Customer Voice Retaliation (CVR) Test: Constructs Verification

Nor Irvoni Mohd Ishar and Rosmimah Mohd Roslin

Introduction

In the past, studies on customer complaining behavior have received a considerable attention in the marketing literature. However, despite the extensive research conducted, our understanding on its overall concept is still scarce, over simplified and does not reflect the full spectrum of the subject. Indeed, past research conducted on customer reaction from dissatisfied consumption experience have only focused on certain aspects of behavioural responses such as switch and complain or commonly known as complaining behavior (Singh 1988). It overlooks other possible aggressive response behaviours that might be performed by customers such as retaliation (Funches et al. 2009; Huefner and Hunt 2000). Therefore, the aim of this study is to explore customer retaliation as an extension to customer complaining behavior.

To help facilitate this study, a framework was developed based on pervious literatures to measure customer retaliation. The theory of equity is employed as the underpinning theory to explain the big picture of customer retaliatory behavior from dis-satisfied experience. According to Equity Theory (ET), people value fair treatment and have their own perception of fairness that serves as the basis to develop beliefs about what is a fair exchange reward (Adams 1963). Equity Theory also contended that when a person feels that the system or process is unjust, they will make attempts to achieve fairness or equitable relationship (Adams 1963; Pritchard 1969), or in other words they retaliate. There are many different ways that a customer can retaliate to achieve fairness. One of it is by voicing dissatisfaction aggressively. Therefore, for this study the term customer voice retaliation (CVR) will be used. However, in order for voice retaliation to take place, it requires

N.I. Mohd Ishar (✉) · R. Mohd Roslin
Faculty of Business and Management, Universiti Teknologi MARA,
Shah Alam, Malaysia
e-mail: irvoni@gmail.com; irvoni@salam.uitm.edu.my

Q. Zhang (ed.), *Pacific Rim Objective Measurement Symposium (PROMS) 2015 Conference Proceedings*,
DOI 10.1007/978-981-10-1687-5_18

a strong trigger. Often times the trigger is in the form of dissatisfied experience, and emotion. Therefore, in this study we will be investigating customer's dissatisfied experience attribution (DExSA) and emotional experience (EMEx) in relation to CVR.

Traditionally, researchers have been applying the Classical Test Theory (CTT) to analyze data. This scenario however has changed. Although many testing and measurement textbooks coined CTT as the only way to determine the quality of an assessment, the Item Response Theory (IRT) such as Rasch Measurement Model, does offer a sound alternative to the classical approach (Idowu et al. 2011). Indeed, Rasch Analysis has gained considerable attention among the social scientists and has been applied in various area of studies (De Battisti et al. 2005; Brentari and Golia 2008; Nor Irvoni and Mohd Saidfudin 2012; Nor Irvoni and Rosmimah 2016).

Rasch Analysis for Constructs Verification

Conventionally, social scientists rely on EFA to assess dimensionality, preliminary validity, and reliability aspects of their research instrument (Yau et al. 2007; Yoon et al. 2010). However, to ascertain the psychometric properties of an instrument, it demands for rigor and robustness in the methodological approach. It should not be restricted to domain specification but instead should aim at covering a range of the construct as wide as possible (Salzberger 2000). On that note, Rasch is seen as more appropriate method of analysis for construct verification process. Rasch represents a different philosophy of construct operationalization, provides superior foundation for assessing content validity as well as construct validity (Salzberger 2000).

Indeed, Rasch equips construct validation process with more rigor and robust analysis as it emphasizes for the items to cover different intensity level so that the entire breadth of the construct is represented (Ganglmair and Lawson 2003). Another important property of Rasch is its ability to provide researchers with interval level data. Although research using CTT has been assuming that the Likert-scale is interval, often times the response categories have a rank order, in which the intervals between the value labels cannot be presumed as equal (Jamieson 2004). Hence, the application of Rasch Rating Scale Model is most appropriate as it transforms the counts of endorsement in each response category of the rating scale into an interval scale (known as logits) based on the actual data (Grodin and Blais 2010). As a result, measures will be more meaningful, and the features of validity in terms of interpreting measures especially construct and content validity can be investigated within the Rasch measurement framework (Abdullah and Lim 2013; Smith 2005).

