The Journal of Social Sciences Research
ISSN(e): 2411-9458, ISSN(p): 2413-6670
Vol. 5, Issue. 1, pp: 28-39, 2019
URL: https://arpgweb.com/journal/journal/7
DOI: https://doi.org/10.32861/jssr.51.28.39

Academic Research Publishing Group

Original Research Open Access

The Psychometric Properties of an Intrinsic Motivation Scale in Conducting Research: The Application of Rasch Measurement Model

Kamal Badrasawi*
Department of Curriculum and Instruction Kulliyyah of Education International Islamic University Malaysia

Faizah Idrus
Department of Language and Literacy Kulliyyah of Education International Islamic University Malaysia

Afareez Abd Razak
Department of Social Foundations and Leadership Kulliyyah of Education International Islamic University Malaysia

Nik A. Hisham
Department of Psychology and Counselling Kulliyyah of Education International Islamic University Malaysia

Abstract

Research culture a system that posits importance to conducting and communicating scholarly research, is highly expected from academic staff in the higher learning institutions. Skillful academics with high levels of (intrinsic) motivation in conducting research are most likely needed to achieve the desired research culture. Therefore, a research project has been conducted to examine the possibility of understanding and promoting research culture at a public university in Malaysia. Distinctive instruments were developed to achieve the purpose. This paper, thus, is aimed at examining the psychometric properties of a 44-item survey instrument developed to delve into the academic staff's intrinsic motivation in conducting research. The survey was administered to 326 academic staff from various faculties in a public university. The Rasch Measurement Model, which provides evidences on the fundamental measurement requirements of research instrument, was used to analyse the psychometric properties of the survey instrument using the Winsteps software program. The results of the Rasch analysis combined with qualitative investigation showed that the survey measured two distinct subscales or sub-dimensions of intrinsic motivation (namely positive and negative), which should be carried out separately in the final analysis. The resulted two subscales met the measurement requirements as evidenced by the individual analyses of the Rasch Model. Three misfit items were deleted from the second subscale (i.e. positive). Further items could be added to the two subscales to target the respondents with high ability measures. Recommendations were also given to revise the 5-point Likert scale used in the surveys in other related studies.

Keywords: Higher education; Research culture; Intrinsic motivation; Psychometric properties; Polytomous data; Rasch measurement model.

1. Introduction

Conducting research plays a fundamental and vital role at the higher learning institutions (HEIs). However, research culture amongst academics is highly debatable and ambiguous. Many academics claimed that they are doing what are required of them by the organization in relations to research. Thus, it is unclear whether the culture of research is genuinely in existence in HEIs.

The outcome of this investigation could help HEIs accomplish their missions, attract and maintain teaching staff of top professionalism, create and sustain a stimulating environment for the transmission and rapid assimilation of knowledge, transfer experience from one generation to another, and establish relations with other institutions of higher education, with professionals and specialists in various fields (Gregory *et al.*, 2009). Hence, research on creating and promoting 'research culture' at universities has been conducted in different parts of the world (Clemeña and Acosta, 2007; Evans, 2007;2009; Lewis and Simmons, 2010; Lodhi, 2012; Naoreen and Adeeb, 2014; Studman and Tsheko, 2007; Teshcko, 2007). To achieve this noble purpose, all parties involved in higher education should work together to ensure or secure all the physical, psychological and moral requirements of the research culture (Evans, 2007;2009; Lodhi, 2012).

An essential step is the thorough investigation of the current situations at the higher learning institutions, including attitudes, perceptions, opinions, knowledge, skills, and motivations of both the academic and non-academic staff. Survey and interview instruments have been commonly used to collect information in these areas (Creswell, 2013). However, measuring such constructs or traits is challenging due to the nature of the construct (i.e. how it is operationally defined), development of valid instruments, and models used to examine the psychometric properties of the instruments (Bond and Fox, 2015; Green and Frantom, 2002). Using tests or instruments that are valid and reliable (key indicators) to measure constructs is a crucial component of research quality (Kimberlin and Winterstein, 2008).

*Corresponding Author

1.1. The Rasch Measurement Model

With respect to the measurement models, they should ensure the properties that conform to the requirements of fundamental measurements so that valid inferences could be generated (Bond and Fox, 2015; Engelhard, 2000; Hambleton *et al.*, 1991; Wright, 1997). These requirements include a clear definition of the measured construct; using valid items that define the construct; providing consistent results; and using valid response patterns (Wright and Stone, 1979). The Rasch Measurement model is among those models that provide evidences on the fundamental measurement requirements of research instruments (Bond and Fox, 2015; Conrad *et al.*, 2010; Green and Frantom, 2002; Ingebo, 1997; Wright and Stone, 1979; Wright and Linacre, 1994). (Curtis and Boman, 2007) maintained that the Rasch model provides developers and researchers with more detailed diagnostic information on surveys and assessment instruments. Kimberlin and Winterstein (2008) said that "Rasch models and item-response theory (IRT) or latent-trait models have provided an alternative framework for understanding measurement and alternative strategies for judging the quality of a measuring instrument" (p.2281)

The Rasch Measurement Model, developed by the Danish mathematician George Rasch, could be used in both physical and psychological sciences. Mathematically, it relates the probability of a correct response on an item to the difference between a person parameter (ability) and an item parameter (item difficulty) (Bond and Fox, 2015; Stenner, 2004). Meaning that, the person's response to an item is dominated by the item difficulty and the person's ability (Bond and Fox, 2015; Green and Frantom, 2002). Two key propositions underlie the theoretical concept of the Rasch model (Bond and Fox, 2015; Wright and Stone, 1979). First, persons who are able have a greater likelihood of correctly answering all the items; second, easier items are likely to be answered or reached correctly by all persons (Bond and Fox, 2015). This dichotomous model has been extended to include other data types including the rating scale as in the current study (Boone and Noltemeyer, 2017). They further highlight that Rasch Model has been increasingly used in the field of psychology, education and other fields.

> Rasch measurement has been applied in a variety of ways in education, school psychology, and many other felds. It has been used to (a) develop, evaluate, and improve surveys and tests, and (b) facilitate the computation of Rasch "measures" that lead to data analysis and interpretation of greater confidence (because equal interval data are being utilized) (p.2).

