EPRA International Journal of Environmental Economics, Commerce and Educational Management
Journal DOI: 10.36713/epra0414 |ISI I.F Value: 0.815|SJIF Impact Factor (2021): 7.743 ISSN: 2348 – 814X
Volume: 9 | Issue: 4 | April 2022

COMPARATIVE PSYCHOMETRIC ANALYSIS OF LIKERT AND EXPANDED SCALE FORMATS USING GRADED RESPONSE MODEL

Amadioha Ambrose (PhD)[1], Longjohn Ibiene Tandi (PhD)[2], Ajala Ibrahim (PhD)[3]

[1]*Department of Educational Psychology, Guidance and Counselling, University of Port Harcourt, Port Harcourt, Nigeria*
[2,3]*Department of Educational Psychology, Guidance and Counselling, Ignatius Ajuru University of Education, Port Harcourt, Rivers State*

ABSTRACT

Despite the popularity of the Likert scale response format for data collection in the social sciences including education, its continued use has raised some serious methodological questions. It was therefore against this background that the current study established the psychometric properties of the Likert scale in relation to an alternative response format called the expanded scale format. Using the instrumentation research design, the study compared the extended scale format and Likert scale format using the graded response model of the Item Response Theory (IRT). The study was guided by four research questions, while a sample of 2742 undergraduate students were used for the study. The sample was drawn using the convenience sampling technique. The instrument used for data collection was the Rosenberg Self-Esteem Scale, which was converted from the original Likert scale format to the expanded format. The instrument was assessed for validity and reliability and was shown to possess preliminary psychometric tools for further investigation. Data collection was done using an online tool, Question Pro® to administer the two version of the instrument to programmed respondents through social media platforms. Data analysis was done using exploratory factor analysis, confirmatory factor analysis, difficulty and discrimination parameters where applicable. The result showed that across all psychometric indicators, the expanded scale format was more suitable than the Likert version. On the basis of the result obtained, it was recommended that psychometricians develop future scales using more of the expanded format.

KEYWORDS: ***Grade response model, Item Response Theory, Likert scale, Expanded response scale***

INTRODUCTION

In the field of psychometrics, three broad approaches are used in the collection of data for analysis according to Leary (2012). These approaches are observational, physiological and self-reports. Observational measures involve the direct observation and recording of anything, activity or event a participant does that is of interest to a researcher (or someone assigned to observe such as peers, parent or teacher). This includes activities like pressing a bar (as in Pavlov classic experiment), aggression by children in the playfield, typing speed of a computer student, fidgeting by someone making a speech, or presentation skill of students defending their project, dissertation, or thesis. Physiological measures are used to gauge bodily (biological) processes and responses such as salivation of a dog (as in Thordike's seminal experiment), sweating, heart rate, brain activity, hormonal changes, sleep pattern etc.

Self-report measures involves the replies people provide to questionnaires and instruments regarding their thoughts, feelings and behavior. Leary (2012) further provided three classifications of self-report measures which are cognitive, affective and behavioral self-reports. Cognitive self-report measures what people think about something. For example, an educational psychologist may ask children which cup contains more content among three cups of different shapes (or sizes) with equal content. Affective self-report involves participants' responses regarding how they feel about themselves, others, events, or objects. Most psychological researches involving topics

EPRA International Journal of Environmental Economics, Commerce and Educational Management
Journal DOI: 10.36713/epra0414 |ISI I.F Value: 0.815|SJIF Impact Factor (2021): 7.743 ISSN: 2348 – 814X
Volume: 9 | Issue: 4 | April 2022

such as stress, anxiety, grief, satisfaction, depression, mental quality, school attitude, vocational interest etc. involve the use of affective self-report. Behavioral self-reports involves participants responding on their actions. Participants may be asked how often they listen to news, have sex, use condom, or react when threatened with violence.

The various approaches to measurement (observational, physiological, and self-reports) have their respective positives and pitfalls. This is because any particular measurement procedure provides only a rough (and imperfect) picture of any given construct. Therefore, researchers sometimes measures a psychological attribute by employing and integrating two or more of the approaches through the process called converging operations (Sternberg & Grigorenko, 2001). However, due to practical considerations such as time limit limitation, ethical bottlenecks and financial constraints, researchers are mostly restricted to the utilization of only one approach, with the easiest and most popular being self-report, also called rating scale. The ease of constructing self-report measures, or rating scales, has also lead to a corresponding burgeoning of instruments with various types and design. According to Taylor (2017), there are over 20 types of self-report scale format which can be broadly classified into graphic scale, likert scale and sematic differential scale. Of these self-report scale types, the likert scale is the most preferred and utilized.

Originally introduced by Rensis Likert (1932), likert scaling (LS) is the most widely used method of measuring personality, social and psychological attitudes and attriutes (Hodge & Gillespie, 2007). It has been used to measure various psychological constructs such as educational aspiration (Nwaialu & Iruloh, 2014), school dropout among Nigerian male (Agbakuru & Izuchi, 2009), marital adjustment (Paul-Cookey & Opara, 2015) etc. In the original scale format proposed by Likert (1932), the scale is composed of a stem – a positively or negatively stated proposition; and a response key which respondents are expected to provide their opinion based on the extent to which they approve or disapprove the proposition made in the stem. The popularity of LS can be attributed to a number of factors including ease of construction and application, intuitive appeal, adaptability to various setting and suitable reliability. In contemporary usage, respondents to a LS are required to indicate their level of agreement or disagreement to proposition on a graded scale (often but not necessarily four- or five-point scale). There appears no consensus on the maximum number of response categories to be used on a scale (Leung, 2011).

As previously stated most likert scale contains both positively and negatively keyed items. In fact most scholars (Orluwene, 2012; Iweka, 2014) strongly recommends that a scale should contain a balanced set of positive and negative keyed items. In this study, positively keyed items are items in which the stems or proposition are phrased in the direction of the construct or that in which agreement to the proposition indicates a greater presence of the construct being assessed or measured. On the other hand, negatively-keyed items in are those which the stems or item propositions are phrased in the opposite direction or in which agreement to the proposition indicates a lesser presence of the construct being measured. While positively keyed items are straight forward to construct, three types of negatively keyed LS items in English Language have been identified according to Salazar (2015). These are (i.) regular or direct negation; (ii.) polar opposites and; (iii.) negation of polar opposites. Using the assessment of happiness as an example, a positive keyed item may read " I am happy". For negative items it may read "I am not happy" (regular or direct negation), "I am unhappy" (polar opposite) or "I am not unhappy" (negation of polar opposites).

Apart from the commonsensical explanation for including negatively keyed items into scales in order two create balanced assessment tools, it has been advocated as a means for controlling or minimizing bias when people tend to agree or disagree to scale items without regard to their actual content, due to laziness, indifference, or response style. This is because negatively keyed items are believed to reduce response speed and increase cognitive reasoning needed to adequately respond to the items. Also, inclusion of negative items helps to cancel out bias as the response to these items are reversed scored during data analysis stage (Baumgartner & Steenkamp, 2001).

However, the logic and utility of negatively-keyed items have received strong empirical bashing recently. One problem associated with negatively keyed items is that they are not responded to on a consistent manner because acquiescence bias (the tendency to only agree to items) is an individual difference variable and not an item variable (Coleman, 2013). Furthermore, the result from analyses that includes negatively worded items have been shown to contaminate the factor structure of instrument when principals component analysis (PCA), confirmatory factor analysis (CFA) and structural equation modelling are the purpose of the investigation. Thirdly, the inclusion of negatively keyed items may cause confusion and carelessness in responding such as when respondents mistakenly read "I am happy" instead of "I am *un*happy". (Zhang & Savalei, 2015). These problems although viewed as inconsequential, have been shown to be inimical to the psychometric properties of instruments. For example empirical evidence reveal that inclusion of negative items result in low level of correlation between negatively-

EPRA International Journal of Environmental Economics, Commerce and Educational Management
Journal DOI: 10.36713/epra0414 |ISI I.F Value: 0.815|SJIF Impact Factor (2021): 7.743 ISSN: 2348 – 814X
Volume: 9 | Issue: 4 | April 2022

keyed items and total score, even after reverse scoring. They also resulted in loss of reliability in the scales and the introduction of method effects in scales (Ye & Wallace, 2014). Also, DiStefano & Molt (2006), have shown that the introduction of negatively keyed items have resulted in the emergence of multiple components or dimension for a scale that is supposed to measure only one underlying construct.