Rasch analysis is a method to obtain measures which are objective, fundamental, and linear and has been widely applied in education related research. In education, it is used to separate the ability of respondents and the quality of a test. It predicts the likelihood of how a person of different ability level for a particular trait should

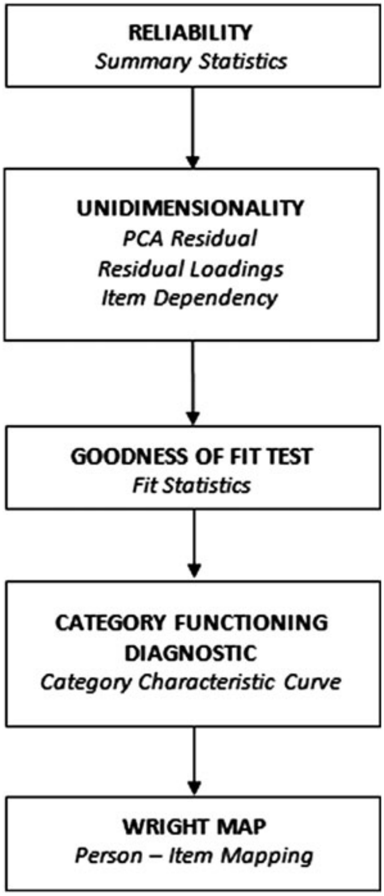


Fig. 1 CVR construct verification process

respond to an item of a certain level of difficulty. The probability of success depends on the difference between the ability of the person and the difficulty of the item (Bond and Fox 2015). From a marketing context, the term difficulty can be replaced by how hard it is to endorse an item or how extreme the item is (Ganglmair and Lawson 2003). Therefore, for CVR constructs verification, we followed the process as illustrated in Fig. 1.

Methodology

A. The Instrument

The instrument for this study comprises of three constructs represented by 66 items adapted from various sources where items were modified to suit the local context of the study. Six point Likert rating scales were used across all three constructs. Out of 66 items, 30 items are related to dissatisfied service experience attribution (DExSA), 16 items are on negative emotional experience (NEMEx), and 20 items are on customer voice retaliation (CVR). None of the items are negatively worded.

B. Pre-test Study

Prior to pilot testing, the instrument has undergone a pre-test for content and face validation. The pre-test was conducted in two phases. The first phase involves three domain area experts, three industry practitioners, and three Rasch Measurement practitioners. Feedbacks from experts and practitioners were used to revise unclear terms, and poorly worded questions. The second stage of the pre-test involves distributing the questionnaire to 27 subscribers in three separate sessions. From the pre-test, although alpha correlation score was high (0.96) and none of the Point Measure Correlation (PMC) register a negative value; the instrument did not fulfill the unidimensionality test of Rasch Analysis.

Further, although, the variance in data explained by measures is at 66.2 %, the eigenvalue unit is at 10.9, suggesting the existence of a second dimension with the strength of 10 items and hence need further examination. Other than that, there were also feedbacks from the respondents on certain terminology used in the questionnaire. Terms such as 'blow the whistle' and 'denigrated' were highlighted as unfamiliar. These, feedbacks were taken into consideration and the wordings were changed accordingly to suit the local context. A language expert from the International Islamic University of Malaysia was consulted and necessary corrections were made before the questionnaire was released for the pilot study.

C. Pilot Study

In identifying the suitable respondents for the study, the researchers applied the mall intercept technique. At selected locations, passers-by were asked if they would like to participate in the survey. Interested participants were then asked two qualifying question prior to answering the survey. The questions are; (i) have they experienced any dissatisfaction with their mobile telco service provider, and (ii) have they shared or spoken to anyone about their dissatisfaction. Only those who answered 'yes' to both questions qualified to answer the survey questionnaire. For this pilot test, a total of 66 mobile telco subscribers fulfilled both criterion, and were handed a set of questionnaire to fill in. They were given approximately 15–20 min and survey instrument were collected immediately after the completion time. From a total of 66 questionnaires, only 53 were Rasch analyzed using Winsteps 3.80.1. The other 13 were excluded due to straight lining response pattern.