The Rasch Measurement Model demonstrates if there are enough items that are positioned and spread along a unidimensional interval scale and targeted the persons with different abilities by providing reliability and separation indices for both items and persons (Bond and Fox, 2015; Green and Frantom, 2002; Stenner, 2004; Wright, 1996;1997; Wright and Stone, 2003). The Rasch Measurement Model also provides objectivity, one of the basic measurement requirements as it estimates the item difficulty and person ability independently (Bond and Fox, 2015; Green and Frantom, 2002; Wright and Stone, 1979; Wright and Linacre, 1987). The Principal Component Analysis of residuals and item and person fit statistics are used to assess or measure scale unidimensionality (Bond and Fox, 2015). Primarily, the fit statistics identify responses of unexpected patterns of item and person performance i.e. the items which are not working as intended and persons whose performance is inconsistent. To acquire these evidences, two types of fit indexes are used: INMSQ (Infit Mean Square) and OUTMSQ (Outfit Mean Square); see (Bond and Fox, 2015; Green and Frantom, 2002; Hula *et al.*, 2006; Wright and Stone, 1979; Wright and Linacre, 1987; Wright, 1993). In this respect, the data should fit the model.

In sum, since the Rash one-parameter model meets the requirements of fundamental measurement, it is commonly used to assess or examine the psychometric properties of instruments including test and surveys. It ensures that the used rating scale possesses "the highest-quality measures for the construct of interest" (Bond and Fox, 2015). They further explicate that "validity is an argument that entails judgement about meaning. Meaning stems from theory: from the type of questions we ask, the way that we ask them, of whom we ask them, and how we summarize and interpret the data" (p. 245).

1.2. Motivation and Intrinsic Motivation

To be motivated means to be moved to do something (Ryan and Deci, 2000). A motivated person is a person who feels, stimulates, inspires and energizes to do something and to work towards an end. In theory, if the performance is more attractive, the greater will be the effort of an individual. The performance is more attractive when it is very difficult and in contrast, it will be the least attractive if it is easy to achieve (Costin, 2014). Thus, motivation has been recognized as one of the major factors in improving individuals' performances (Guilloteaux and Dornyei, 2008).

Motivation can be divided into two parts namely intrinsic and extrinsic motivations. Extrinsic motivation refers to performance of a task in order to receive observable and tangible reward, recognition, respect and appreciation. In theory, extrinsic motivation constitutes four elements which are; 1) desire for affiliation; 2) fear of consequences; 3) ambitions; and 4) normative trend (Moldovan, 2014). The author also suggests that there is the possibility of the transformation of extrinsic motivation into intrinsic motivation through exposure and stimulations of the subjects' environment. Intrinsic motivation is the doing of an activity for its inherent satisfactions and enjoyment for reasons that lie within the activity itself rather than its consequences (Ryan and Deci, 2000; Wigfield and Eccles, 1992). As supported by (Cerasoli *et al.*, 2014) and (Pinder, 2011) on the importance of intrinsic motivation as not being instrumental towards other object of value rather than it motivates behaviors to be enjoyable, purposive and provide sufficient reasons for an individual to persist. In an academic situation, intrinsic motivation leads to a profounder processing, greater mastery, and better implementation of strategies (Covington, 2000). Thus, intrinsic motivation is a very significant drive for performing achievement-related activities because erudition comes as a by-product of engaging in an enjoyed task and one's self-determination. Evidently, intrinsic motivation remains a strong predictor

of performance regardless to the availability of incentives (Cerasoli *et al.*, 2014). (Hannam and Narayan, 2015) assert that intrinsic motivation helps individuals to have positive perception towards their nature of work, more creative and imaginative minds in solving problems. According to Shalley *et al.* (1987), individuals demonstrate high intrinsic motivation when they attain difficult goals and anticipate no external evaluation. This is to show that individual's intrinsic motivations depends on the difficulty of the goals rather than expectations.

1.3. Objective of the Study

The purpose of this paper is to provide a psychometric analysis of a scale used to measure the intrinsic motivation using the Rasch measurement model. The scale consists of 44 items related to intrinsic motivation on a five-point Likert-type scale ranged from Strongly Disagree (1), Disagree (2), Neutral (3), Agree (4), and Strongly Agree (5) completed by university academic staff.

2. Research Method

The research survey design was utilized in this study. A 44-item survey was individually administered to 326 academic staff randomly selected from various faculties in a public university in Malaysia to collect data on their intrinsic motivation in conducting research (Creswell, 2013). The survey was developed based on a thorough literature review and interviews with academic staff in psychology and education. Several items were added and others were modified in line with the feedback given by the academic staff in the content and face validity process, given a total of 44 items listed in the final survey instrument. For the sample size, when using the Rasch Model it is recommended that the most reliable interpretation comes from a sample closer to 100 (Bond and Fox, 2015; Green and Frantom, 2002; Linacre, 1994; Wright and Stone, 1979). In this study, the one parameter Rasch modeling for polytomous data was used to examine the psychometric properties of the instrument. As mentioned earlier, Rasch Modeling converts the ordinal raw data collected through surveys into interval measures(Bond and Fox, 2015; Curtis and Boman, 2007). It provides necessary diagnostic information about the instruments at the macro level (e.g. item and scale reliability indices); meso level (e.g. location parameters and item fit indices); and micro levels (e.g. individual item thresholds through their locations and standard errors(Curtis and Boman, 2007) . The Rasch analysis was conducted using Winsteps software, 3.72.1 (Linacre, 2011). The results were depicted in Tables and Figures. It is important to highlight that the negative items were recoded in the first run. However, many of the items showed misfit values and negative correlations. Below is the analysis for all the survey items without recoding, followed by the analysis for the resulted two scales (negative, 20 items and positive, 24 items).

3. Results and Discussion

3.1. Screening and Cleaning Data

Each survey was examined to ensure the integrity of the data collected. The incomplete surveys were excluded from the analysis. Then the individual responses and demographic information were keyed into SPSS file (version 16) and checked for data entry errors (Pallant, 2013). The final SPSS data set was imported to Winsteps software program, 3.72.1 for Rasch modelling (Linacre, 2011).

