Development and Validation of A Multidimensional Measure: Manufacturing Practices, Knowledge Transfer and Manufacturing Capabilities in Malaysian Manufacturing Companies

Jusoh, M.S., PhD, 1Amlus, M.H., PhD, 1Osman, A., Dba and 2Abidin, R.

1University Malaysia Perlis (UniMAP), School of Technopreneurship and Business Innovation, 01000 Kangar, Perlis, Malaysia
2University Utara Malaysia (UUM), Technology Management and Logistic, 06400, Sintok, Kedah, Malaysia

ABSTRACT

Instrument construct is one of the most important issues in conducting research. Without proper consideration in tackling the issue, it is difficult for the instrument to be considered as valid and reliable. If construct validity is accurate, then it will provide a clearer and more precise descriptive analysis on the concepts being investigated. The most important criteria that need to be considered in answering the construct validity are reliability and validity. In traditional measurement model, the understanding of reliability and validity is totally different from the one offered by Rasch model. Regardless of the difference, an internal consistency is still widely being used as the primary method to measure the item sampling. Using this model, local dependence and item fit are most considered in getting valid, reliable, and consistent, hence its significance. Finally the study applies goodness of measure purposely to answer issues related to validity and reliability tests. Principal components analysis was carried out to test the construct of questionnaires used in the study.

INTRODUCTION

Reliability relates to ability of a measure to remain the same; consistently over time (Sekaran, 2003) or the same result is obtained when the same research is repeated or does it again once more. In order to get the reliability of the test, Cronbach-α is used as common value in estimating the internal consistencies of the items (Onwuegbuzie & Danial, 2002). The Cronbach-α value should considerably be higher than the acceptance level of 0.60 (Garson, 1998; Gliem & Gliem, 2003; Leedy & Ormrod, 2005) to be measured as reliable. Rasch Analysis further provides the reliability of person and item arising from the interaction between both subjects in the test. If the Person reliability registered 0.76, it is described as ‘Fair’ reliability. Otherwise if the Person reliability calculated at 0.81-0.90 is considered as good, 0.91-0.94 reflected as very good and >0.94 indicated as excellent (Fisher Jr., 2007). This specifies the Person or sample involved in the assessment with reliability equal to or greater than 0.6, stand a high chance to produce a repeat outcome for the next time if the same set of questionnaire are given to them (Andrich, 1988). Azrilah (2010) further explained, if the selected group of respondents were to be given a different set of quality management-cost of quality-organizational performance tool or questionnaire with the same dimension, the probability of the ability pattern (in person measure order table) or the location of the quality management-cost of quality-organizational performance on the person-item distribution map would remain similar.

The purpose of having validity is to make sure the instrument in use is measuring what it is supposedly to measure (Sekaran, 2003; Zickmund et al., 2010). The criterion validity of such findings represents the actual likelihood of the situation. Since most of the questions were adapted from previous studies, the issue of face validity is to assure the meanings of the questions given are measuring the underlying concept (Sekaran, 2003). Though some researchers believed that face validity is not a valid component of content validity, it remains a very important process to determine the suitability of the questions given posted to the respondents. The content validity itself has to be done as it will ensure the questionnaire include the adequate and represent or sufficient and enough items to represent the subject matter or the concept ushered in the study (Sekaran, 2003). This will later be validated by the model hence answering the content validity issue. Similarly with the item reliability.
provided, it will determine whether the instrument is having sufficient number of questions for all range of respondents.

The Rasch’s generate important information about the items whether it is measuring in a single direction or behaving erratically by functioning in the opposite direction. In this case, it can be reported that the instrument is behaving in bi-direction. Rasch ‘zero-set’ the instrument when the item is at a threshold point equals to midpoint 0.5 hence a situation of 50:50. Then, it calibrating the rating scale to ascertain the assumed rating is valid for use. If the threshold separation between any two ratings is less than 1.4, then the assumed rating is collapsed and re-calibrated to ensure better Infit Standard Deviation or invariance obtained. This is the crucial test involve as the procedures itself determines the reliability of the respondents and construct validity of the instrument hence valid data (Andrich, 1988; Bond & Fox, 2007; Fisher Jr., 2007; Linacre, 2008).