Results and Discussions

A. Reliability

Summary statistics on 66 items and 53 persons are tabulated in Table 1. The Cronbach Alpha (KR-20) Person Raw Score Test Reliability was used to test for the internal consistency of the respondents' responses and can be considered a perfectly adequate index of the interitem consistency reliability (Sekaran 2003). For this study, it was found that the Cronbach Alpha value for the three constructs is considered as 'good' (Fisher 2007) with the value of 0.90, 0.94 and 0.91 for DExSA, EMEx, and CVR respectively. This is an indication that instrument has good internal consistency in measuring the latent traits.

Further, to ensure that the person fit the Rasch model reasonably well, the data need to fulfill the fit test conditions. From Table 1, results indicate that both the person Outfit Mean square, and z-standard values are very close to the expected value of '1' and '0' which is to be expected at the norm. Hence, it can be said that respondents for this pilot test do fit the Rasch model. Other than that, from the separation index, it can be concluded that the instrument is able to reliably separate the respondents apart into 3 to 4 distinct groups. In addition, person reliability index indicates that this order of item hierarchy will be replicated with a high degree of probability if the items were given to other comparable cohorts.

Table 1 Summary statistics (Items and person reliability coefficients)

		DExSA	EMEx	CVR
Items	Reliability	0.95	0.94	0.96
	Outfit MnSq	1.02	1.0	1.0
	Outfit Z-std	0.0	−0.1	−0.1
	Separation	4.54	3.82	4.81
	Max measure	1.58	2.00	1.23
	Min measure	−1.00	−0.87	−1.32
	Model error (Mean)	0.15	0.15	0.15
Persons	Cronbach alpha (KR-20)	0.90	0.94	0.91
	Reliability	0.88	0.93	0.88
	Outfit MnSq	1.02	1.0	1.0
	Outfit Z-std	−0.1	−0.2	0.0
	Separation	2.73	3.67	2.76
	Max measure	1.02	3.47	1.57
	Min measure	−2.06	−1.99	−2.37
	Model error (Mean)	0.21	0.28	0.24

Table 2 Principle component analysis

Constructs variance explained	DExSA			EMEx			CVR		
	Empirical		Model	Empirical		Model	Empirical		Model
Total raw variance in observations	54.2	100	100	38.9	100	100	43.4	100	100
Raw variance explained by measures	24.2	44.6	44.6	22.9	58.9	58.8	23.4	53.9	55.2
Raw variance explained by persons	4.2	7.7	7.7	9.9	25.4	25.4	5.8	13.4	13.7
Raw variance explained by items	20.0	36.9	36.9	13.0	33.4	33.4	17.5	40.5	41.4
Raw unexplained variance (total)	30.0	55.4	55.4	16.0	41.1	41.2	20.0	46.1	44.8
Unexplained variance in 1st contrast	4.3	8.0	14.4	3.1	7.9	19.3	4.6	10.6	23.1

B. Unidimensionality

An important part of Rasch validity analysis is unidimensionality. It is based on the value of raw variance explained by measure and unexplained variance in 1st contrast produced by Principle Component Analysis (PCA). The results of PCA for all 3 constructs are tabulated in Table 2.

From Table 2, raw variance explained by measures for all three constructs register a value of more than 40 %, and are nearly identical to the variance expected by the model, suggesting a strong principal measurement dimension (Conrad et al. 2011). Meanwhile, unexplained variance in 1st contrast showed an acceptable fair percentage because it was less than 15 % (Fisher 2007). Both percentages of raw variance explained by measures and unexplained variance in 1st contrast indicated that the 66 items-instrument used for measuring all three construct achieved the good criteria as it met the unidimensionality trait and was able to measure what it was intended to measure. As such, the analysis showed that the data for the 66 items had a very good fit to the Rasch measurement model and supports unidimensionality. Indeed, the PCA of the Rasch Model residual indicated that the underlying items for each constructs in the instrument are assessing a unidimensional measurement model.