3.2. Adequacy of Research Instrument – Overall

The initial analyses of the collected data were conducted to check the adequacy of scale measurement developed to inquire data on intrinsic motivation in conducting research among academic staff in a public university. Item and person reliability indices were examined followed by determining the validity of items through three indicators: Item polarity, Item Fit, and Unidimensionality (Bond and Fox, 2015). In case any sub-dimension is identified, a separate measure for this dimension should be created(Bond and Fox, 2015). It is important to mention that the negative items were recoded, but no improvement on the measures were observed. Below is the analysis for all items; followed by the resulted two sub-scales.

Based on summary statistics (after deleting most misfit persons), Table 1 shows that reliability of item difficulty measures is very high (1.00). This suggests that the ordering of item difficulty is highly replicable with other comparable sample. The item separation index was > 2. Table 1 also reveals that the person reliability is fair .74, suggesting that it is likely high that the ordering of students can be replicated with other items of the same difficulty. The person separation index is 1.69, indicating that the items can divide respondents into two levels. Table 1 also displays item polarity and item fit statistics.

The Item polarity (i.e. point-measure correlation coefficient) indicates the extent to which scale items are working in the same direction to define the measured construct. Negative and zero values indicate that items or examinees are working in the wrong direction, and relatively high positive values are desired (Linacre, 2010). Table 1 shows that the point measure correlation (PTMEA CORR.) for the 44 items are positive, but some items are below 0.3 (between 0.12 - 0.29).

Fit statistics are used to ensure that the items are contributing meaningfully to the measurement of the variable or construct as expected by the model (Bond and Fox, 2015). The two major fit statistics used are the infit and outfit Mean-square statistics. These statistics indicate the amount of "distortion of measurement system" (Linacre, 2010). Among the recommended fit statistics ranges for scale items (0.5-1.5). The items within the recommended range are considered productive or meaningful to the measurement; values below this range indicate that the items are considered as over fitting, while those above this range are considered as misfitting (Bond and Fox, 2015; Wright *et al.*, 1994). Table 1 shows that the infit and outfit mean-square of individual items are within the specified range (0.5 - 1.5), indicating that they were working as expected by the model.

Table-1. Items measures, fit statistics and point-measure correlation coefficient

Item		Measure	S.E	Infit MNSQ	Outfit MNSQ	PT-Measure CORR
1	I enjoy doing research	-1.18	0.08	1.16	1.19	0.13
2	I think doing research is a boring activity	1.64	0.07	1.33	1.3	0.27
3	Doing research does not hold my attention at all	1.54	0.06	1.29	1.29	0.3
4	I believe that I am good at doing research	-0.52	0.06	0.73	0.74	0.22
5	I believe I am doing well in research compared to my colleagues	0.12	0.06	0.63	0.64	0.29
6	After working on my research, I feel competent	-0.71	0.07	0.65	0.69	0.25
7	I am satisfied with my performance in doing research	-0.3	0.06	0.84	0.86	0.25
8	I am quite skilled at doing research	-0.44	0.06	0.76	0.77	0.26
9	Research is an activity that I cannot do very well	1.31	0.06	0.85	0.85	0.34
10	I always complete my research on time	-0.06	0.06	0.72	0.73	0.23
11	I put a lot of effort into doing research	-0.88	0.07	0.78	0.78	0.25
12	I do not try very hard to do well in my research	1.3	0.06	1.1	1.11	0.31
13	I try very hard to complete my research	-0.77	0.07	1.00	1.04	0.12
14	It is important for me to do well in my research	-1.55	0.08	0.9	1.01	0.15
15	I do not put much energy into completing my research	1.37	0.06	1.3	1.33	0.13
16	It is important for me to publish my research in a journal	-1.32	0.08	1.08	1.12	0.18
17	I feel rewarded when my research is recognized by my senior	-1.29	0.08	1.00	1.05	0.27
18	I do not feel nervous at all while doing research	-0.09	0.06	1.36	1.36	0.14
19	I feel very tense while doing research	0.61	0.05	0.91	0.91	0.39
20	I feel anxious while completing on my research	0.23	0.06	0.93	0.93	0.46
21	I cannot cope with my research while having extra workload at my workplace	-0.36	0.06	1.2	1.26	0.23
22	My superior always pressures me into doing research	0.34	0.05	0.87	0.89	0.34
23	I feel unhappy to find out that too much attention is given to research	0.67	0.05	1.06	1.06	0.23
24	I feel pressured if my research could not be published within six months after completion	0.07	0.06	0.84	0.85	0.35
25	I believe I have some choice in teaching or doing research	0.08	0.06	1.08	1.09	0.29
26	I feel like is not my own choice to do research	0.67	0.05	1.04	1.04	0.27
27	I do research because I have no choice	1.07	0.06	1.07	1.07	0.34
28	I do research because I want to do it	-0.77	0.07	0.97	0.99	0.17
29	I wish to teach without doing research	0.91	0.06	1.26	1.26	0.28
30	I wish to do research without teaching	0.83	0.06	1.05	1.04	0.35
31	I believe doing research is valuable for me	-1.18	0.08	0.94	0.97	0.25
32	I think that doing research is useful for my career development	-1.45	0.08	1.04	1.03	0.30
33	I think research is important to promote professionalism	-1.51	0.08	1.17	1.13	0.36
34	I believe doing research could be a benefit to me in monetary terms	-0.11	0.06	1.02	1.05	0.33
35	I believe publishing in a research journal is an advantage for me	-1.38	0.08	1.00	0.98	0.33
36	I value research more than publication	-0.09	0.06	0.96	0.97	0.35
37	I use my research findings in my teaching	-0.74	0.07	0.88	0.87	0.25
38	I feel bad about doing research	1.75	0.07	0.94	0.99	0.22
39	I feel depressed doing research	1.56	0.07	0.97	0.97	0.3
40	I am motivated in doing research	-0.7	0.07	0.92	0.95	0.18
41	I feel my future is hopeless and things cannot improve even with research	1.42	0.06	1.09	1.07	0.38
42	I feel like a failure with respect to research and publication	1.31	0.06	0.98	1	0.36
43	I get as much satisfaction out of doing research	-0.48	0.06	1.06	1.06	0.32
44	I am very happy doing research	-0.93	0.07	1.06	1.05	0.26
Means		.00	.06	1.00	1.01	
Reliability and Separation						
Item Reliability		1.00				
Item separation		15.31				
Person Reliability		.74				
Person Separation		1.69				

In the Rasch model, the data must fit the model usefully and items must work together to measure a single unidimensional construct. The principal component analysis of residuals is used to test unidimensionality. Table 2 shows that unidimensionality is violated. Though the variance explained by the measure is 51% (accepted), the factor analysis of the residuals in the first contrast indicates high unexplained variance was (10.9 units or 12.1 %). Meaning that the largest factor extracted from the residuals was equivalent to 10.9 units which have the strength of about 11 items (more than 3 items needed to be considered as a second factor) (Linacre, 2010). So, the PCA likely defined two subscales of sub-dimensions of the scale as shown in Table 2.