With the identified problems associated with negatively keyed items, some psychometricians have asseverated that eliminating negatively worded items or replacing them with only positive items will result in greater validity and reliability (Roszkowsky & Soven, 2010). However, this suggestion is wanting as doing so will further increase acquiescence bias another problem of positive-only keyed items. To balance the demand for controlling acquiescence bias and addressing the plethora of problems associated with negative keyed items, alternative scale formats have been suggested by scholars including forced choice response format (Brown & Maydeu-Olivares, 2011), phrase completion response format (Hodge & Gillespie, 2007) and the expanded response format (Zhang & Savalei, 2015).

The expanded format of scale has a major advantage over the forced choice format in that in the forced choice format, item response are only available in binary response (true of me/not true of me, yes/no), thereby making it difficult to compare it to the likert format which usually uses three, four, five or more responses. The expanded format is a relatively recent and less utilized scaling format in which full sentences are used to reflect each response category. To the best of this researcher's knowledge only one instrument have used this scaling format which is the Beck Depression Inventory (BDI, Beck, Ward, Mendelson, Mock & Grossland, 1989). For instance, item 1 of the BDI which measures sadness was stated as follows:

- I do not feel sad.
- I feel sad much of the time.
- I am sad all the time.
- I am so sad or unhappy that I can't stand it.

From the sample item shown above, it can be seen that for each item, a response statement similar to both positive and negative item is presented. It therefore can be seen that each item has been posited in such a manner that has the possibility of eliminating acquiescence bias associated with positively keyed items and reducing methodological effects arising from the use of negative keyed items. This is because expended format forces respondents to pay increased attention to item content and notice subtle difference between options. Also the expanded format has the possibility of including the three forms of negative keyed response format into one items in a manner that provides grading, intensity or frequency. Before opting for this format by researchers and measurement specialist, it is imperative that evidence from empirical studies be provided showing the robustness of the format over the fairly established and popular liker format. However, research comparing these two formats is virtually non-existent with that of Zhang and Savelei (2016) being the only exception.

In the present study, the effectiveness of the expanded format will be compared with the likert format to ascertain the item, scale and person properties of a standardized instrument using the Graded Response Model (GRM). Graded Response Model is one of the polytomous models of Item Response Theory (IRT). Graded Response Model was developed by Samejima (1969) and is used an extension of the 2-PLM of dichotomously scored items to an polytomous item response constructed in an ordered category such as in most self-reports with grading of 1, 2, 3, 4and 5 as used in likert scales or other rating scale formats with more than two response alternatives. In conducting this investigation, GRM will be applied on a popular scale in two alternative formats, specifically the expanded format and likert format through a web-survey collection format. This scale is the Rosenberg Self-Esteem Scale (RSE, Rosenburg 1965).

STATEMENT OF THE PROBLEM

While likert scale formats have a fairly established history and great application due to their ease of construction, scoring and usage, the wordings of items and the introduction of negative keyed items may introduce method effects and response biase which often contaminate the final score obtained. Furthermore, considering the fact that data collected from instruments using likert scales are not end in themselves, but are used for further statistical analysis such as F-test, t-test, Cronbach Alpha analysis, correlational and regression analysis, factor analysis and Structural Equation Modelling etc, it becomes imperative that items making up the scale are carefully worded by test developers, appropriately understood and responded to by respondents and suitably analyzed and used by researchers. However, negatively keyed items may confuse the respondents and results in careless

EPRA International Journal of Environmental Economics, Commerce and Educational Management
Journal DOI: 10.36713/epra0414 |ISI I.F Value: 0.815|SJIF Impact Factor (2021): 7.743 ISSN: 2348 – 814X
Volume: 9 | Issue: 4 | April 2022

responding, create measurement errors or even introduce additional factors, dimension or components into scales structures not originally modeled by scale developers. These concerns which are problematic at the theoretical level portends greater danger at the practical stages of research. This issues may results in further problems like wrong diagnosis of psychological problems as most clinical assessment utilize likert response scale. Also in the educational and profession world, the assessment of non-cognitive competencies or skills such as emotional intelligence, motivation, interest, learning styles, learning disability and grit is dominated by the use of likert scale instruments. With poorly constructed instruments, decisions on students and employees placement are prone to being flawed because they may not fit in with the result obtained from the instruments.

It is therefore pertinent that the wisdom of employing negative-worded items be questioned, as well as provide evidence to ascertain the feasibility of an alternative response formats that is more efficacious in providing item, scale and person characteristics that are valid. Reliable and usable in an indigenous basic and applied research setting. Ascertaining these very vital characteristics are not dependent or feasible with CTT as item characteristics such as difficulty and discrimination are based on the sample data is obtained from, with little or no concern for item statements or response category. Unlike CTT, IRT provides the possibility of ascertaining item, scale and person parameters. More specifically, GRM is robust enough to permit the modelling of traits and abilities along an ordered response scaling format. The problem of this study therefore is to conduct a psychometric investigation of the item, scale and person parameters on both the likert and expanded format through the application of GRM.

RESEARCH QUESTIONS

The understated questions were answered to guide the present study:

1. What is the dimension of both the expanded and likert response scale of RSES?
2. To what extent are the expanded and likert response scale formats maintain local independence?
3. What is the spread of categorical difficulties for the expanded and likert response scale formats of the RSES using GRM?
4. What is the discrimination of each item of the RSES for the likert and expanded response scale formats?

METHODOLOGY

The instrumentation research design was used for this study. Instrumentation research design according to Kpolovie (2010) deals with ascertaining the psychometric properties of a psychological scale or tool. While this includes the generation of items, it is also concerned with ascertain the extent to which the items meets specified criteria on the basis of specific theories. The population for the study includes all undergraduate students in the University of Port Harcourt during the 2017/2018 academic session. From the population, a sample of 2742 students were drawn using convenience sampling technique. The instrument for data collection was the Rosenberg (1965) Self-Esteem Scale (RSES).

The instrument was developed by Rosenburg (1965) and the original version of the instrument contained five positively worded item and five negatively worded items all measured on a four-point scale, where 4 corresponds to *Strongly agree* and 1 corresponds to *Strongly disagree*. In the corresponding 10-item Expanded version of RSES, each item consists of four sentences to choose from, and all items are always arranged from the highest to the lowest self-esteem. For each item, participants were asked to select one of the four options that best describes them. The text of the corresponding Likert version item was used to create the four options for each item in the Expanded format. For most items in the Expanded version, the original Likert item was included as one of the four options; for the rest of the items, the options were created by adding a modifier to the original Likert item. The face and content validities of the instrument was established, for both the Likert and extended versions of the instrument. Reliability of the instrument was done using the Cronbach Alpha method of internal consistency. Result of the analysis showed that the expanded scale format had an internal reliability coefficient of 0.83, while the likert version had an internal consistency of 0.73. The values showed high level of reliability according to the Ukwuijie (2009).

