SERDEC Education Journal -SEJ
ISSUE. 2, 2021 Online ISSN: 2710-4796 https://sej.so/

Rasch Model Analysis on the Exam of Social Work Research Methods

Said Abubakar Sheikh Ahmed,PH.D in Education,Curriculum
E-mail:baashaa4@gmail.com/ said@mu.edu.so

DOI: 10.70595/sej88

Abstract

This study aimed at determining the quality of research methodology course test for social work department students, third year at Mogadishu University. Rasch model in WINSTEPS was used to examine 17 items for 42 test-takers based on Separation, Reliability, Dimensionality, Model fit, Item difficulty Point measure correlation, and Person items distribution map (right map). The majority of the items met the acceptable requirements, however some did not. As a result, the study's findings revealed that the Rasch measurement model in WINSTEPS is a useful tool for analysing tests in higher education.

Key Words: Rasch Model, Analysis , Test of Research Methods

Introduction

Rasch, G (1960) in the human sciences, the Rasch model and its extended models have been frequently used to test data. To calibrate item and person parameters, several computer programs have been developed (Wang & Chen, 2005).

Rasch modeling provides a variety of immediate practical benefits, including the quantification of question difficulty and system capability on a single scale with a standard metric, equating (calibration) of different question corpora, evaluation of the degree to which independent evaluations assess the same system abilities, and availability of rigorous statistical techniques such as data analysis, test reliability, and identification of unmodeled sources of variation in the data(Lange et al., 2004).

This study is the first work or one of the primary studies was done in Somalia according to the knowledge of the author. The main aim of this study is to analyse the test results of research methods for the students of the social work department, third year at Mogadishu University using Rasch model in WINSTEPS to determine the quality of the test.

Methodology

The study is a descriptive method designed to examine the students' results of a research methodology course to explore the students' performance . Test takers made up 42 students, 30 male, and 12 female. 17 items of the objective test were applied. WINSTEPS version 4.8.0.0, and SPSS version 22.0 were used to analyse the data based on separation, and reliability of both persons and items, dimensionality**,** item difficulty, and fit (infit and outfit), as well as person-item map analysis.

Results and Discussion

Reliability

To determine the validity and reliability of the data, WINSTEPS software version 4.8.0.0 was used. Separation and reliability levels for both persons and items were displayed in table (1), and table (2) which the consistency responses examined by the Rasch model interpretation on the person and item reliability were explained with Cronbach's alpha equivalent to Kuder-Richardson (KR-20).

Table 1 shows the reliability and separation of 42 measured persons. The person reliability indicates 0.68 which is fair according to Fisher (2007), less than the required and acceptable value by Sekaran (2003) cited by (Maat, 2015). However, in social science, by Ghazali (2008), the acceptable α value is 0.60 (Mohamad et al., 2015). Based on the separation index value (1.44) shown in the table is less than the minimum level (<2.0). This value indicates that 1.44 statistically distinct groups can be identified in the data or distinct strata (i.e., low and high ability) as a discrimination level.

Table 1. Person Reliability and Separation Results

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	9.5	17.0	.43	.61	.98	-.02	.97	.09
SEM	.5	.0	.18	.01	.05	.17	.07	.15
P.SD	3.2	.0	1.14	.10	.30	1.10	.47	.97
S.SD	3.2	.0	1.15	.10	.30	1.11	.48	.98
MAX.	16.0	17.0	3.30	1.05	1.79	2.86	2.28	2.45
MIN.	4.0	17.0	-1.52	.56	.57	-2.20	.39	-1.27

REAL RMSE .65 TRUE SD .94 SEPARATION 1.44 PERSON RELIABILITY .68
MODEL RMSE .62 TRUE SD .96 SEPARATION 1.54 PERSON RELIABILITY .70
S.E. OF PERSON MEAN = .18
MEDIAN = .37

The separation value of items shown in table (2) designates (2.78) is less than the required value (<3), however, this value shows that the items, somewhat can be divided into 3 levels of difficulty. Kubiszyn and Borich (2000) opined that the reliability of items is acceptable if the alpha is within 0.70 and 0.99 (Mohamad et al., 2015). The value of item reliability shown in table (2) reveals (0.89) which is very close to (0.9).Thus, it is an acceptable value.