Table-2. Standardized residual variance (in eigenvalue units) (All Items)

	Empirical		**Modeled**	
Total raw variance in observations	90.5	100.00%		100.00%
Raw variance explained by measures	46.5	51.4%		51.0%
Raw variance explained by persons	4.3	4.8%		4.7%
Raw Variance explained by items	42.2	46.6%		46.3%
Raw unexplained variance (total)	44.0	48.6%	100.00%	49.0%
Unexplained variance in 1st contrast	10.9	12.1%	24.8%	

STANDARDIZED RESIDUAL LOADINGS FOR ITEM (SORTED BY LOADING)

CON-TRAST	LOADING	INFIT MEASURE	INFIT MNSQ	OUTFIT MNSQ	ENTRY NUMBER	ITEM		LOADING	INFIT MEASURE	INFIT MNSQ	OUTFIT MNSQ	ENTRY NUMBER	ITEM
1	.77	-.93	1.06	1.05	A 44	IM45		-.76	1.75	.94	.99	a 38	IM39
1	.66	-.70	.92	.95	B 40	IM41		-.72	1.07	1.07	1.07	b 27	IM27
1	.64	-.77	.97	.99	C 28	IM28		-.71	1.56	.97	.97	c 39	IM40
1	.63	-1.18	1.16	1.19	D 1	IM1		-.67	.91	1.26	1.26	d 29	IM29
1	.62	-1.18	.94	.97	E 31	IM31		-.66	1.31	.85	.85	e 9	IM9
1	.61	-.44	.76	.77	F 8	IM8		-.60	1.31	.98	1.00	f 42	IM43
1	.60	-.88	.78	.78	G 11	IM11		-.60	.67	1.06	1.06	g 23	IM23
1	.56	-1.51	1.17	1.13	H 33	IM33		-.55	.67	1.04	1.04	h 26	IM26
1	.56	-.52	.73	.74	I 4	IM4		-.52	1.30	1.10	1.11	i 12	IM12
1	.53	-1.55	.90	1.01	J 14	IM14		-.48	1.42	1.09	1.07	j 41	IM42
1	.52	-.71	.65	.69	K 6	IM6		-.48	1.64	1.33	1.30	k 2	IM2
1	.50	-.48	1.06	1.06	L 43	IM44		-.43	1.54	1.29	1.29	l 3	IM3
1	.48	-.30	.84	.86	M 7	IM7		-.42	1.37	1.30	1.33	m 15	IM15
1	.42	-1.45	1.04	1.03	N 32	IM32		-.41	.61	.91	.91	n 19	IM19
1	.42	.12	.63	.64	O 5	IM5		-.33	-.36	1.20	1.26	o 21	IM21
1	.39	-1.32	1.08	1.12	P 16	IM16		-.31	.34	.87	.89	p 22	IM22
1	.38	-.74	.88	.87	Q 37	IM38		-.27	.23	.93	.93	q 20	IM20
1	.38	-.06	.72	.73	R 10	IM10		-.12	.08	1.08	1.09	r 25	IM25
1	.36	-.77	1.00	1.04	S 13	IM13		-.10	.83	1.05	1.04	s 30	IM30
1	.35	-1.38	1.00	.98	T 35	IM35		-.09	.07	.84	.85	t 24	IM24
1	.29	-.09	1.36	1.36	U 18	IM18							
1	.29	-1.29	1.00	1.05	V 17	IM17							
1	.09	-.11	1.02	1.05	v 34	IM34							
1	.09	-.09	.96	.97	u 36	IM36							

Remedial solutions were conducted as deleting misfit persons and few items and recoding the negative items. No improvements were observed. Therefore, the recommendation was to analyze the scale separately. Further qualitative investigation carried out by experts in psychology, highlighted that this scale has two subscales or sub-dimensions. The first dimension with 24 items is representing the positive items, and the other one with 20 items is representing the negative items.

The category statistics of the scale showed that all categories were used by the respondents more than recommended, > 10 (Bond and Fox, 2015). The average measure increases monotonically across the categories. The outfit mean squares are less than recommended value < 2 (Bond and Fox, 2015), indicating no noise in the measurement process. However, distances between the adjacent thresholds are less than1.4 (Bond and Fox, 2015), which needs more investigation. It is also noticed that the mid-category (neutral) is used in the scales very often. It might attract many of the respondents from different abilities.

From the analysis above, it could be concluded that the scale items are related to intrinsic motivation and have meaningful contribution to the construct. Nevertheless, the PCA analyses showed violation in scale unidimensionality. The scale could be measuring two sub-dimensions (positive and negative). Other solutions were performed to examine the unidimensionality by deleting items and misfit persons, rescoring items, and collapsing categories, but no improvement was observed. Qualitative investigation by expert in psychology asserted that the items are measuring two sub-scales (positive and negative)

Having the initial Rasch analysis of the overall scale shown that the scale consists of two dimensions, it is recommended to analyze each scale individually for better measurement (Bond and Fox, 2015). Below are the analyses of the two scale; (negative with 20 items and positive with 24 items respectively).

3.3. First Subscale (Negative Items)

The Rasch Model analyses of the 20-item subscale were conducted to check the psychometric measurement properties of this subscale. Item and person reliability indices were examined followed by determining the validity of items through three indicators: Item polarity, Item Fit, and Unidimensionality (Bond and Fox, 2015).