Table 3 highlights the items that are suspected as problematic in which might contribute to a secondary dimension. Problematic items are, items with high residual loadings value (contrast loading >+0.6 and <−0.6). Overall, 10 out of 66 items did not fulfill the loading criteria. Items are COR_3 and SOC_3 from DExSA, rag_1 and rag_2 from EMEx, and PAt_4, WOM_3, WOM_4, VIN_2, VIN_4 and 3P_4 from CVR. Therefore, there is a need to cross check if any of these items are listed in the misfitting item list that violate the goodness of fit conditions before excluding them from the final survey.

Table 3 Standardized residual loadings for items

Constructs	Loading	Measure	MnSq		Item—Domain
			Infit	Outfit	
DExSA	0.72	0.56	0.98	1.00	20 CORS_3
	0.70	0.95	0.74	0.75	27 SOC_3
EMEx	0.67	0.38	0.65	0.64	5 rag_1
	0.62	0.22	0.88	0.84	6 rag_2
CVR	0.71	−0.79	0.59	0.59	20 PAt_4
	0.65	−1.32	0.85	0.86	3 WOM_3
	0.64	−0.47	0.69	0.68	4 WOM_4
	−0.68	0.39	0.95	0.88	6 VIN_2
	−0.61	0.32	1.07	0.97	8 VIN_4
	−0.61	0.44	1.35	1.27	20 3P_4

Rasch analysis has a mean of showing redundancy of items in which can be an indication for item deletion. Indeed, redundancy enables items reduction in order to shorten the length of an instrument (Green and Frantom 2002). Raw score residual correlations are used to detect dependency between pairs from the same domain. Items that are highly locally dependent (correlation >+0.7) share more than half of their "random" variance, suggesting that only one of the pair is needed for measurement. Hence, pairs form the same domain that have Large Standardised Residual Correlations are candidates for deletion. In this study, none of the items in Table 4 violates the local dependency criteria. Hence, none will be considered for deletion.

Another form of redundancy can be detected from items that have "same measure and same domain". In Table 5, is a list of items from each constructs by measures. Any two or more items that have the same measure and also testing on the same domain are not allowed to co-exist as they are measuring the same thing at the same difficulty level. To avoid the redundancy, item that is of lower quality (with negative PMC values) needs to be eliminated from the instrument. Scrutiny of items from the same dimension having the same measure indicates that although there are items in the respective constructs having the same measures (DExSA: 7 & 30, 11 & 28, 23 & 10; EMEx: 10 & 15; CVR: 1 & 21), none are measuring in the same domain indicating no redundancy.

C. Goodness of Fit

In assessing goodness of fit, Rasch requires the items to satisfy all three important fit index conditions. The results of fit indices for suspected problematic items for each constructs are tabulated in Table 6.

Table 4 Item dependency according to largest standardized residual correlations

Construct	Correlation	Item—Domain	Item—Domain
DExSA	0.68	12 PRO_1	13 PRO_2
	0.65	7 CON_2	30 SOC_6
	0.64	14 PRO_3	17 PRO_6
	0.59	25 SOC_1	10 CON_5
	0.55	1 eSE_1	26 SOC_2
	0.54	24 CORS_7	9 CON_4
	0.52	20 COR_3	29 SOC_5
	0.49	20 COR_3	27 SOC_3
	0.44	21 COR_4	27 SOC_3
EMEx	0.48	3 ang_3	9 dis_1
	0.43	5 rag_1	7 rag_3
	0.42	5 rag_1	6 rag_2
	0.40	15 sad_3	16 sad_4
	0.39	6 rag_2	7 rag_3
	0.37	2 ang_2	13 sad_1
	0.36	1 ang_1	2 ang_2
	−0.44	5 rag_1	12 dis_4
	−0.42	6 rag_2	14 sad_2
	−0.41	2 ang_2	7 rag_3
CVR	0.68	7 VIN_3	12 COL_1
	0.67	17 3P_1	20 3P_4
	0.66	6 VIN_2	8 VIN_4
	0.59	18 3P_2	20 3P_4
	0.58	7 VIN_3	14 COL_4
	0.52	4 WOM_4	22 PAt_2
	0.51	7 VIN_3	16 COL_4
	0.50	5 VIN_1	18 3P_2
	−0.53	3 WOM_3	6 VIN_2
	−0.50	5 VIN_1	22 PAT_2

One of the first things to observe in the data is the value of Point Measure Correlation (PMC). For this pilot study, all 66 items in the instrument have a positive PMC indicating that the instrument is measuring in the right direction. However, although the responses are moving in the same direction, 7 items (see Table 6) were identified as needing closer investigation as it violated the MnSq ($0.5 > y > 1.5$), and z-std ($-2 > z > +2$) conditions. An item is a misfit when its' MnSq and z-standard values are not within the stipulated acceptable range.