Data collection was done using an online tool called Question Pro®. This program ensured that when the link to the instrument is place on students' social media platform, a student can only access the instrument once and can only have access to one version of the instrument (either likert or expanded version). The software was synchronized with Google Sheet where the data was collected and exported to appropriate statistical tool either SPSS® or STATA®. For answering research question, the Exploratory Factor Analysis (EFA) was used to identify the factor structure of the data. Research question two was answered using Confirmatory Factor Analysis (CFA). Research question three was answered by determining the difficulty parameter of the items, while the last research question

EPRA International Journal of Environmental Economics, Commerce and Educational Management
Journal DOI: 10.36713/epra0414 |ISI I.F Value: 0.815|SJIF Impact Factor (2021): 7.743 ISSN: 2348 – 814X
Volume: 9 | Issue: 4 | April 2022

was answered by determining the discrimination parameter of the items. The first research question was answered using the SPSS® version 21 software, while the other research questions were answered using Stata® software.

RESULT PRESENTATION

Research Question One: What is the dimension of both the expanded and likert response scale of RSES?
One of the basic assumption of IRT is unidimensionality. Unidimensionality of the two scale formats were assessed using the SPSS statistics software version 18 through employing the Exploratory Factor Analysis. The result of the analysis is shown in figures 1 and 2

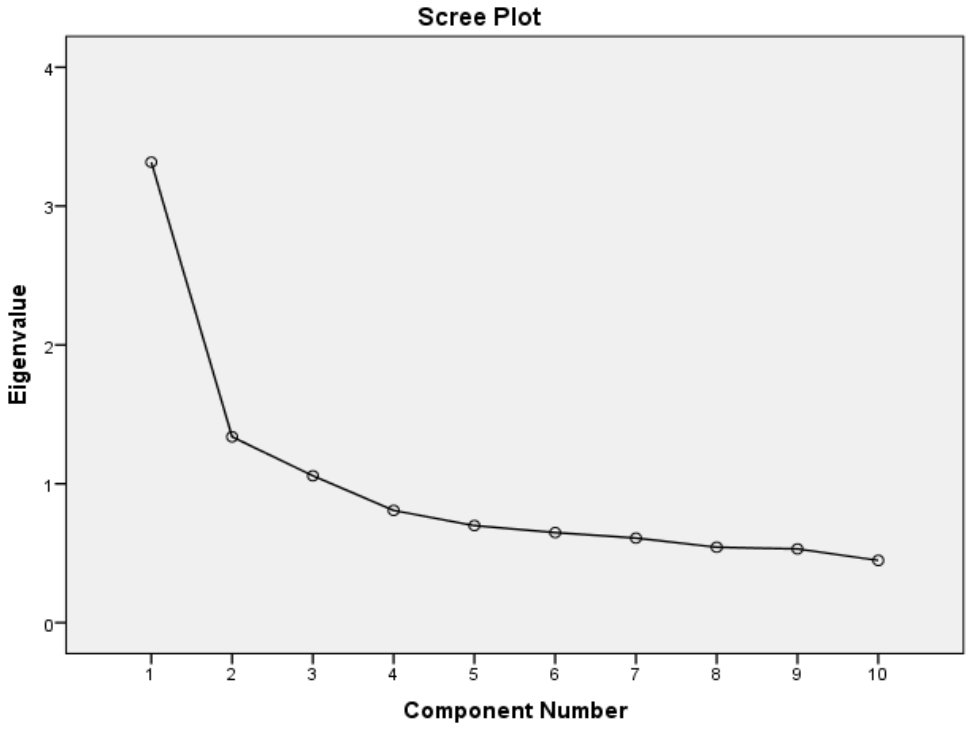


Fig 1: Scree plot for the EFA of the Likert Format

EPRA International Journal of Environmental Economics, Commerce and Educational Management
Journal DOI: 10.36713/epra0414 |ISI I.F Value: 0.815|SJIF Impact Factor (2021): 7.743 ISSN: 2348 – 814X
Volume: 9 | Issue: 4 | April 2022

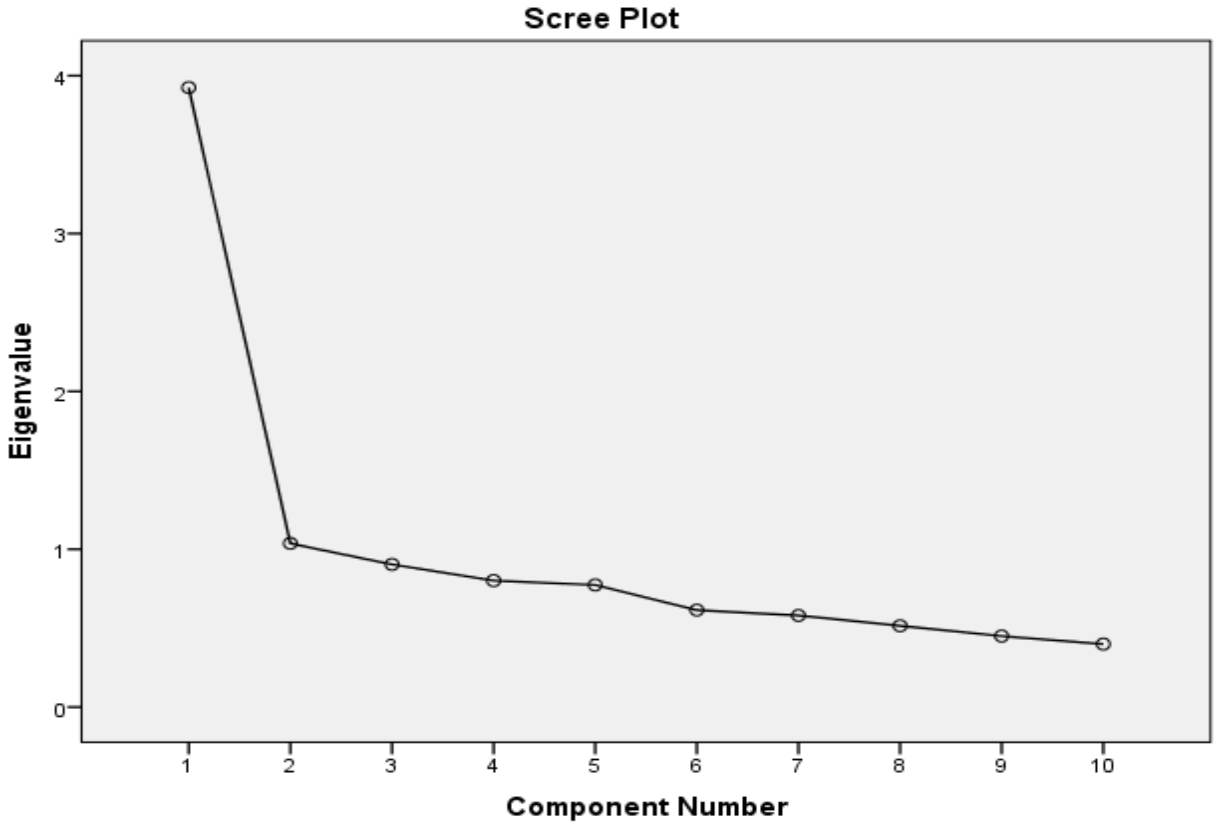


Fig 2: Scree plot for the EFA of the Likert Format

From the result displayed pictorially in Fig 1 and Fig 2, it can be seen that that there was a drop in the first and second eigen values followed by a leveling off of the subsequent eigen values. This suggests that there is only one dominant component in the data collected. In this case, self-esteem. Furthermore, the result showed that there was a smoother levelling in fig 2, than in figure 1. Hence it can be observed that the assumption of unidimensionality was not violated, as such the data obtained was measuring only one construct, and nothing else. However, the result from the EFA showed that for the expanded scale format, the first component accounted for about 39.28% of the total variance, while for the likert scale format the first component accounted for 33.16% variation in the components of the scale for the Rosenberg Self-Esteem Scale.