There are three indicators need to be fulfilled before one can claim that the instrument employed, i.e. the questionnaire as reliable and valid thus replicable and measuring what we are supposedly to measure. The explanation summarized by Azrilah (2010):

a) Cronbach- a value (should be more than 0.6)
b) Item Reliability value – to answer whether the question is valid or not. If the reliability value is > 0.6; then the questions asked in the study is sufficient for the expected range of respondents. If the score is less than <0.6, then the number of questions asked is insufficient thus invalid instrument construct.
c) Person Separation value – to show the ability of the instrument to discriminate the respondents into distinct groups. If the instrument cannot separate the respondents as expected, then the items need to be reviewed; either rephrased or new item added.
d) Person Reliability value – gives an indication of the person latent trait measures or psychometry. If the score > 0.6 meaning the person involve in the study is reliable and if the score<0.6 meaning that the person is not reliable. Meaning that if the result shows high reliability, it means that the location of person along the ruler will be the same for the second time if an instrument of the same construct were taken by the same respondents.

Methodology:

Goodness of measure is performed in answering the issues related to validity and reliability tests. Principal Components Analysis (PCA) was carried out to test the unidimensionality of the questions used in the study. Basically the purpose in Rasch is to “perform a Principal Components Analysis of the residuals by item correlation matrix. The first factor reported here is really the second factor, because the Rasch dimension is the first factor overall. This second factor identifies the strongest pattern in local dependency among the items as reflected in their correlations” (Wright, 1996). However, Sick (2011) views Principal Component Analysis as “an extension of Rasch fit analysis used to confirm whether the Rasch difficulty dimension; (thus the construct) has adequately accounted all of the non-random variance in the data”.

As the items represent in this study fit the model then it supported the unidimensionality of the scale hence explain the goodness of content validity (Wright, 1996; Sick, 2011). Principal Component Analysis allows the researcher to refine the instrument construct further by conducting the process of elimination to choose which item fits best by looking at the item quality compliance. Hence, it allows only quality items to best describe the variables being used.

Tennant and Pallant (2006) further details the Principal Components Analysis capabilities to support the post-hoc testing, having undertaken the Rasch analysis and supposing fit the Rasch model requirements. Principal Components Analysis constituted of two major steps, started with choosing the item fits best from the Local Item Dependence (LID) requirement and the quality compliance to Item Measure standard. Yen (1993) and Zenisky, et al. (2003) suggested using the local item dependence is to detecting the dependency between pairs of items or persons.

In order to select the relevant items to represent the construct, a criterion for local dependence loading had been discussed by few authors. Yen (1984) and Yen (1993) suggested a small positive adjustment to the correlation of size 1/(L-1) where L is the test length. Local dependence would be large positive correlation, with highly locally dependent items (Correlation > 0.7) suggesting that only one of the two items is needed for measurement based Table Principal Component Analysis : Largest Standardized Residual Correlations in Winsteps 2006 (Wright et al., 2000).

Further analysis in fulfilling the Rasch criterion through misfit analysis will show the items that are inconsistent with the construction of a single measure. The first step involve in this process is sharing more than half of their random variance (Wright, 1999). As suggested by Wright (1999), the only one of these two items is needed for measurement thus one of the items has to be discarded or removed. Item from the same domain which exhibit measure of MNSQ further away than 1 and the z-Std further than ‘0’ shall be dropped. Further justification by Tennant and Pallant (2006) suggested, the analysis of the residuals was conducted in detecting the second factors after the Rasch factor is removed due to understanding that, “originally interpretation of this was difficult as the proportion of variance attribute to the first residual factor was reported but the total
variation in the data was unknown”. Subsequently, Linacre (2006) had incorporated the total variation into presentable and understandable reporting as shown in Table 1.

Table 1: Principal Component Analysis of Standardized Residual variance (in Eigenvalue units).