Table 3 shows that reliability of item difficulty measures is very high (.99). This suggests that the ordering of item difficulty is highly replicable with other comparable sample. The item separation index is > 2. Table 3 also reveals that the person reliability is .83, suggesting that it is likely high that the ordering of students can be replicated with other items of the same difficulty. The person separation index is 2.23, indicating that the items can divide respondents into two levels. Table 3 also displays item polarity and item fit statistics.

Table 3 shows that the point measure correlation (PTMEA CORR.) for the 20 items are positive, ranged (.28-.72), indicating the items are working in the same direction to define the measured construct. Table 3 also shows that the infit and outfit mean square of individual items are within the specified range (0.5 - 1.5), indicating that there working as expected by the model. However, only one item has outfit mean square (1.54) due to misfit persons. The analysis ensures that the items are contributing meaningfully to the measurement of the construct as expected by the model (Bond and Fox, 2015).

Table-3. First subscale items measures, fit statistics and point-measure correlation coefficient

Item		Measure	S.E	Infit MNSQ	Outfit MNSQ	PT-Measure CORR
2	I think doing research is a boring activity	0.88	0.07	1.32	1.26	0.51
3	Doing research does not hold my attention at all	0.76	0.07	1.34	1.41	0.48
9	Research is an activity that I cannot do very well	0.47	0.07	0.64	0.62	0.67
12	I do not try very hard to do well in my research	0.46	0.07	1.01	1.01	0.57
15	I do not put much energy into completing my research	0.55	0.07	1.28	1.54	0.42
19	I feel very tense while doing research	-0.38	0.06	0.9	0.91	0.56
20	I feel anxious while completing on my research	-0.83	0.06	1.04	1.07	0.5
21	I cannot cope with my research while having extra workload at my workplace	-1.5	0.06	1.15	1.19	0.43
22	My superior always pressures me into doing research	-0.71	0.06	0.95	1	0.47
23	I feel unhappy to find out that too much attention is given to research	-0.31	0.06	0.84	0.85	0.61
24	I feel pressured if my research could not be published within six months after completion	-1.02	0.06	1.15	1.18	0.31
25	I believe I have some choice in teaching or doing research	-1.01	0.06	1.4	1.49	0.28
26	I feel like is not my own choice to do research	-0.3	0.06	0.9	0.97	0.57
27	I do research because I have no choice	0.18	0.06	0.76	0.73	0.72
29	I wish to teach without doing research	-0.01	0.06	0.97	0.96	0.65
30	I wish to do research without teaching	-0.11	0.06	1.4	1.48	0.32
38	I feel bad about doing research	1.01	0.08	0.65	0.63	0.68
39	I feel depressed doing research	0.78	0.07	0.72	0.69	0.69
41	I feel my future is hopeless and things cannot improve even with research	0.61	0.07	1.06	1.08	0.56
42	I feel like a failure with respect to research and publication	0.48	0.07	0.81	0.88	0.64
Means		.00	.07	1.02	1.05	
Reliability and Separation						
Item Reliability		.99				
Item separation		10.85				
Person Reliability		.83				
Person Separation		2.23				

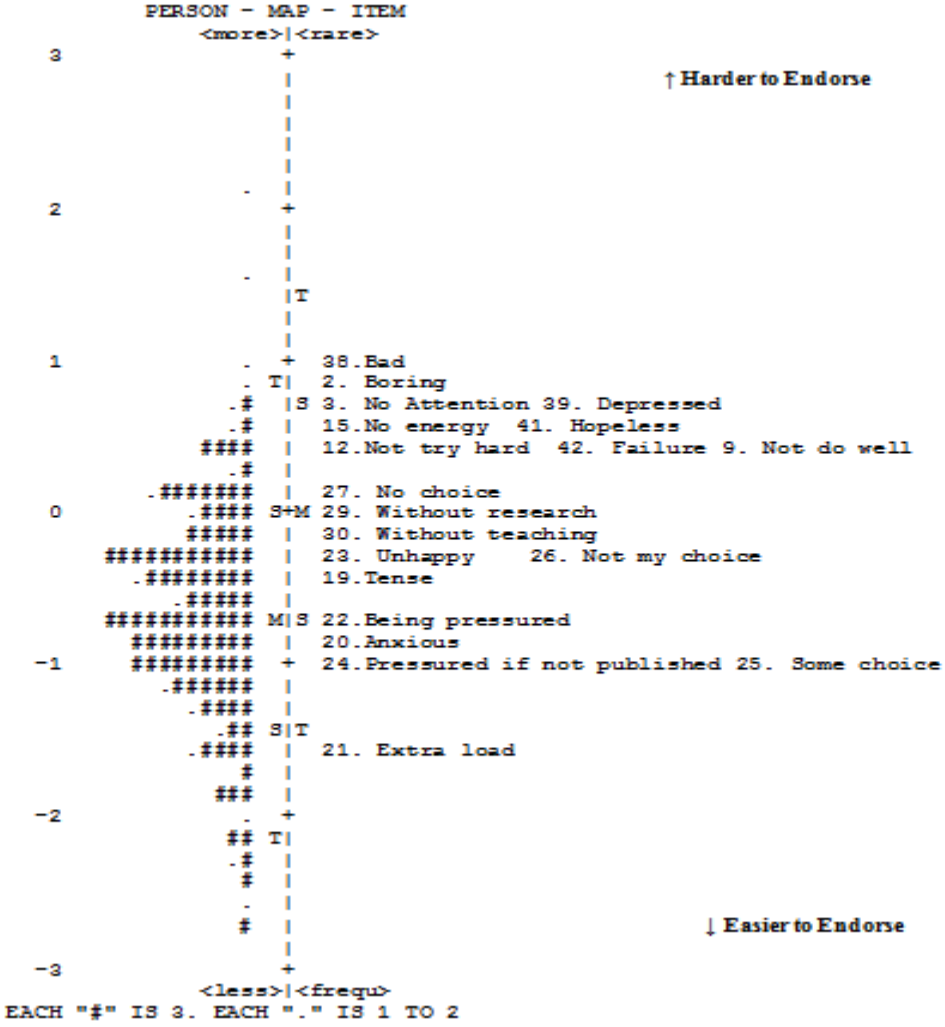
The principal component analysis of residuals is used to test unidimensionality i.e. to ensure that data fit the model usefully and items are working together to measure a single unidimensional construct. Table 4 shows that unidimensionality is not violated. The variance explained by the measure is 44.2, and the largest factor extracted from the residuals is equivalent to 2.6 units which have the strength of about 3 items (Linacre, 2010).