Research Question Two: To what extent are the expanded and likert response scale formats maintain local independence?

To assess local independence which is one of the assumption for the conduct of IRT, including GRM a one factor CFA solution was conducted using Stata 14 package for both formats of the instrument, results of which are shown in table 1 and 2 below

EPRA International Journal of Environmental Economics, Commerce and Educational Management
Journal DOI: 10.36713/epra0414 |ISI I.F Value: 0.815|SJIF Impact Factor (2021): 7.743 ISSN: 2348 – 814X
Volume: 9 | Issue: 4 | April 2022

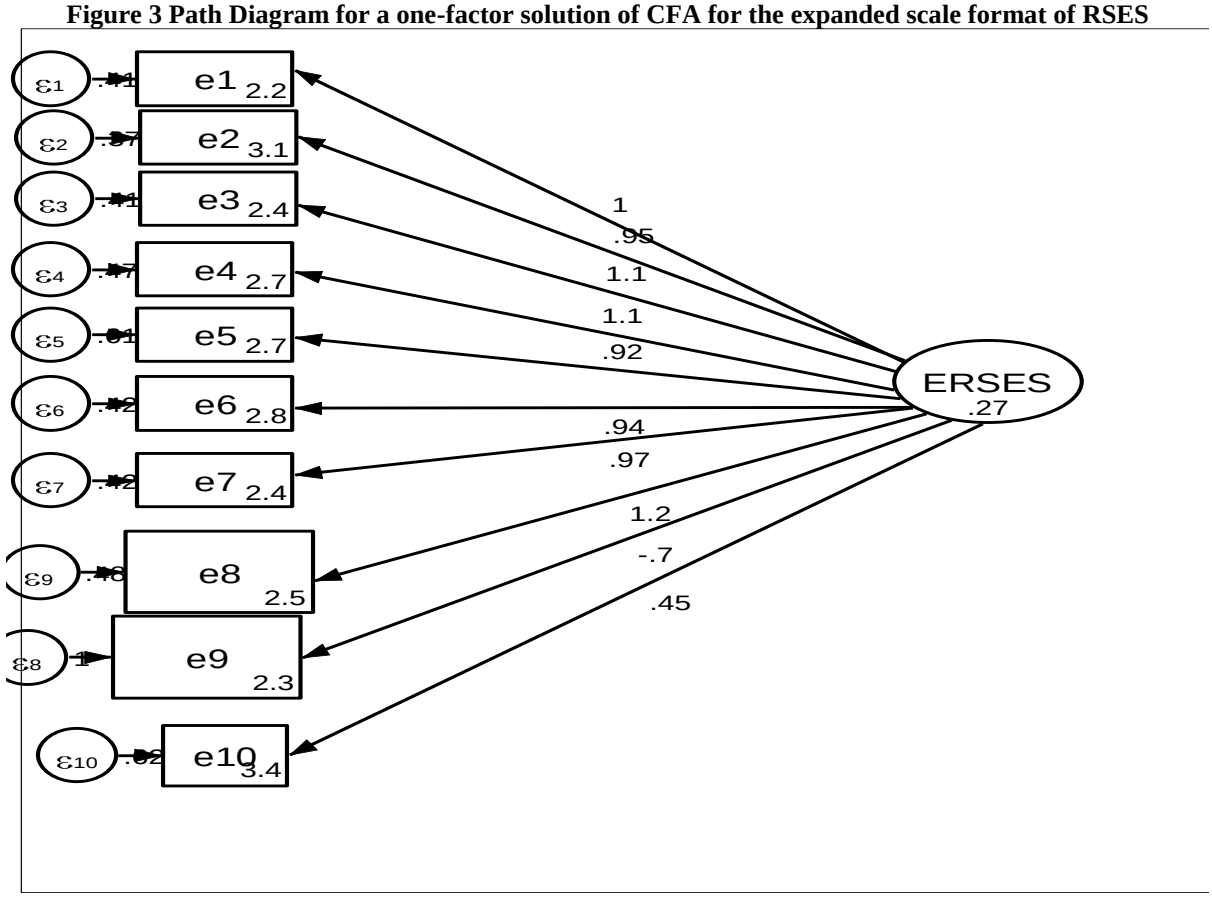
Table 1 Model Fit statistics of the one factor solution CFA of the Expanded format

Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(35)	383.060	model vs. saturated
p > chi2	0.000	
chi2_bs(45)	3546.207	baseline vs. saturated
p > chi2	0.000	
Population error		
RMSEA	0.085	Root mean squared error of approximation
90% CI, lower bound	0.078	
upper bound	0.093	
pclose	0.000	Probability RMSEA <= 0.05
Information criteria		
AIC	31923.217	Akaike's information criterion
BIC	32079.916	Bayesian information criterion
Baseline comparison		
CFI	0.901	Comparative fit index
TLI	0.872	Tucker-Lewis index
Size of residuals		
SRMR	0.043	Standardized root mean squared residual
CD	0.843	Coefficient of determination

According to the information shown in table 1 above on the confirmatory factor analysis of the expanded format scale, mixed result was obtained indicating the goodness of fit of the format. The chi-square values yielded a coefficient that was statistically significant (383.060, p <0.000). However the Root Mean Squared error of approximation (RMSEA) showed a value that with a failing good fit of 0.085, while the comparative fit index showed that a value of 0.901 was gotten which implies that suitable level of fit. This result indicates that the data has a fairly reasonable level of fit and there are no excess covariation in the residual matrix which indicated that the items for the expanded format are not correlated when the construct is held constant. Hence for the expanded format, the assumption of local independence was not violated. The one-factor CFA path diagram for the expanded format is displayed in figure 4.3 below

EPRA International Journal of Environmental Economics, Commerce and Educational Management
Journal DOI: 10.36713/epra0414 |ISI I.F Value: 0.815|SJIF Impact Factor (2021): 7.743 ISSN: 2348 – 814X
Volume: 9 | Issue: 4 | April 2022

Figure 3 Path Diagram for a one-factor solution of CFA for the expanded scale format of RSES



EPRA International Journal of Environmental Economics, Commerce and Educational Management
Journal DOI: 10.36713/epra0414 |ISI I.F Value: 0.815|SJIF Impact Factor (2021): 7.743 ISSN: 2348 – 814X
Volume: 9 | Issue: 4 | April 2022