<table>
<thead>
<tr>
<th></th>
<th>Observed</th>
<th>Modeled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total raw variance in observations</td>
<td>170.1 100.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Raw variance explained by measures</td>
<td>55.1 32.4%</td>
<td>34.3%</td>
</tr>
<tr>
<td>Raw variance explained by persons</td>
<td>19.5 11.5%</td>
<td>12.1%</td>
</tr>
<tr>
<td>Raw variance explained by items</td>
<td>35.6 20.9%</td>
<td>22.1%</td>
</tr>
<tr>
<td>Raw unexplained variance (total)</td>
<td>115.0 67.6%</td>
<td>100.0% 65.7%</td>
</tr>
<tr>
<td>Unexplained variance in 1st contrast</td>
<td>21.3 12.5%</td>
<td>18.5%</td>
</tr>
<tr>
<td>Unexplained variance in 2nd contrast</td>
<td>16.0 9.4%</td>
<td>13.9%</td>
</tr>
<tr>
<td>Unexplained variance in 3rd contrast</td>
<td>12.6 7.4%</td>
<td>10.9%</td>
</tr>
<tr>
<td>Unexplained variance in 4th contrast</td>
<td>9.7 5.7%</td>
<td>8.4%</td>
</tr>
<tr>
<td>Unexplained variance in 5th contrast</td>
<td>7.9 4.7%</td>
<td>6.9%</td>
</tr>
</tbody>
</table>

The magnitude data of the first residual factor using the Rasch factor can be determined easily as compared to other measurement model (Tennant & Pallant, 2006; Sick, 2011; Wright, 1999). Reckase (1979) accept raw variance explained by measures greater than 20% but Rasch requires 40% as an indicator of uni-dimensionality. Generally as shown in table 1, it can be considered since the modeled variance is 34.3% and the unexplained variance is measured at 12.5% nearing the limit of 15%. Thus, further test is needed in order to improve the uni-dimensionality of the instrument.

The next steps will examine the item redundancy or possible multi-collinearity through item pairs. It is important to note the local dependency specifies that the value of one data has no influence on another once the underlying data has been accounted (Wright, 1996). As such, data from this study were fitted to the model and tested for appropriate category ordering if local dependency and principal component analysis were done since this two methods are equally answered the uni-dimensionality and multicollinearity methods in classical test theory (Yen, 1993; Zenisky et al., 2003; Salzberger & Sinkovics, 2006; Pallant et al., 2006).

Further investigation is needed to further evaluate the person’s / respondent’s responses towards best fit to the model. The method known as fit statistic is implemented to execute the process (Boone & Scantlebury, 2005). Further literature by Bond and Fox (2007), explained the fit statistics from the technical justification as, “the use of chi-square fit statistic to determine how well any set of empirical data met the requirements of this model”. Same approaches being recommended by Wright (1977) and Linacre (2002), where both of them recommended similar steps toward detecting the dissimilarity among items. It is important to identify how respondent’s pattern accurately or predictably fit the model by converting the mean-square statistics to the normally-distributed z-standardized. To abridge the findings, Azrilah (2010) recommended four (4) criteria as to check for any outliers or misfits data, as any misfits pattern to be considered are focused on the requirements given, that are:

(1) Point Measure Correlation: (Pt-Mea Corr); 0.4 < Pt-Mea Corr value < 0.85.

(2) Point Measure Correlation (PT-Mea Corr); gave a negative value (meaning that the person is predicted a misfit due to careless respond or guessing).

(3) Outfit Mean Square (MNSQ); 0.5 < Outfit MNSQ value < 1.5

(4) Outfit Z-Standard (Z-STD); -2 < Outfit Z-Std value <+2

If the items under investigation do not meet the above criteria thence, the items can be discarded due to poor quality fit.

Conclusions and implications:

Generally, the findings from the Rasch measurement model in answering the reliability, validity and the level of significance are totally different from the classical test theory (CTT) approach. In CTT, it is generally focused on items (or test) rather than persons or items independently (Bell, 1982). However, the Rasch model is concerned about what people do in testing or estimating the person’s ability. This supported by Andrich’s (1982), argue on the shortfalls of CTT computation for Cronbach-alpha based on raw score as compared to Rasch analysis using the probabilistic model. Additionally, the observed variance among the person parameter estimates and the result thereof are used to construct measurement scales which yield a better value of reliability as compared to Cronbach alpha (KR-20). With the execution of Rasch’s principal component analysis and misfits pattern, there are total 30 unfit items are discarded from the original 115 items. Further checking on the content validity shows that the person reliability and item reliability increase as compared to before the elimination being made.
REFERENCES