Table 2. Item Reliability and Separation Results

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	23.5	42.0	.00	.39	1.01	-.11	.97	-.15
SEM	2.2	.0	.31	.02	.06	.30	.09	.27
P.SD	8.8	.0	1.24	.07	.23	1.20	.37	1.10
S.SD	9.1	.0	1.27	.07	.24	1.24	.38	1.13
MAX.	39.0	42.0	1.55	.62	1.59	2.58	2.12	2.91
MIN.	12.0	42.0	-2.59	.35	.68	-2.25	.49	-2.11

REAL RMSE .42 TRUE SD 1.16 SEPARATION 2.78 ITEM RELIABILITY .89
MODEL RMSE .40 TRUE SD 1.17 SEPARATION 2.94 ITEM RELIABILITY .90
S.E. OF ITEM MEAN = .31
MEDIAN = .39

Dimensionality

Table.3. Standardized Residual Variance (in Eigenvalue Units)

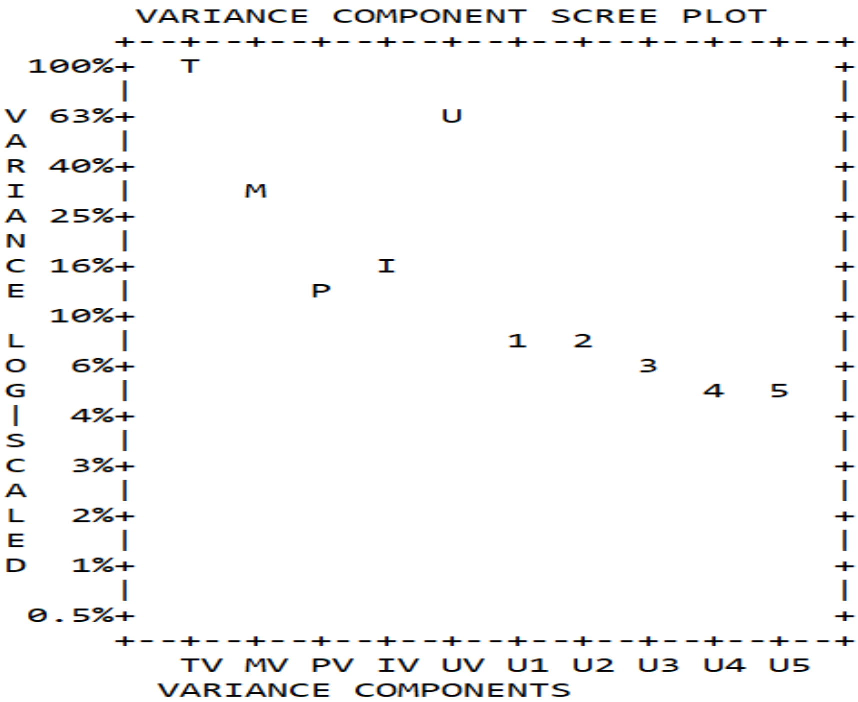
	Eigenvalue	Observed		Expected
Total raw variance in observations	25.6411	100.0%		100.0%
Raw variance explained by measures	8.6411	33.7%		33.4%
Raw variance explained by persons	3.6232	14.1%		14.0%
Raw Variance explained by items	5.0179	19.6%		19.4%
Raw unexplained variance (total)	17.0000	66.3%	100.0%	66.6%
Unexplned variance in 1st contrast	2.5181	9.8%	14.8%	
Unexplned variance in 2nd contrast	2.3949	9.3%	14.1%	
Unexplned variance in 3rd contrast	1.7959	7.0%	10.6%	
Unexplned variance in 4th contrast	1.5785	6.2%	9.3%	
Unexplned variance in 5th contrast	1.4810	5.8%	8.7%	

Table (3) and figure (1) illustrate the dimensionality of the instrument in terms of raw variance explained by measures, persons and items as well as unexplained of 1^{st}, 2nd, 3rd, 4th, and 5^{th} contrasts.

According to Lord and Novick (1968), dimensionality is defined as the total number of abilities required to satisfy the assumption of local independence. If only one ability is required for the responses to a set of items to meet the assumption of local independence, then that set is referred to as a unidimensionality (Council & Measurement, 1994). Dimensionality analysis attempts to ascertain whether test items target the same latent trait. If they do, the test is said to be unidimensional (Aryadoust, 2012).