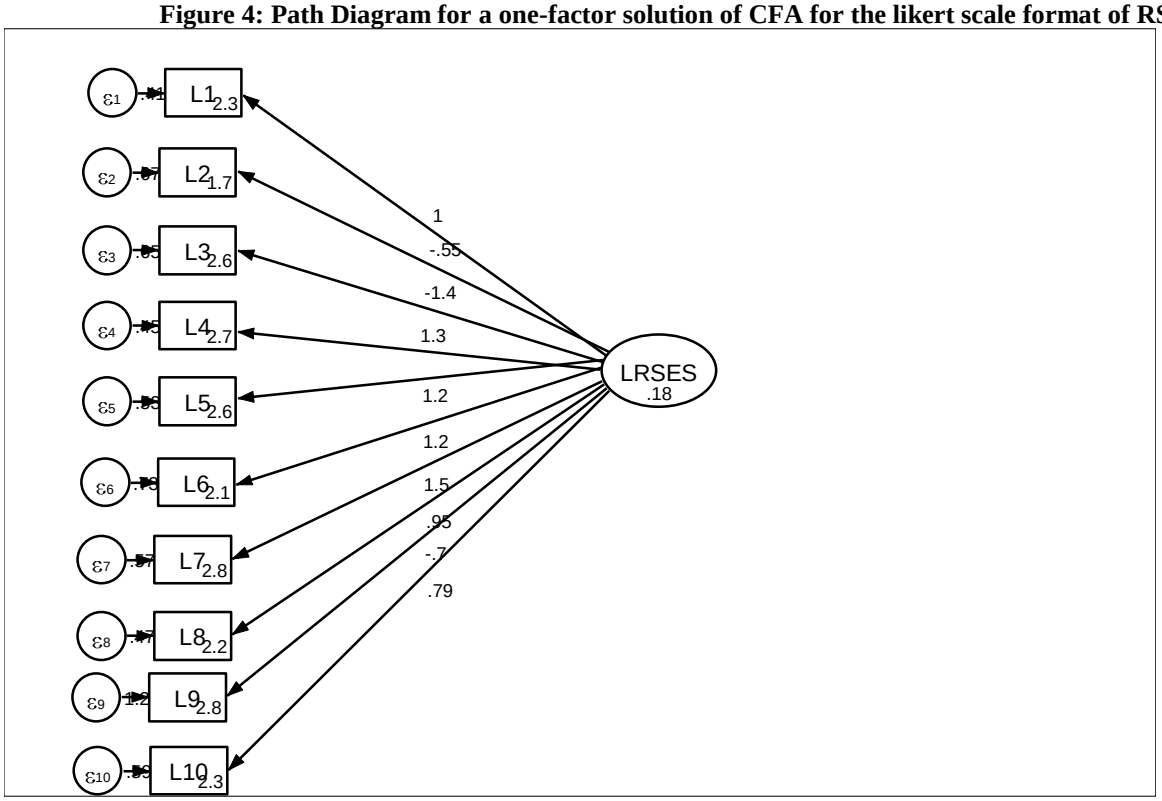
Table 2 Model Fit statistics of the one factor solution CFA of the Expanded format

Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(35)	566.879	model vs. saturated
p > chi2	0.000	
chi2_bs(45)	2740.460	baseline vs. saturated
p > chi2	0.000	
Population error		
RMSEA	0.105	Root mean squared error of approximation
90% CI, lower bound	0.098	
upper bound	0.113	
pclose	0.000	Probability RMSEA <= 0.05
Information criteria		
AIC	34187.213	Akaike's information criterion
BIC	34343.912	Bayesian information criterion
Baseline comparison		
CFI	0.803	Comparative fit index
TLI	0.746	Tucker-Lewis index
Size of residuals		
SRMR	0.065	Standardized root mean squared residual
CD	0.796	Coefficient of determination

From the data displayed in table 3, it can be seen that the chi-square value obtained was 566.879 ($p < 0.05$) which suggest a poor fit of the data. Furthermore, the Root Mean Square Error of Approximation obtained was greater than the acceptable standard of 0.05 – 0.08, indicating that the data was not properly fit for the analysis suggesting that the data obtained from the likert format had excess covariation in the residual matrix, thus indicating that the principal of local independence was violated. However from the recommendation of Chen, Hwang and Lin (2013) that if in an EFA, the first factor accounted for above 30% of the total variance, such is an indication of local independence. On the basis of this information, further analysis was conducted. The one factor CFA path diagram for the likert scale format is displayed in table 4

EPRA International Journal of Environmental Economics, Commerce and Educational Management
Journal DOI: 10.36713/epra0414 |ISI I.F Value: 0.815|SJIF Impact Factor (2021): 7.743 ISSN: 2348 – 814X
Volume: 9 | Issue: 4 | April 2022

Figure 4: Path Diagram for a one-factor solution of CFA for the likert scale format of RSES



Research Question Three: What is the spread of categorical difficulties for the expanded and likert response scale formats of the RSES using GRM?

To answer research question one, the scores of respondents on both scale formats were fitted into a graded response format, to identify their difficulty parameters. The values for both test formats are shown in table 3 and 4 below

EPRA International Journal of Environmental Economics, Commerce and Educational Management
Journal DOI: 10.36713/epra0414 |ISI I.F Value: 0.815|SJIF Impact Factor (2021): 7.743 ISSN: 2348 – 814X
Volume: 9 | Issue: 4 | April 2022

Table 3: Item and Calibrated Item Parameters of the Expanded RSES

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
e1						
Discrim	1.696145	.0976817	17.36	0.000	1.504693	1.887598
Diff						
>=2	-1.342112	.0732406	-18.32	0.000	-1.485661	-1.198563
>=3	.7648638	.0559767	13.66	0.000	.6551515	.8745761
=4	1.979438	.0995758	19.88	0.000	1.784273	2.174603
e2						
Discrim	1.707702	.0991355	17.23	0.000	1.5134	1.902003
Diff						
>=2	-2.932601	.1584315	-18.51	0.000	-3.243121	-2.622081
>=3	-1.025452	.0629297	-16.30	0.000	-1.148792	-.9021125
=4	.6968304	.0552176	12.62	0.000	.5886059	.805055
e3						
Discrim	1.931532	.1072248	18.01	0.000	1.721375	2.141689
Diff						
>=2	-1.521965	.0739401	-20.58	0.000	-1.666885	-1.377045
>=3	.3523566	.0461482	7.64	0.000	.2619077	.4428055
=4	1.509868	.0744825	20.27	0.000	1.363885	1.655851
e4						
Discrim	1.622648	.0922517	17.59	0.000	1.441838	1.803458
Diff						
>=2	-2.222485	.1127649	-19.71	0.000	-2.443501	-2.00147
>=3	-.303092	.0496823	-6.10	0.000	-.4004675	-.2057164
=4	1.06803	.0663792	16.09	0.000	.9379296	1.198131
e5						
Discrim	1.209275	.0764332	15.82	0.000	1.059469	1.359082
Diff						
>=2	-2.531094	.1514382	-16.71	0.000	-2.827907	-2.23428
>=3	-.1048424	.0569261	-1.84	0.066	-.2164154	.0067307
=4	1.22074	.0861794	14.17	0.000	1.051831	1.389648

EPRA International Journal of Environmental Economics, Commerce and Educational Management
Journal DOI: 10.36713/epra0414 |ISI I.F Value: 0.815|SJIF Impact Factor (2021): 7.743 ISSN: 2348 – 814X
Volume: 9 | Issue: 4 | April 2022

Table 3: Item and Calibrated Item Parameters of the Expanded RSES Cont'd

e5						
Discrim	1.209275	.0764332	15.82	0.000	1.059469	1.359082
Diff						
>=2	-2.531094	.1514382	-16.71	0.000	-2.827907	-2.23428
>=3	-.1048424	.0569261	-1.84	0.066	-.2164154	.0067307
=4	1.22074	.0861794	14.17	0.000	1.051831	1.389648
e6						
Discrim	1.532543	.0894807	17.13	0.000	1.357164	1.707922
Diff						
>=2	-2.785333	.151815	-18.35	0.000	-3.082885	-2.487781
>=3	-.4928134	.0536102	-9.19	0.000	-.5978875	-.3877393
=4	1.173688	.0721495	16.27	0.000	1.032278	1.315099
e7						
Discrim	1.603023	.093556	17.13	0.000	1.419656	1.786389
Diff						
>=2	-1.966619	.100906	-19.49	0.000	-2.164391	-1.768847
>=3	.4062883	.051117	7.95	0.000	.3061007	.5064758
=4	1.694463	.0901059	18.81	0.000	1.517858	1.871067
e8						
Discrim	1.810111	.0999846	18.10	0.000	1.614144	2.006077
Diff						
>=2	-1.445364	.0739404	-19.55	0.000	-1.590285	-1.300444
>=3	.0470881	.0456019	1.03	0.302	-.04229	.1364661
=4	1.357044	.0712473	19.05	0.000	1.217402	1.496686
e9						
Discrim	-.664392	.0609336	-10.90	0.000	-.7838198	-.5449643
Diff						
>=2	1.51167	.1558394	9.70	0.000	1.206231	1.81711
>=3	-.7209082	.1077931	-6.69	0.000	-.9321787	-.5096377
=4	-2.45276	.2296522	-10.68	0.000	-2.90287	-2.00265
e10						
Discrim	.6540523	.0660699	9.90	0.000	.5245577	.7835469
Diff						
>=2	-6.071494	.6303178	-9.63	0.000	-7.306895	-4.836094
>=3	-2.606731	.2571751	-10.14	0.000	-3.110785	-2.102677
=4	-.4712315	.0977586	-4.82	0.000	-.6628347	-.2796282