Table 5 Item dependency by measure order

DExSA			EMEx			CVR		
Item no	Measure	Domain	Item no	Measure	Item	Item no	Measure	Domain
15	1.58	PRO_4	8	2.00	rag_4	15	1.23	COL_3
21	1.43	COR_4	7	0.49	rag_3	7	0.84	VIN_3
14	1.11	PRO_3	5	0.38	rag_1	19	0.79	3P_3
19	1.01	COR_2	14	0.31	sad_2	16	0.71	COL_4
27	0.95	SOC_3	6	0.22	rag_2	14	0.63	COL_2
17	0.92	PRO_6	11	−0.01	dis_3	18	0.60	3P_2
29	0.69	SOC_5	13	−0.03	sad_1	13	0.56	COL_1
20	0.56	COR_3	10	−0.05	dis_2	20	0.44	3P_4
5	0.50	eSE_5	15	−0.05	sad_3	5	0.41	VIN_1
18	0.42	COR_1	16	−0.18	sad_4	6	0.39	VIN_2
7	0.40	CON_2	1	−0.23	ang_1	8	0.32	VIN_4
30	0.40	SOC_6	4	−0.29	ang_4	17	0.24	3P_1
8	0.05	CON_3	3	−0.48	ang_3	4	−0.47	WOM_4
26	0.03	SOC_2	2	−0.60	ang_2	2	−0.76	WOM_2
22	−0.02	COR_5	9	−0.62	dis_1	22	−0.78	PAt_2
1	−0.09	eSE_1	12	−0.87	dis_4	24	−0.79	PAt_4
13	−0.46	PRO_2	–	–	–	23	−0.83	PAt_3
2	−0.48	eSE_2	–	–	–	1	−1.11	WOM_1
11	−0.53	CON_6	–	–	–	21	−1.11	PAt_1
28	−0.53	SOC_4	–	–	–	3	−1.32	WOM_3
24	−0.61	COR_7	–	–	–	–	–	–
3	−0.63	eSE_3	–	–	–	–	–	–
6	−0.67	CON_1	–	–	–	–	–	–
9	−0.69	CON_4	–	–	–	–	–	–
12	−0.73	PRO_1	–	–	–	–	–	–
4	−0.85	eSE_4	–	–	–	–	–	–
25	−0.90	SOC_1	–	–	–	–	–	–
23	−0.94	COR_6	–	–	–	–	–	–
10	−0.94	CON_5	–	–	–	–	–	–
16	−1	PRO_5	–	–	–	–	–	–

From Table 6, only 1 item (sad_2) registers MnSq value that is outside the stipulated range (Infit MnSq: 1.66 logits/Outfit MnSq: 1.59 logits). Meanwhile, for z-std, 4 items (PRO_4, PRO_3, sad_2, and PAt_3) were found to not satisfying the acceptable range of −/+ 2 logits. Therefore, all these items will be checked against the residual loadings, and item dependency values. If the same item appears as problematic, the items are a candidate for deletion and will not be included in the actual study. However, if the items were not identified as problematic, the items can be considered as fit and will be retained for actual survey.