A Principal Component Analysis (PCA) on the raw data and residuals, is used to check the data-model fit. (Alavi & Bordbar, 2016). Based on that, Table3 shows the dimensionality of items where 33.7% of the raw variance was explained by the measurement model which is approximately equal to the expected which is 33.4%. However, 33.7% of the variance for the Rasch measurement model is measured as weak according to Fisher (2007) as a rating the Instrument Quality Criteria (Mokshein et al., 2019). The minimum requirement value of variance is 40% (Isa et al., 2016). Nevertheless, Reckase (1979) opined unidimensional if the amount of variance explained by measures is > 20% (Alavi & Bordbar, 2016).According to Rasch model, the eigenvalue of unexplained variance of the 1st contrast should be (< 2) and it's observed percentage not larger than the variance explained by the item difficulties (Linacre, 2011). Therefore, the eigenvalue of unexplained variance of the 1st contrast is (2.51) ,and the observed present of unexplained variance of the 1st contrast 9.8% is not larger than the variance explained by the item difficulties 19.6%.

Figure 1.Variance Component Scree Plot and Logit Scaled



Infit and Outfit Analysis of the Items

Fit statistics were examined to determine differences between what is observed and what is expected through infit-outfit analysis to explore the items that do not fit the model's predictions. Infit-outfit of mean square (MNSQ) and ZSTD are used to analyse item fit indices. INFIT is a t standardized information-weighted mean square statistic, that is more sensitive to unexpected behaviour affecting responses to items near the person's measure level. OUTFIT is a t standardized outlier-sensitive mean square fit statistic, more sensitive to unexpected behaviour by persons on items far from the person's measure level. The standardization is approximate. Its success depends on the distribution of persons and items (Linacre, 2011).

Table 4 shows Statistics of Item Difficulty, Standard Error, Infi-Outfit, and Point Measure Correlation. All 17 items- except items - fall in the acceptable range 0.7-1.3 (Alavi & Bordbar, 2016). Item 1 is considered misfit due to (1.59) for infit MNSQ and (2.12) for the outfit MNSQ are under-fitting beyond the range of productive measurement which is between. Items Q10, Q12, Q13, and Q15 are considered to overfit less than(0.7). For the Point Measure Correlation for the items illustrated in table 4, the items (Q1, Q2, Q3, Q5, Q7, Q14, and Q15) are less than the required point (0.4), while the rest of items have met the acceptable criteria (<0.4). For the item difficulty, items (Q1, Q2, Q16, Q8, and q12)are the most difficult items, while,(Q3, Q5, Q13,andQ15) are the easiest.

Table 4. Statistics of Item Difficulty, Standard Error,Infi-Outfit ,and Point Measure Correlation

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
1	12	42	1.55	.39	1.59	2.58	2.12	2.91	-.04	.46	66.7	77.7	Q1
2	12	42	1.55	.39	1.31	1.48	1.23	.81	.24	.46	71.4	77.7	Q2
16	12	42	1.55	.39	1.08	.46	.98	.02	.41	.46	71.4	77.7	Q16
8	16	42	1.00	.36	.88	-.71	.83	-.71	.55	.46	83.3	72.4	Q8
12	16	42	1.00	.36	.68	-2.25	.58	-2.11	.72	.46	83.3	72.4	Q12
9	17	42	.87	.35	.91	-.59	.88	-.51	.53	.46	76.2	70.9	Q9
4	19	42	.63	.35	.90	-.73	.81	-.90	.55	.46	71.4	69.4	Q4
6	20	42	.51	.35	.97	-.16	.92	-.32	.48	.45	69.0	68.8	Q6
10	21	42	.39	.35	.75	-2.04	.67	-1.75	.65	.45	81.0	68.7	Q10
11	25	42	-.09	.35	.85	-1.18	.85	-.56	.54	.43	73.8	68.9	Q11
17	27	42	-.34	.36	.96	-.27	1.20	.75	.41	.41	78.6	71.1	Q17
14	29	42	-.60	.37	1.02	.19	1.00	.12	.37	.39	73.8	73.4	Q14
7	30	42	-.74	.37	1.12	.74	.95	-.02	.32	.38	66.7	74.6	Q7
13	33	42	-1.19	.40	.76	-1.19	.58	-.93	.54	.35	83.3	79.3	Q13
3	36	42	-1.75	.47	1.30	1.03	1.38	.75	.04	.29	85.7	85.7	Q3
5	36	42	-1.75	.47	1.21	.78	1.03	.25	.15	.29	85.7	85.7	Q5
15	39	42	-2.59	.62	.91	-.03	.49	-.34	.32	.22	92.9	92.8	Q15
MEAN	23.5	42.0	.00	.39	1.01	-.1	.97	-.1			77.3	75.7	
P.SD	8.8	.0	1.24	.07	.23	1.2	.37	1.1			7.4	6.7	

Persons Items Distribution Map (Wright Map)

Persons Items Distribution Map (Wright Map) allows both the person and the item to be mapped together, but on the same logit scale, offering a clearer picture of correlates to each individual item (Hashim et al., 2012). It shows all and more importantly the logical hierarchy of difficulty based on the conceptual theory put under test. This will be the evidence of the instrument construct validity acceptance (Ishar & Masodi, 2012).