Table-4. Standardized residual variance (in eigenvalue units) (First Sub Scale Items)

	Empirical		Modeled	
Total raw variance in observations	35.9	100.00%		100.00%
Raw variance explained by measures	15.9	44.2%		45.4%
Raw variance explained by persons	4.0	11.2%		11.4%
Raw Variance explained by items	11.9	33.1%		33.9%
Raw unexplained variance (total)	20.0	55.8%	100.00%	54.6%
Unexplained variance in 1st contrast	2.6	7.3%	13.2%	

Figure 1 (Item-Person Map) shows the distribution of all items and persons on one logit scale. The item difficulty measure spanned from -1.50 logits to 1.10 logit, while the person ability measure spanned from -2.71 to 2.12 logits. There were no significant visible gaps between item distribution, except for two gaps at the upper and lower ends of the scale. Looking at the overlapping items, it is found that those item are not measuring the same aspects of the measured constructs. It is recommended to add a few more items at the top and end of the scale for more precise person estimates. Overall, it could be noticed that targeting of this sample's levels has been well. The hierarchy of the items is almost as expected due to qualitative investigation with experts in psychology i.e. the order of the items make sense in this context.

Figure-1. Item-person map (first subscale)



For the category functions, the category use statistics (i.e. category frequencies and average measures) for each options were examined. The former indicates how many respondents actually chose any particular response category. A recommended minimum number of responses per category is 10. Table 5 shows that each category has met this criterion. The average measures are defined as the average of the ability estimates for all persons in the sample who chose that particular response category, with the average calculated across all observations in that category as cited in Bond and Fox (2015). They further elaborated that these average measures should increase monotonically. Table 5 shows that the average measures for this scale are increasing monotonically (-1.72 < -0.86 < -0.25 < 0.40). The thresholds are also increasing monotonically (-1.66 < -.22 < 0.23 < 1.65). However, the distance between adjacent categories, mainly between categories 3 and 4 is less than 1.4 (Bond and Fox, 2015). These could be collapsed for better improvement (Bond and Fox, 2015; Green and Frantom, 2002; Linacre, 2002). For the fit statistics, they provide another criterion for assessing the quality of rating scales. Table 5 the fit statistics of each rating scale are less than 2, indicating that there is no noise in the measurement process.

Table-5. Summary of category structure. Model="R"

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	STRUCTURE CALIBRATN	CATEGORY MEASURE	
1	1	1473	23	-1.72	-1.56	.81	.87	NONE	(-2.92)	1
2	2	2222	34	-.86	-.95	1.01	.94	-1.66	-1.16	2
3	3	1430	22	-.25	-.36	.89	.87	-.22	.00	3
4	4	1056	16	.20	.23	1.05	1.09	.23	1.16	4
5	5	339	5	.47	.82	1.51	1.74	1.65	(2.91)	5

The same analysis process was utilized to examine the items defined the second subscale (positive items).

3.4. Second Subscale (Positive Items)

The Rasch Model analyses of the 24-item subscale (positive) was conducted to check psychometric measurement properties of the scale. Item and person reliability indices were examined followed by determining the validity of items through three indicators: Item polarity, Item Fit, and Unidimensionality.

It is essential first to highlight that the analysis for this subscale showed that the there is a problem in the category functions. The respondents seem that they did not use the categories as intended as shown in Table 6 and Table 7. Table 6 shows that though the category frequencies have more than 10 responses as recommended by Bond and Fox (2015), the average measures do not increase monotonically (0.97 > 0.24 < 0.60 < 1.38 < 2.47 across the rating scale response categories (1,2,3,4,5). In addition, the outfit mean squares for category 1 is greater than 2, indicating that this category is introducing more noise more than meaning into the measurement (Bond and Fox, 2015).

Table-6. Summary of category structure.

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	STRUCTURE CALIBRATN	CATEGORY MEASURE	
1	1	89	1	.79	-.41	1.99	3.01	NONE	(-3.09)	1
2	2	444	6	.24*	.10	1.16	1.27	-1.77	-1.46	2
3	3	1500	19	.60	.72	.88	.92	-.82	-.20	3
4	4	3612	46	1.38	1.45	.93	.81	.20	1.40	4
5	5	2179	28	2.47	2.34	.87	.90	2.39	(3.56)	5

According to Bond and Fox (2015) it is recommended to collapse adjacent categories (i.e. 1 and 2) to improve the variable interpretation. The collapsing categories 1 and 2 has improved the rating scale diagnostics as shown in Table 7.

Table-7. Summary of category structure. Model="R"

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	STRUCTURE CALIBRATN	CATEGORY MEASURE	
1	1	533	7	-.33	-.73	1.38	1.61	NONE	(-2.71)	1
2	2	1500	19	-.08	.02	.91	.97	-1.39	-.97	3
3	3	3612	46	.77	.86	.94	.84	-.44	.80	4
4	4	2179	28	1.95	1.83	.87	.90	1.83	(3.01)	5

Having deleted the three misfit items, the reliability of item difficulty measures is still very high (.99) as shown in Table 8. This suggests that the ordering of item difficulty is highly replicable with other comparable sample. The item separation index is > 2. Table 8 also reveals that the person reliability is .87, suggesting that it is likely high that the ordering of students can be replicated with other items of the same difficulty. The person separation index is 2.59, indicating that the items can divide respondents into two levels. Table 8 also displays item polarity and item fit statistics.

