn: Lawrence Erlbaum Associates: Mahwah, New Jersey.

Loewenthal, K. M. (2001). *An introduction to psychological tests and scales.* 2nd edn: Taylor and Francis: Philadelphia.

Lomax, R. G. and Hahs-Vaughn, D. L. (2012). *An introduction to statistical concepts.* 3rd ed edn: Taylors and Francis: New York.

Mohd, E. E. M. M. and Ahmad, Z. K., 2015. "Psychometric assessment on Adversity Quotient instrument (IKBAR) among polytechnic students using Rasch model," in Proceedings of the International Conference on Education and Educational Technologies EET 2015." pp. 52–57.

Mohd, E. E. M. M. and Ahmad, Z. K. (2016). Correlation between adversity quotient (aq) with iq , eq and sq among polytechnic students using rasch model. *Indian J. Sci. Technol,* 9(47).

Mohd, E. E. M. M., Ahmad, Z. K. and Nordin, A. R. (2018a). Examining the Education Dimension of MPCL for Polytechnic Students Using Rasch Measurement Model. *Adv. Sci. Lett,* 24(5): 3427–31.

Mohd, E. E. M. M., Ahmad, Z. K., Nordin, A. R. and Siti, M. M. (2018b). Assessing the quality of items measuring Adversity Quotient among mechanical engineering students using Rasch model. *In Proceedings of Mechanical Engineering Research Day*: 189–90.

Mohd, E. E. M. M., Ahmad, Z. K., Mariani, M. N. and Mashitah, R. M. (2018c). Psychometric properties on isis in measuring spiritual quotient (sq) among polytechnic students using rasch measurement model. *Adv. Sci. Lett,* 24(7): 5310–15.

Mohd, E. E. M. M., Siti, M. M., Haryanti, M. A., Suziyani, M. and Ahmad, Z. K. (2018d). Assessment of psychometric properties for Raven Advanced Progressive Matrices in measuring intellectual quotient (IQ) using Rasch model. *Asian J. Sci. Res,* 11(3): 393–400.

Mohd Rafi Yaacob (2013). *SPSS 20 for business and social science students.* Eduserve Resources: Kota Bharu.

Muhamad, A. A., Sarimah, I., Rohana and Ros, E. A. Z. (2018). Content reliability measurement of holistic approach training module. *Int. J. Pure Appl. Math,* 118(24): 1–13.

Nornadiah, M. R. and Wah, Y. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *J. Stat. Model. Anal,* 2(1): 21–33.

Park, H. M. (2008). Univariate Analysis and Normality Test Using SAS, Stata, and SPSS. North Park Avenue Bloomington, Indiana.

Peat, J. and Barton, B. (2005). *Medical Statistic, A guide to data analysis and critical appraisal.* Blackwell publishing Ltd: USA.

Smith, E. V., Conrad, K. M., Chang, K. and Piazza, J. (2002). An introduction to Rasch measurement for scale development and person assessment. *J. Nurs. Meas,* 10(3): 189–206.

Wright, B. D. (1993). Logits'? *Rasch Meas. Trans,* 7(2): 288.