EPRA International Journal of Environmental Economics, Commerce and Educational Management
Journal DOI: 10.36713/epra0414 |ISI I.F Value: 0.815|SJIF Impact Factor (2021): 7.743 ISSN: 2348 – 814X
Volume: 9 | Issue: 4 | April 2022

Table 4: Item and Calibrated Item Parameters of the Likert RSES

Graded response model Number of obs = 1,371
Log likelihood = -15631.341

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
L1						
Discrim	1.342907	.0880816	15.25	0.000	1.17027	1.515543
Diff						
>=2	-2.013443	.1177784	-17.10	0.000	-2.244284	-1.782602
>=3	.5225628	.0591933	8.83	0.000	.4065461	.6385795
=4	2.341203	.136124	17.20	0.000	2.074405	2.608001
L2						
Discrim	-.5660721	.0671288	-8.43	0.000	-.6976422	-.4345021
Diff						
>=2	-.3184591	.1073461	-2.97	0.003	-.5288535	-.1080646
>=3	-3.283101	.3822026	-8.59	0.000	-4.032204	-2.533997
=4	-5.34561	.627856	-8.51	0.000	-6.576185	-4.115034
L3						
Discrim	-1.449851	.090354	-16.05	0.000	-1.626942	-1.27276
Diff						
>=2	1.45157	.0853327	17.01	0.000	1.284321	1.618819
>=3	.3064976	.0528775	5.80	0.000	.2028595	.4101357
=4	-1.186212	.0755853	-15.69	0.000	-1.334356	-1.038068
L4						
Discrim	1.646577	.1008742	16.32	0.000	1.448867	1.844287
Diff						
>=2	-2.215138	.1156035	-19.16	0.000	-2.441717	-1.988559
>=3	-.2190517	.0487508	-4.49	0.000	-.3146015	-.1235019
=4	1.152562	.0692151	16.65	0.000	1.016903	1.288221
L5						
Discrim	1.38341	.0882751	15.67	0.000	1.210394	1.556426
Diff						
>=2	-2.438179	.1391595	-17.52	0.000	-2.710926	-2.165431
>=3	.0484458	.0525715	0.92	0.357	-.0545925	.1514841
=4	1.28083	.081927	15.63	0.000	1.120256	1.441404

EPRA International Journal of Environmental Economics, Commerce and Educational Management
Journal DOI: 10.36713/epra0414 |ISI I.F Value: 0.815|SJIF Impact Factor (2021): 7.743 ISSN: 2348 – 814X
Volume: 9 | Issue: 4 | April 2022

Table 4: Item and Calibrated Item Parameters of the Likert RSES Cont'd

L6						
Discrim	1.191071	.08251	14.44	0.000	1.029354	1.352787
Diff						
>=2	-1.015238	.0801928	-12.66	0.000	-1.172413	-.8580626
>=3	1.00858	.0794619	12.69	0.000	.8528377	1.164323
=4	1.798266	.1172227	15.34	0.000	1.568514	2.028018
L7						
Discrim	1.661426	.105317	15.78	0.000	1.455009	1.867844
Diff						
>=2	-1.97885	.1041466	-19.00	0.000	-2.182973	-1.774726
>=3	-.2505637	.0487361	-5.14	0.000	-.3460847	-.1550428
=4	.7036262	.057268	12.29	0.000	.5913831	.8158693
L8						
Discrim	1.22465	.0825237	14.84	0.000	1.062906	1.386393
Diff						
>=2	-1.714922	.1079885	-15.88	0.000	-1.926575	-1.503268
>=3	1.046444	.0788173	13.28	0.000	.8919653	1.200923
=4	2.405618	.1487928	16.17	0.000	2.113989	2.697247
L9						
Discrim	-.4760639	.0602504	-7.90	0.000	-.5941525	-.3579753
Diff						
>=2	3.129974	.3994084	7.84	0.000	2.347148	3.9128
>=3	1.047768	.1743734	6.01	0.000	.7060025	1.389534
=4	-.9889997	.1683229	-5.88	0.000	-1.318907	-.6590929
L10						
Discrim	.8744695	.0709798	12.32	0.000	.7353516	1.013587
Diff						
>=2	-2.377421	.18405	-12.92	0.000	-2.738152	-2.01669
>=3	1.087215	.1050458	10.35	0.000	.881329	1.293101
=4	2.629855	.2051809	12.82	0.000	2.227708	3.032002

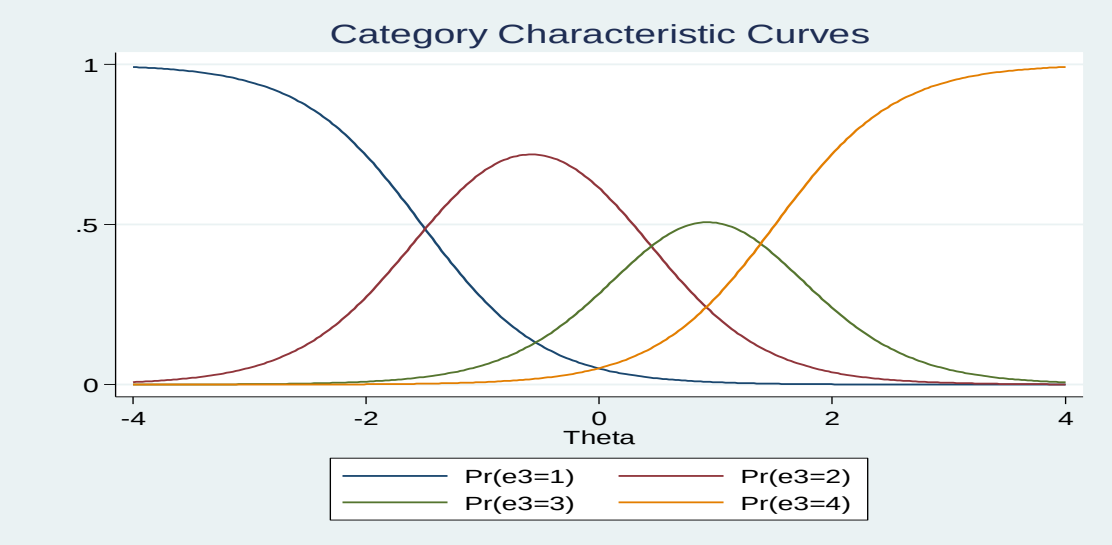
.