Figure (2) shows Persons Items Distribution Map (Wright Map) where the majority test items are centered on the mean, which the majority of test takers positioned. In addition, most items were logical hierarchy of difficulty, indicating the sufficient items to estimate test takers' ability.

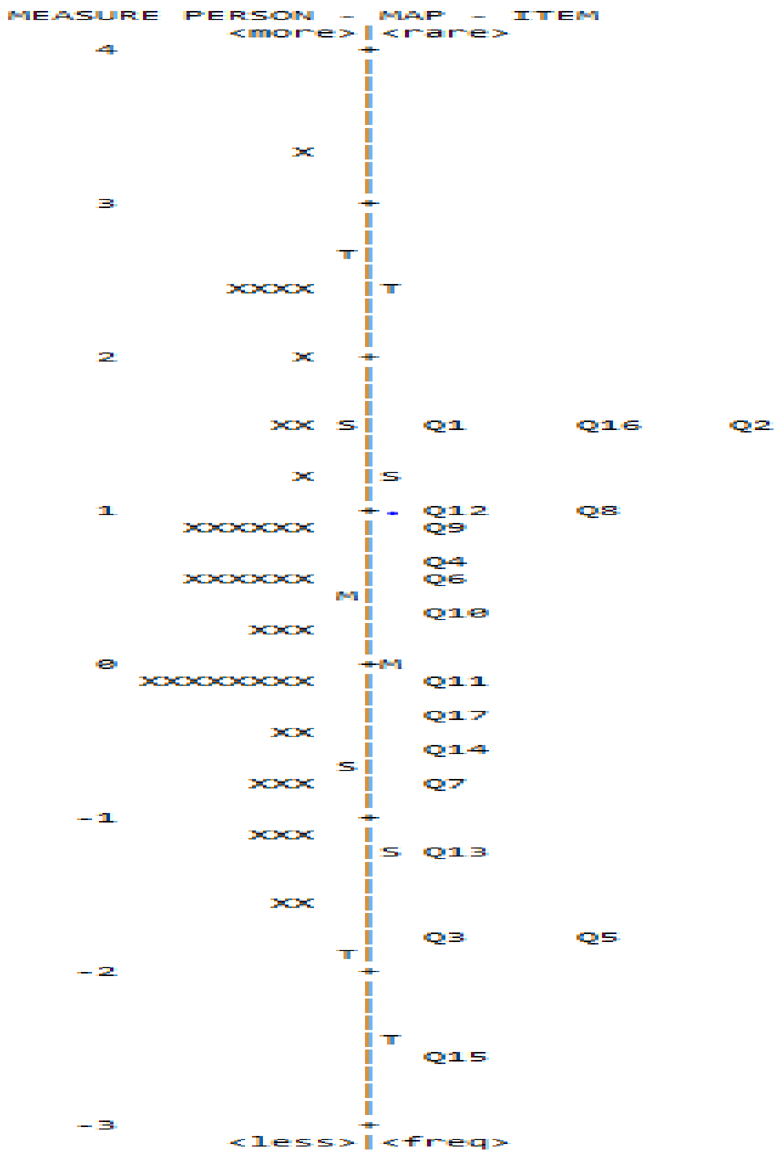


Figure 2.Persons Items Distribution Map (Wright Map)

Table 5. Results of *T*-Test Analysis on Each Item Separation/Discrimination

	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Q1	.670	40	.506	.095	.142
Q2	-2.108	40	.041	-.286	.136
Q3	.000	40	1.000	.000	.111
Q4	-3.017	40	.004	-.429	.142
Q5	-1.789	40	.081	-.190	.106
Q6	-2.609	40	.013	-.381	.146
Q7	-2.108	40	.041	-.286	.136
Q8	-1.947	40	.059	-.286	.147
Q9	-2.283	40	.028	-.333	.146
Q10	-4.985	40	.000	-.619	.124
Q11	-3.069	40	.004	-.429	.140
Q12	-5.970	40	.000	-.667	.112
Q13	-2.811	40	.008	-.333	.119
Q14	-1.685	40	.100	-.238	.141
Q15	-1.826	40	.075	-.143	.078
Q16	-2.941	40	.005	-.381	.130
Q17	-1.622	40	.113	-.238	.147

Based on the separation items previously discussed was scored up (2.78) is less than the required point (>3). To find out the separation level of each item is used t-test analysis in SPSS.25.Table 5 indicates that the *P*-values of items (Q1, Q3, Q5, Q8, Q14, Q15, and Q 17) are greater than (0.05), therefore, these items are not discriminant, while the rest of the 10 items are discriminating ability.