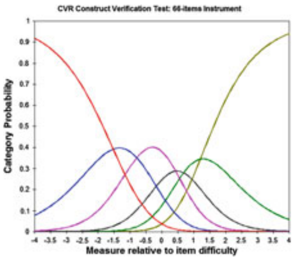
Table 6 Misfitting items by construct

Construct	Measure	Infit		Outfit		PMC	Item—Domain
		MnSq	z-STD	MnSq	z-STD		
DExSA	1.58	1.53	2.4	1.50	2.3	0.38	15 PRO_4
	1.11	1.50	2.3	1.51	2.4	0.37	14 PRO_3
EMEx	0.31	1.66	2.9	1.59	2.7	0.53	14 sad_2
	−0.87	1.29	1.5	1.56	2.4	0.60	12 dis_4
CVR	−0.83	1.42	2.1	1.61	2.9	0.07	7 PAt_3
	0.63	1.48	2.1	1.29	1.3	0.54	14 COL_2
	0.56	1.47	2.1	1.33	1.5	0.47	13 COL_1

D. Category Functioning Diagnostic

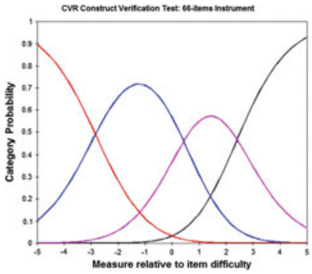
The final step of our construct verification process is rating scale calibration. It is a process by which the categories used in the instrument were analyzed for its functionality. This process is very crucial in any measurement because its validity significantly affects measurement precision. Indeed, rating scale is the way how researchers communicate with respondents as they attempt to use the response restrictions (Bond and Fox 2015). However, this process it is often overlooked.

A valid scale is when all categories are functioning optimally in which enough data are represented in each thresholds. The difference in the threshold should be 1.4 logits apart but not exceeding 5 logits (Bond and Fox 2015; Linacre 1999). Rasch rating scale model (RSM) has the capacity to provide evidence for such claims as it allows researcher to extract the most meaning from the data. However, if the categories are not functioning as expected, then collapsing will take place as remedy. Therefore, the 6 response categories used in this pilot study will be re-examined to determine which categorization of responses yielded higher-quality measures. Indeed, revision of the rating scale should be done at the pilot phase in the development of the measure (Bond and Fox 2015), prior to actual data collection. Figure 2 depicts the probability curves for CVR construct.




```
SUMMARY OF CATEGORY STRUCTURE.  Model="R"
-------------------------------------------------------------
|Cat    Observed  |Average|Infit  Outfit||         |Category|
|Label   Count    |Measure| MNSQ   MNSQ ||Threshold| Measure|
|-----------------+-------+-------------++---------+--------|
|  1      537     | -1.42 |   .87    .91||  NONE   |( -2.89)| 1
|  2      830     |  -.79 |  1.00    .99||   -1.53 |  -1.33 | 2
|  3     1007     |  -.34 |  1.04   1.02||    -.79 |   -.30 | 3
|  4      563     |   .05 |  1.05   1.10||     .43 |    .48 | 4
|  5      360     |   .37 |  1.09   1.09||     .68 |   1.29 | 5
|  6      198     |   .81 |   .98   1.05||    1.20 |(  2.63)| 6
|-----------------+-------+-------------++---------+--------|
|MISSING    3    0|  -.11 |             ||         |        |
-------------------------------------------------------------
```

Fig. 2 Six-rating scale category characteristic curve (CCC)




```
SUMMARY OF CATEGORY STRUCTURE.  Model="R"
---------------------------------------------------------
|Cat   Observed |OBSVD  |Infit Outfit||Threshold|Category|
|LABEL  Count   |Measure| MNSQ  MNSQ||         | Measure|
|---------------+-------+------------++---------+---------
|  1     537    | -2.38 |   .93   .94||  NONE   |( -3.97)| 1
|  2    1837    |  -.85 |   .99   .97||   -2.84 |  -1.23 | 2
|  3     923    |   .30 |  1.05  1.06||     .43 |   1.43 | 4
|  4     198    |  1.39 |  1.01  1.04||    2.41 |(  3.60)| 6
|---------------+-------+------------++---------+--------|
|MISSING   3   0|  -.24 |            ||         |        |
---------------------------------------------------------
```

Fig. 3 Four-rating scale category characteristic curve (CCC)

Analysis revealed that the original six-rating scale does not function effectively as depicted in Fig. 2. Overlapping peaks indicates that respondents are not able to differentiate between the categories in which will disrupt construct definition. This is further supported by the threshold estimates values between the categories which are less than 1.4 logits. Therefore, collapsing problematic categories would be a good remedy to overcome such problem. There are two general rules for collapsing categories. First it has to be logical, and second it should create a more uniform frequency distribution (Bond and Fox 2015). Figure 3 depicts a four-rating scale CCC after collapsing process took place. Note that the thresholds estimates value between categories have improved to more than 1.4 logits indicating a well-functioning scale.