From the calibrated items displayed in table 3 and 4 respectively, the category threshold each display the point at which 50% of respondents would endorse the designated option or higher in the RSES for each scale format. Every respondent has a 100% probability of endorsing the first option so there is no threshold for that option. The metric established for the RSES is set to a mean of 0 and standard deviation of 1. As seen from table for the expanded format scale, the difficulty parameter of the item of the first category is was -1.34, while that of the likert format is -2.01, for the second category the difficulty parameter was 0.76 for the expanded format while for the likert format it

EPRA International Journal of Environmental Economics, Commerce and Educational Management
Journal DOI: 10.36713/epra0414 |ISI I.F Value: 0.815|SJIF Impact Factor (2021): 7.743 ISSN: 2348 – 814X
Volume: 9 | Issue: 4 | April 2022

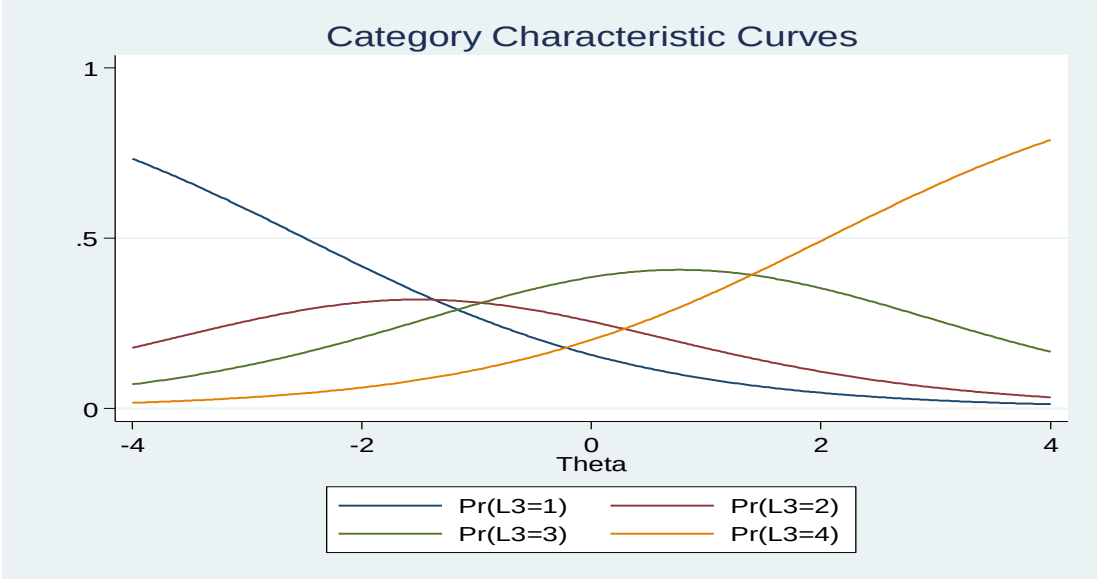
was 0.52. Finally, the data showed that for the fourth category the difficulty parameter was 1.98 for the expanded format while the likert format yielded a difficulty parameter of 2.34. On the whole the result showed that the threshold of the likert scale ranged from -5.34 to 3.12. Conversely, the result showed that the expanded format had difficulty parameters that ranged from -6.07 to 1.97. This result suggest that the expanded format has a broader range of item difficulty from the calibrated items than the likert scale. Further indication of this is shown in the category response curve in figure 5 and 6 for item 3 of both the expanded and likert scale formats.

Fig 5: Category response curve for item 3 in expanded format



a = 1.93, -1.52, b1 = 0.35, b2 = 0.35, b3 = 1.51

Figure 6: Category response curve for item 3 in likert scale



a = -1.45, b1 = 1.45, b2 = 0.30, b3 = -1.18

From the values in figures 5 and 6, it can be seen that the expanded format has a narrower slope curve than the likert scale format. Also the item threshold for the expanded scale format ranged from -1.5 to 1.51 which indicates 3.01 standard deviation apart which account for a narrower curve, than that of the Likert scale format with a range of -

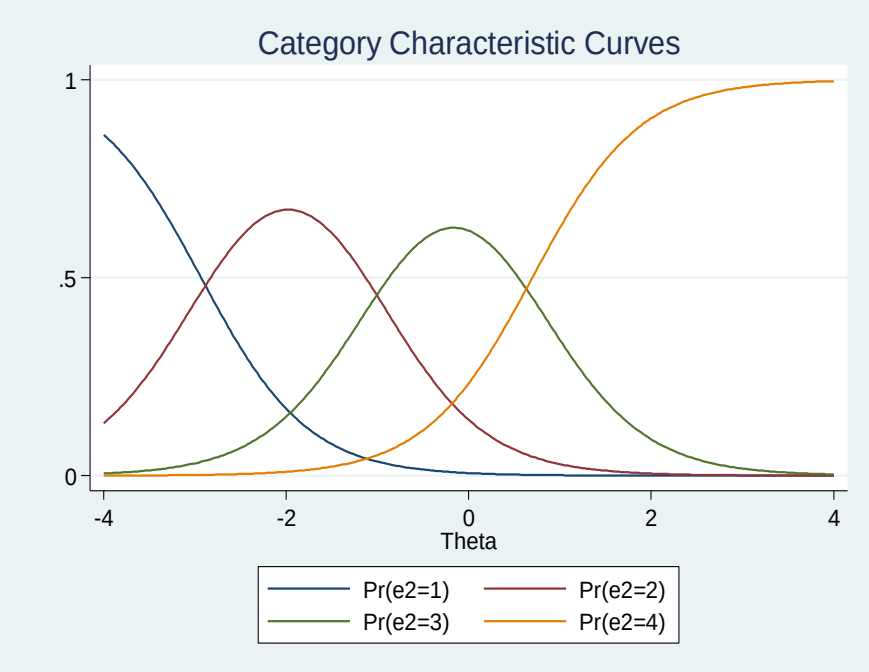
EPRA International Journal of Environmental Economics, Commerce and Educational Management
Journal DOI: 10.36713/epra0414 |ISI I.F Value: 0.815|SJIF Impact Factor (2021): 7.743 ISSN: 2348 – 814X
Volume: 9 | Issue: 4 | April 2022

1.18 to 1.45 with 2.8 standard deviations apart. From appendix (insert number here) , it can be seen that items in the expanded format had narrower slopes than that of the likert scale format, indicating that the expanded format is a better response format than the likert scale format.

Research Question Four: What is the discrimination of each item of the RSES for the likert and expanded response scale formats?

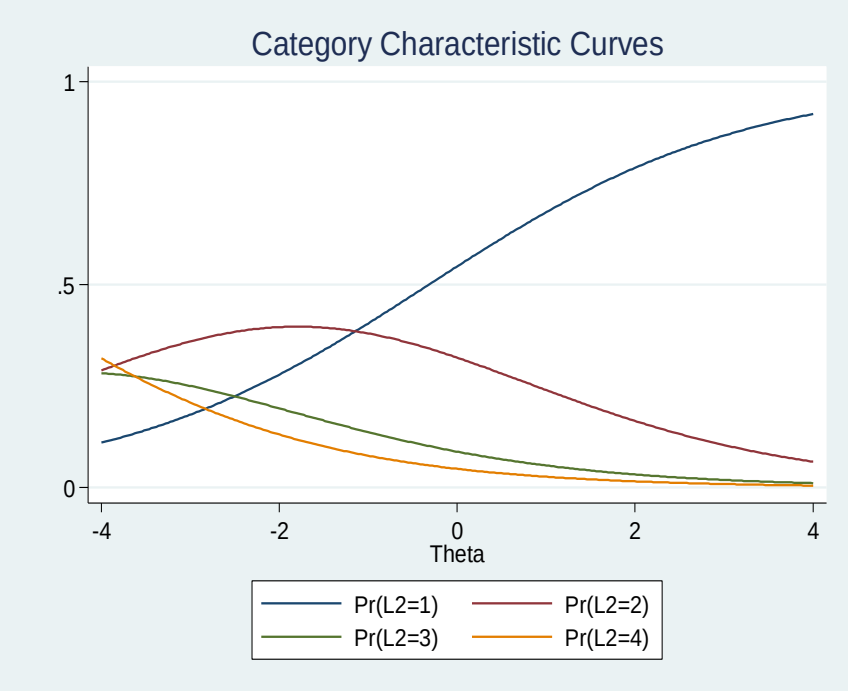
The discrimination parameter **b,** refers to the probability of a respondent endorsing a higher level of the item response on the basis of the level of their self-esteem. It refers to the strength of association between the item and the construct of self-esteem. From the value displayed in table 4.2 and 4.3, it can be seen that item discrimination of the expanded format ranged from 1.69 to 0.65, while that of the likert scale format ranged from -1.44 to 1.66. In more specific terms, the discrimination parameters of the expanded and likert scale formats 1.69 and 1.34 for item 1, 1.70 and -0.56 for item 2, 1.93 and -1.45 for item 3, 1.62 and 1.64 for item 4, 1.20 and 1.38 for item 5, 1.53 and 1.19 for item 6, 1.60 and 1.19 for item 7, 1.81 and 1.22 for item 8, -0.66 and – 0.48 for item 9, and 0.65 and 0.67 for item 10. This result indicates that with the exception of items 8, 9 and 10, the items in the expanded format displayed a greater level of discrimination than those of likert scale formats. This is shown from the pictorial representation shown in Figures 7 and 8 respectively.