Table-8. Second subscale items measures, fit statistics and point-measure correlation coefficients

Item		Measure	S.E	Infit MNSQ	Outfit MNSQ	PT-Measure CORR
1	I enjoy doing research	-0.55	.09	1.05	1.06	0.60
4	I believe that I am good at doing research	0.55	.09	.83	.84	0.60
5	I believe I am doing well in research compared to my colleagues	1.76	.08	.98	1.00	0.53
6	After working on my research, I feel competent	0.23	.09	.80	.81	0.58
7	I am satisfied with my performance in doing research	0.95	.08	1.08	1.10	0.57
8	I am quite skilled at doing research	0.69	.08	.76	.79	0.64
10	I always complete my research on time	1.42	.08	1.15	1.18	0.48
11	I put a lot of effort into doing research	-0.05	.09	.77	.75	0.65
13	I try very hard to complete my research	0.1	.09	1.26	1.31	0.45
14	It is important for me to do well in my research	-1.09	.10	.85	.84	0.56
16	It is important for me to publish my research in a journal	-0.76	.10	1.30	1.39	0.44
17	I feel rewarded when my research is recognized by my senior	-0.71	.10	1.34	1.44	0.41
18	I do not feel nervous at all while doing research	Deleted item /misfit				
28	I do research because I want to do it	0.1	.09	.90	.99	0.61
31	I believe doing research is valuable for me	-0.55	.09	.89	.88	0.62
32	I think that doing research is useful for my career	-0.97	.10	1.05	1.06	0.53

	development					
33	I think research is important to promote professionalism	-1.06	.10	1.03	.97	0.61
34	I believe doing research could be a benefit to me in monetary terms	Deleted item/misfit				
35	I believe publishing in a research journal is an advantage for me	-0.85	.10	1.26	1.35	0.44
36	I value research more than publication	Deleted item/misfit				
37	I use my research findings in my teaching	0.17	.09	1.20	1.30	0.46
40	I am motivated in doing research	0.22	.09	.75	.79	0.68
43	I get as much satisfaction out of doing research	0.58	.08	1.20	1.35	0.58
44	I am very happy doing research	-0.17	.09	.80	.78	0.73
Means		-0.55	.09	1.05	1.06	0.60
Reliability and Separation						
Item Reliability		.99				
Item separation		8.23				
Person Reliability		.87				
Person Separation		2.59				

Table 8 shows that the point measure correlation (PTMEA CORR.) for the 24 items are positive, ranged (.41-.73), indicating the items are working in the same direction to define the measured construct. Table 8 also shows that the infit and outfit mean-square of 21 individual items are within the specified range (0.5 - 1.5), indicating that they are working as expected by the model. However, three items, namely 18,34,36 have infit and outfit mean squares above (1.5), which were deleted from the analysis. Qualitative investigation of these items showed that item 18 (*did not feel nervous*) has double negations, which might cause confusion/difficulty to the participants to endorse this item; while item 34 (*benefits in monetary*) might be related to extrinsic motivation rather than intrinsic motivation. For item 36 (*value research than publication*), the participants might be confused about the difference between publication and research in this item or they might value both research and publications.

The principal component analysis of residuals is used to test unidimensionality i.e. to ensure that data fit the model usefully and items are working together to measure a single unidimensional construct. Table 9 shows that unidimensionality is not violated. When deleted persons with negative correlation, a little improvement was observed as shown in Table 9. The variance explained by the measure is 43.9 and the largest factor extracted from the residuals is equivalent to 2.8 units which have the strength of about 3 items (Linacre, 2010).

Table-9. Standardized residual variance (in eigenvalue units) (Second Sub Scale Items)

	Empirical		**Modeled**	
Total raw variance in observations	37.4	100.00%		100.00%
Raw variance explained by measures	16.4	43.90%		44.30%
Raw variance explained by persons	7.4	19.90%		20.00%
Raw Variance explained by items	9	24.00%		24.20%
Raw unexplained variance (total)	21	56.10%	100.00%	55.70%
Unexplained variance in 1st contrast	2.8	7.50%	13.40%	

Figure 2 (Item-Person Map) shows the distribution of all items and persons on one logit scale. The item difficulty measure spanned from -1.09 logits to 1.76 logit, while the person ability measure spanned from -1.25 to 6.7 logits. There was a significant visible gap at the upper end of the scale. Looking at the overlapping items, it is found that those item are not measuring the same aspects of the measured constructs. It is recommended to add a few more items at the top end of the scale for more precise person estimates. Overall, the hierarchy of the items is almost as expected due to qualitative investigation with experts in psychology.

The analyses also showed that there are respondents with negative correlation and infit and outfit mean squares above the recommended value, 1.5. They were deleted for better measurement. As a result, little improvement was observed mainly on the variance explained by measures.

Figure-2. Item-person map (second subscale)

4. Implications and Conclusions

In order to make research culture as a success, the researchers must be willing to commit themselves towards the mission of the HEIs. Intrinsic motivation is a good indicator to show the level of commitment of researchers and whether the research culture will be nurtured and established within institutions of higher learning. There is a need for academic staff who possess research skills, and with high levels of (intrinsic) motivation in conducting research to promote a research culture at universities. Valid and reliable instruments to measure the staff intrinsic motivation are highly demanded to provide more accurate and reliable results. The initial analysis of Rasch Model analysis showed that the survey instrument used to measure the intrinsic motivation among academic staff was measuring two subscales or sub-dimensions of intrinsic motivation (positive and negative). The individual Rasch Model analysis of the two scales met the fundamental measurement requirements in terms of person and item reliability and separation indices, item correlation, fit statistics, unidimensionality, and targeting. However, three misfit items were deleted from the scale with positive items. Overall, the hierarchical order of the items that defined the two subscales make sense and sounds reasonable. It has been recommended to add further items mainly at the top and bottom ends of the scales to target the respondents with high and low ability measures. Recommendations were also given to revise the 5-point Likert scale used in the surveys in other related studies.

References

Bond, T. and Fox, C. M. (2015). *Applying the rasch model, Fundamental measurement in the human sciences.* Routledge: New York and London.

Boone, W. J. and Noltemeyer, A. (2017). Rasch analysis, A primer for school psychology researchers and practitioners. *Cogent Education*, 4: Available: https://doi.org/10.1080/2331186X

Cerasoli, C. P., Nicklin, J. M. and Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin*, 140(4): 980–1008. Available: https://doi.org/10.1037/a0035661

Clemeña, R. and Acosta, S., 2007. "Developing research culture in philippine higher education institutions, Perspectives of university faculty." In *Paper presented at the regional seminar competition, cooperation and change in the academic profession, Shaping higher education's contribution to knowledge and research, september18–19 in hangzhou, in china.*

Conrad, K., Conrad, K., Riley, B., Funk, R. and Dennis, M. (2010). *Validation of the current withdrawal scale (cws) to the rasch measurement model, Gain methods report 1.0. Chicago, il, Chestnut health systems. Retrieved from tba, 3.*

Costin, A. (2014). *Aspecte privind protecţia copilului în regim rezidenţial, Aspects regarding the protection of the child in residential regime.* Pro Universitaraia Publishing House: Bucharest. 82.