E. Wright Map

This is the heart of Rasch Analysis and will be the premise of the instrument construct validity acceptance. It shows the logical hierarchy of item difficulty based on the conceptual theory put under test. A good item construct is evident when it is represented in the vertical direction. Being in horizontal direction shows redundancy of items measuring the same thing and is not desirable. Only when the item difficulty hierarchy is in place, then it is said the instrument has construct validity. Figure 4 depicts the item hierarchy map for each of the constructs for this study.

From the Wright Map it can be clearly seen that overall, items has a good hierarchical order with item measuring range of a mere 2.58, 2.87 and 2.55 logits for each of the construct respectively. It is also evident that a large number of items can be found along the continuum on which the majority of respondents' abilities fall. However, there are gaps in the hierarchical order for items belonging to EMEx construct which would require more items in between item rag_4 and rag_3 so that better meaning and measures can be achieved.

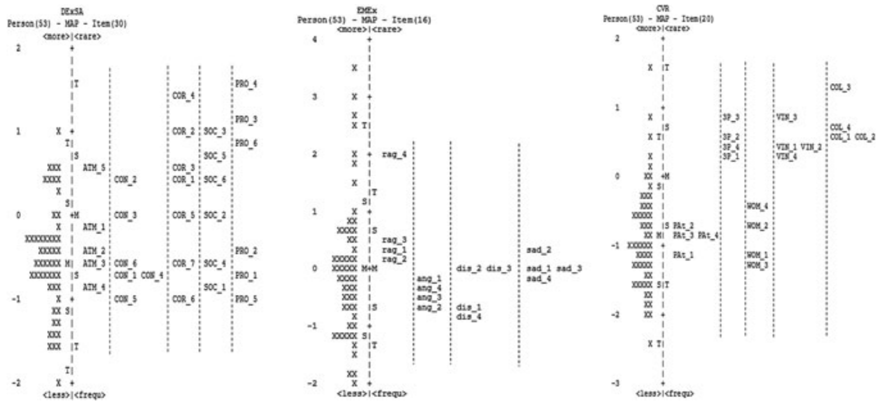


Fig. 4 Wright map for DExSA, EMEx and CVR

Other than that, there are also items that appear in horizontal order. Items are CON_1 & CON_4 from DExSA, dis_2 & dis_3, sad_1 & sad_3 from EMEx, and PAt_3 & PAt_4, VIN_1 & VIN_2, COL_1 & COL_2 from CVR. These items should be checked against the item dependency table (Table 4). If the pairs are not listed in the table, then the items is not a candidate for deletion.

Conclusion

In light of the aforementioned evidences, it seems that the application of Rasch analysis in refining research instrument facilitates the development of a more powerful tool for measurement. Rasch exposed the items to series of rigorous tests, producing measures with interval level data, which is an important requirement for high level analysis. As a result, the instrument yielded measures that have better fit and quality. Therefore, are more likely to produce more reliable and valid findings. Indeed, cross checks on the analyses confirmed that none of the suspected problematic items should be eliminated from the final instrument. However, analysis on category functioning curve suggests that the 6-point Likert rating scale should be collapsed to 4 categories to produce better measures. Therefore, the 66-items instrument with 4-point Likert rating will be used for actual study.

Other than that, the findings of this study would be very significant for organization in measuring customers' complaining behaviour as it provides a basis for a valid instrument construct that gives a better and true linear measure. This paper will also facilitate in adding new knowledge to existing literature in relation to consumer behavioral study.

Regarding future research, the logit measures obtained from this study should be imputed to other software such as smartPLS to investigate the relationship among the constructs as by doing so could help to produce better analysis and more accurate results with interval level data. It cannot be denied that without the application of Rasch, good measurement is hampered in the absence of reliable instrument.