Figure 7: Discrimination parameter curve of item 2 for expanded scale format



EPRA International Journal of Environmental Economics, Commerce and Educational Management
Journal DOI: 10.36713/epra0414 |ISI I.F Value: 0.815|SJIF Impact Factor (2021): 7.743 ISSN: 2348 – 814X
Volume: 9 | Issue: 4 | April 2022

Figure 8: Discrimination parameter curve of item 3 for likert scale format



DISCUSSION OF FINDINGS

From research question one that seeks to investigate the unidimensionality of the both the likert and expanded format options, it was shown that both formats had suitable level of unidimensionality. However, from the analysis of the factor structure of both scale formats, the result showed that the expanded scale format accounted for a higher level of variance that the Likert format. The scree plot from the EFA shows that the data was unidimensional in agreement with Embretson and Reises (2010) and Edelen and Reeve (2007) who agreed that if the first factor of an eigenvalue from an explanatory factor analysis is considerably larger than the orders then it is an indication of unidimensionalty. The result that the first component of the expanded format accounted for a greater percentage of the variance than the other components indicates that the expanded format has a greater level of precision in measuring the construct of self-esteem than the likert scale format. The result from this study is not surprising but expected because as Brown and Maydey-Olivares (2011) argued, in most negatively-worded items there is the tendency for it to create confusion and result in method effects among for respondents.

For research question two, the Confirmatory Factor Analysis showed that showed that the expanded scale format yielded a greater model fit than the expanded scale format. The result showed that from the Akaike Information Criterion and the comparative fit index, the expanded scale format yielded a better model fit than the likert scale. This result further confirmed that the expanded scale format had a better model fit statistics than the likert scale format. The result is not surprising to this researcher because as has been shown by Sonderen et al (2013), the possibility of reverse worded items in likert scale causes confusion and leads to acquiescence bias which causes covariance structure of the data, one of the limiting factors of likert scales.

Answer to research question three indicates that from the calibrated items, the difficulty parameter of the expanded scale formats showed a narrower slope curve than the likert scale. Across the four response options, the result showed that the expanded scale format yielded a better difficulty parameter than the likert scale format. This result is not surprising because the options in the expanded formats were more direct and did not consider the negative formats of the wordings.

Also the discrimination parameter of the study showed that the result from the expanded format showed greater discrimination than the likert scale format. Across all the items, with the exception of the last three items, the result showed the expanded scale format had a higher level of discrimination than the likert scale formats. This

result is not surprising but expected because as Hills and Argyle (2002) showed the fact that respondents clearly understand and can readily respond to the response option in the scale.

CONCLUSION

The conclusion drawn from the analysis of data showed that the expanded scale response format has better psychometric properties than the likert scale format.

Recommendation

On the basis of the result obtained, the following recommendations were made

1. Psychometricians should learn the rudiments of Item Response Theory, especially those of the polytomous response model and use the principles espoused in the model to construct scales.
2. Test experts should endeavor to develop items that adopt the expanded scale formats, especially for non-cognitive test items.
3. The result that Likert scale formats had lower level of unidimensionality and local independence than expanded format showed that the negative worded items may have introduced method effect into the data structure, as such they should not be used in scale construction.

REFERENCES

1. *Agbakwuru, C. & Izuchi, R. N. (2009). Psychological causes of school drop-out among Nigerian males: a wakeup call to teachers and guidance counsellors. Trends in Educational Studies, 4,(1), 57-61.*
2. *Baumgartner, H. & Steenkamp, J. B. E. M. (2001). Response styles in marketing research: A cross-national investigation. Journal of Marketing Research: 38, 2, 143-156.*
3. *Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. Archives of General Psychiatry, 4, 561-571.*
4. *Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaire. Educational and Psychological Measurement, 71, 460-502*
5. *Chen, S.-K., Yeh, Y.-C., Hwang, F.-M., & Lin, S. S. J. (2013). The relationship between academic self-concept and achievement: A multicohort–multioccasion study. Learning and Individual Differences, 23, 172–178. https://doi.org/10.1016/j.lindif.2012.07.021*
6. *Coleman, C. M. (2011). Effects of negative keying and wording in attitude measures: A mixed-methods study. Doctoral Thesis. James Madison University*
7. *DiStenfano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. Structural Equation Modeling, 13(3), 440-464.*
8. *Edelen, M.O. & Reeve, B.B. (2007) Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. Quality of Life Research, 16, 5-18.*
9. *Embretson, S.E., & Reise, S.P. (2010). Item response theory for psychologists. Mahwah, NJ: Erlbaum*
10. *Hills, P., & Argyle, M. (2002). The Oxford Happiness Questionnaire: a compact scale for the measurement of psychological well-being. Personality and Individual Differences, 33, 1073-1082.*
11. *Hodge, D. R. & Gillespie, D. F. (2007). Phrase completion scale: A better measurement approach than likert scales. Journal of Social Service Research, 33 (4), 1-12.*
12. *Iweka, F. (2014) Comprehensive guide to test construction and administration. Omoku: Chifas Nigeria.*
13. *Kpolovie, P. J. (2010). Advanced research methods. Owerri: Springfield Publishers Ltd.*
14. *Leung, S. (2011). A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-Point Likert Scales. Journal of Social Service Research, 37 (4), 412-421.*
15. *Nwaialu, C. E. & Iruloh, B. N. (2014). Psychosocial factors as correlates of educational aspiration among secondary school students in Khana LGA of Rivers State. Reiko International Journal of Social and Economic Research,.7 (3A)*
16. *Orluwene, G. W. (2012). Fundamentals of testing and non-testing tools in educational psychology. Herley Publishers*
17. *Paul-Cookey, N.R. &Opara, I.M. (2015). Personality traits and marital adjustment of married teachers in secondary schools in Imo State. Journal of Education in Developing Areas (JEDA) 23 (1); 18-24.*
18. *Roszkowski, M. J., & Soven, M. (2010). Shifting gears: Consequences of including two negative worded items in the middle of a positively worded questionnaire. Assessment & Evaluation in Higher Education, 35, 117–134.*
19. *Samejima, F. (1969). A new family of models for the multiple-choice item research report 79-4 prepared under Office of Naval Research contract N00014-77-C-360, NR 150-402.Department of Psychology, University of Tennessee.*
20. *Sonderen E. V., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. PloS ONE, 8(7). doi:10.1371/journal.pone.0068967*

21. *Sternberg, R. J., & Grigorenko, E. L. (2001). A capsule history of theory and research on styles. In R. J. Sternberg, & L. F. Zhang (Eds.), Perspectives on thinking, learning and cognitive styles (pp. 1–21). Mahwah, NJ: Lawrence Erlbaum Associates.*
22. *Ye, F. & Wallace, T. L. (2014). Psychological sense of school membership scale: Method effects associated with negatively worded items. Journal of Psychoeducational Assessment, 32 (3), 202-215*
23. *Zhang, X. & Savalei, V. (2016). Improve the factor structure of psychological scales: The expanded format as an alternative to likert scale format. Educational and Psychological Measurement, 76 (3), 375-386.*