Conclusion

A Rasch analysis on the test of the research methodology course in Winsteps had been examined. Person reliability and separation of 42 were measured. The person reliability indicated 0.68 which is fair according to Fisher (2007), and acceptable in social sciences according to Ghazali (2008), whereas the separation outcome (1.44) did not meet the criteria of acceptance, less than the minimum level of 2.0. For the item reliability and separation analysis of 17 items were applied where the reliability was within the acceptance level (0.89), while the separation did not less than(<3), but closed to the acceptable point (2.78). Dimensionality, Infit and Outfit analysis of the items, Persons Items Distribution Map (Wright Map), and t-test analysis were examined. The results of most items showed the acceptable criteria, while others did not. Thus, the results of the study showed that the Rasch measurement model in Winsteps is an appropriate tool for analysing tests of courses in higher education.

References

Alavi, S. M., & Bordbar, S. (2016). Differential Item Functioning Analysis of High-Stakes Test in Terms of Gender: A Rasch Model Approach. *Malaysian Online Journal of Educational Sciences*, *5*(1), 10–24.

Aryadoust, V. (2012). Differential Item Functioning in While-Listening Performance Tests: The Case of the International English Language Testing System (IELTS) Listening Module. *International Journal of Listening*, *26*(1), 40–60. https://doi.org/10.1080/10904018.2012.639649

Council, N., & Measurement, E. (1994). *Assessing Dimensionality of a Set of Item Responses : Comparison of Different Approaches Author (s): Ratna Nandakumar Published by : National Council on Measurement in Education Stable URL : https://www.jstor.org/stable/1435095 REFERENCES Linked referen*. *31*(1), 17–35.

Hashim, N., Ishar, N. I. M., Rashid, W. E. W., & Masodi, M. S. (2012). Personality Traits, Work-Family Conflict and Job Satisfaction: Items Validity using Rasch Measurement Approach. *Procedia - Social and Behavioral Sciences*, *65*(ICIBSoS), 1013–1019. https://doi.org/10.1016/j.sbspro.2012.11.235

Isa, M., Bakar, M. A., Sipan, I., Hasim, M. S., Hashim, A. E., & Jalil, M. K. A. (2016). Measuring Instrument Constructs of Return Factors for Green Office Building Investments Variables Using Rasch Measurement Model. *MATEC Web of Conferences*, *66*. https://doi.org/10.1051/matecconf/20166600125

Ishar, N. I. M., & Masodi, M. S. (2012). Students' perception towards quality library service using Rasch measurement model. *ICIMTR 2012 - 2012 International Conference on Innovation, Management and Technology Research*, *May 2012*, 668–672. https://doi.org/10.1109/ICIMTR.2012.6236479

Lange, R., Greiff, W. R., Moran, J., Ferro, L., & acl. (2004). A probabilistic Rasch analysis of question answering evaluations. *Hlt-Naacl 2004: Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, Proceedings of the Main Conference*, 65–72.

Linacre, J. M. (2011). *Winsteps Help for Rasch Analysis*. http://homes.jcu.edu.au/~edtgb/%5Cnpapers3://publication/uuid/D56B724A-62FF-4D00-84E1-ECC888298B70

Maat, S. M. (2015). Psychometric Evaluation on Mathematics Beliefs Instrument Using Rasch Model. *Creative Education*, *06*(16), 1797–1801. https://doi.org/10.4236/ce.2015.616183

Mohamad, M. M., Sulaiman, N. L., Sern, L. C., & Salleh, K. M. (2015). Measuring the Validity and Reliability of Research Instruments. *Procedia - Social and Behavioral Sciences*, *204*(August), 164–171. https://doi.org/10.1016/j.sbspro.2015.08.129

Mokshein, S. E., Ishak, H., & Ahmad, H. (2019). The use of rasch measurement model in English testing. *Cakrawala Pendidikan*, *38*(1), 16–32. https://doi.org/10.21831/cp.v38i1.22750

Wang, W. C., & Chen, C. Te. (2005). Item parameter recovery, standard error estimates, and fit statistics of the WINSTEPS program for the family of Rasch models. *Educational and Psychological Measurement*, *65*(3), 376–404. https://doi.org/10.1177/0013164404268673