Acknowledgments The authors are grateful to all involved at the Faculty of Business and Management, Universiti Teknologi MARA, and to the Ministry of Higher Education Malaysia (MoHE) for the support given in carrying out this study. This study is funded by the FRGS Grant (Ref: FRGS/1/2015/SS01/UITM/01/1).

References

Abdullah, N., & Lim, B. K. (2013). Parallel circuit conceptual understanding test (PCCUT). *Procedia—Social and Behavioral Sciences*, *90*(InCULT 2012), 431–440.

Adams, J. S. (1963). Toward an understanding of inequity. *Journal of Abnormal Psychology, 67*(5), 422–436.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd Edn). Routledge.

Brentari, E., & Golia, S. (2008). Measuring job satisfaction in the social services sector with the Rasch model. *Journal of Applied Measurement, 9*(1), 45–56.

Conrad, K. J., Conrad, K. M., Dennis, M. L., Riley, B. B., & Funk, R. (2011). Validation of the substance problem scale to the Rasch measurement model. *GAIN Methods Report, 1*, 2.

De Battisti, F., Nicolini, G., & Salini, S. (2005). *The Rasch model to measure service quality*.

Fisher, W. J. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transaction, 21*(1), 2007.

Funches, V., Markley, M., & Davis, L. (2009). Reprisal, retribution and requital: Investigating customer retaliation. *Journal of Business Research, 62*(2), 231–238.

Ganglmair, A., & Lawson, R. (2003). Measuring affective response to consumption using Rasch modeling. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior, 16*, 198–210.

Green, K., & Frantom, C. (2002). Survey development and validation with the Rasch model. In *International conference on questionnaire development, evaluation, and testing* (pp. 1–30).

Grodin, J., & Blais, J. -G. (2010). A Rasch analysis on collapsing categories in item's response scales of survey questionnaire: Maybe it's not one size fits all. In *Annual meeting of the American educational research association* (pp. 1–29).

Huefner, J. C., & Hunt, H. K. (2000). Consumer retaliation as a response to dissatisfaction. *Journal of Consumer Satisfaction Dissatisfaction and Complaining Behavior, 13*(1), 61–82.

Idowu, O., Eluwa, A. N., & Abang, B. K. (2011). Evaluation of mathematics achievement test: A comparison between classical test theory (CTT) and item response theory (IRT). *Journal of Educational and Social Research*, *1*(4), 99–106. ISSN 2240-0524.

Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education, 38*(12), 1217–1218.

Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement, 3*(2), 103–122.

Nor Irvoni, M. I., & Mohd Saidfudin, M. (2012). Students' perception towards quality library service using Rasch measurement model. In *2012 international conference of innovation, management and technology research (ICIMTR2012)* (pp. 668–672).

Nor Irvoni, M. I., & Rosmimah, M. R. (2016). Customer voice retaliation (CVR) construct verification : A Rasch analysis approach. In *Procedia economics and finance* (Vol. 37, pp. 214–220).

Pritchard, R. (1969). Equity theory: A review and critique*1. *Organizational Behavior and Human Performance, 4*(2), 176–211.

Salzberger, T. (2000). An alternative way of establishing measurement in marketing research—its implications for scale development and validity. In *ANZMAC 2000 visionary marketing for the 21st century: facing the challenge* (pp. 1111–1117).

Sekaran, U. (2003). *Research method for business—A skill building approach*. (J. Marshall, Ed.) (4th ed.). John Wiley & Sons, Inc.

Singh, J. (1988). Consumer complaint intentions and behavior : Definitional and taxonomical issues. *Journal of Marketing*, 93–107.

Smith, G. T. (2005). On construct validity: Issues of method and measurement. *Psychological Assessment, 17*(4), 396–408.

Yau, O. H. M., Chow, R. P., Sin, L. Y., Tse, A. C., Luk, C., & Lee, J. S. (2007). Developing a scale for stakeholder orientation. *European Journal of Marketing, 41*(11/12), 1306–1327.

Yoon, H. J., Song, J. H., Donahue, W. E., & Woodley, K. (2010). Leadership competency inventory: A systematic process of developing and validating a leadership competency scale. *Journal of Leadership Studies, 4*(3), 39–50.