Covington, M. V. (2000). Goal theory, Motivation, And school achievement, An integrative review. *Annual Review of Psychology,* 51(1): 171–200. Available: https://doi.org/10.1146/annurev.psych.51.1.171

Creswell, J. W. (2013). *Research design, Qualitative, Quantitative, And mixed methods approaches.* Sage publications: London.

Curtis, D. D. and Boman, P. (2007). X-Ray Your Data with Rasch. *International Education Journal,* 8(2): 249-59.

Engelhard, G. (2000). Historical view of the influences of measurement and reading theories on the assessment of reading. *Journal of Applied Measurement,* 2(1): 1-26.

Evans, L., 2007. "Developing research cultures and researchers in he, The role of leadership." In *Paper presented at the Annual Conference of the Society for Research into Higher Education, SRHE.*

Evans, L. (2009). Developing research capacity in the social sciences: a professionality based model. *International Journal for Researcher Development,* 1(2): 134-49.

Green, K. E. and Frantom, C. G., 2002. "Survey development and validation with the Rasch model." In *Paper presented at the International Conference on Questionnaire Development, Evaluation, and Testing, Charleston, SC.*

Gregory, B. T., Harris, S. G., Armenakis, A. A. and Shook, C. L. (2009). Organizational culture and effectiveness, A study of values, Attitudes, And organizational outcomes. *Journal of Business Research,* 62(7): 673-79.

Guilloteaux, M.-J. and Dornyei, Z. (2008). Motivating language learners, A classroom-oriented investigation of student motivation. *TESOL Quarterly,* 42(1): 55–77. Available: https://doi.org/10.1002/j.1545-7249.2008.tb00207.x

Hambleton, R. K., Swaminathan, H. and Rogers, H. J. (1991). *Fundamentals of item response theory.* Sage Publications: London.

Hannam, K. and Narayan, A. (2015). Intrinsic Motivation, Organizational justice, And creativity. *Creativity Research Journal,* 27(2): 214–24. Available: https://doi.org/10.1080/10400419.2015.1030307

Hula, W., Doyle, P. J., McNeil, M. R. and Mikolic, J. M. (2006). Rasch modeling of revised token test performance: Validity and sensitivity to change. *Journal of Speech, Language, and Hearing Research,* 49(1): 27-46.

Ingebo, G. S. (1997). *Probability in the measure of achievement.* Mesa Press: Chicago, IL.

Kimberlin, C. L. and Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *AMJ Health Syst Pharm,* 65(23): 2276-84.

Lewis, T. and Simmons, L. (2010). Creating research culture in caribbean universities. *International Journal of Educational Development,* 30(4): 337-44.

Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions,* 7(4): 328.

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Jouranl of Applied Measurement,* 3(1): 85-106.

Linacre, J. M. (2010). A user's guide to winstep and ministep rasch-model computer programs. *Program Manual,* 3(69): 1.

Linacre, J. M. (2011). *Winsteps rasch measurement, Version 3.72.1, Software.* Chicago: MESA Press.

Lodhi, A. S. (2012). A pilot study of researching the research culture in Pakistani public universities, The academics' perspective. *Procedia-Social and Behavioral Sciences,* 31: 473-79.

Moldovan, O. (2014). Intrinsic and extrinsic motivation to primary school children. *Journal Plus Education,* 10(1): 203-11.

Naoreen, B. and Adeeb, M. A. A. (2014). Investigating academic research culture in public sector universities of Pakistan. *Procedia-Social and Behavioral Sciences,* 116: 3010-15. Available: https://bit.ly/2CfmVR0

Pallant, J. (2013). *SPSS survival manual.* McGraw-Hill Education: UK.

Pinder, C. (2011). *Work motivation in organizational behaviour.* 2nd edn: Taylor and Francis: London.

Ryan, R. M. and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, Social development, And well-being. *American Psychologist,* 55(1): 68–78. Available: https://doi.org/10.1037/0003-066X.55.1.68

Shalley, C. E., Oldham, G. R. and Porac, J. F. (1987). Effects of goal difficulty, Goal-setting method, And expected external evaluation on intrinsic motivation. *Academy of Management Journal,* 30(3): 553–63. Available: https://doi.org/10.2307/256014

Stenner, A. J. (2004). Does the reader comprehend the text because the reader is able or because the text is easy? The 13th international objective measurement workshop.

Studman, C. G. and Tsheko, G. N. (2007). Evaluation of strategies for building a research culture-an empirical case study at an African university. *Journal of Research Administration,* 38(2): 109-18.

Teshcko, G. N. (2007). Evaluation of strategies for building a research culture-an empirical case study at an African university. *Journal of Research Administration,* 38: 109.

Wigfield, A. and Eccles, J. S. (1992). The development of achievement task values, A theoretical analysis. *Developmental Review,* 12(3): 265–310. Available: https://doi.org/10.1016/0273-2297(92)90011-P

Wright, B. (1993). Thinking with raw scores. *Rasch Measurement Transactions,* 7(2): 299-300.

Wright, B. (1996). Reliability and separation. *Rasch Measurement Transactions,* 9(4): 472.

Wright, B. (1997). A history of social science measurement. *Educational Measurement, Issues and Practice,* 16(4): 33-45.

Wright, B. and Stone, M. H. (1979). *Best design test, A handbook for rasch measurement.* Mesa Press: Chicago.

Wright, B. and Linacre, J. (1987). Rasch model derived from objectivity. *Rasch Measurement Transactions,* 1(1): 5-6.

Wright, B. and Linacre, J. M. (1994). *The rasch model as a foundation for the lexile framework.* Unpublished manuscript.

Wright, B. and Stone, M. (2003). Five steps to science, Observing, Scoring, Measuring, Analyzing, And applying. *Rasch Measurement Transactions,* 17(1): 912-13.

Wright, B., Gustafson, J. and Martin-Lof, P. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions,* 8(3): 370.