Research Journal of Applied Sciences 13 (4): 241-244, 2018
ISSN: 1815-932X

Crucial Role of Exploratory Data Analysis (EDA) in Assessing Survey Instrument Reliability

[1]Suzilah Ismail, [2]Rosna Awang Hashim and [3]Norhafezah Yusof
[1]School of Quantitative Sciences,
[2]School of Education and Modern Languages,
[3]School of Multimedia Technology and Communication, Universiti Utara Malaysia,
Kedah, Malaysia

Abstract: Commonly, survey instrument reliability is achieved through Cronbach's alpha and Exploratory Factor Analysis (EFA). However, the existence of different values rule of thumb often creates confusion for novice researchers. Thus, the study aims to investigate the crucial role of Exploratory Data Analysis (EDA) in assisting the reliability testing of a survey instrument via. box plot to clearly understand the instrument's items. The instrument was adapted and adopted from well-established instrument of the National Survey of Student Engagement (NSSE) which consists of 4 constructs, 13 dimensions and total of 59 items. The study focuses on student-faculty interaction dimension which comprises 6 items. The finding shows that item 1a is causing the internal inconsistency which jeopardizing the reliability of the student-faculty interaction dimension. Hence, by using EDA in depth understanding is obtained and cross checked with Cronbach's alpha and EFA values in removing the item from the instrument. Thus, emphasizing the crucial role of EDA in assessing survey instrument reliability where in depth understanding regarding the item offers alternative solution to overcome the problem of different rule of thumb of Cronbach's alpha and EFA values.

Key words: Exploratory data analysis, survey instrument, reliability, Cronbach's alpha, exploratory factor analysis, NSSE

INTRODUCTION

According to Tukey (1977), "Exploratory Data Analysis (EDA) is detective work-numerical detective work-or counting detective work or graphical detective work unless exploratory data analysis uncovers indications, usually quantitative ones, there is likely to be nothing for confirmatory data analysis to consider, it can never be the whole story but nothing else can serve as the foundation stone as the first step". This quote by the founder of EDA reflects the importance of EDA as the first step in understanding data. Commonly in survey instrument development requires the validity and reliability assessing process as in Fig. 1. The content and face validity involves feedback from expert and respondent of small focus group, respectively (Zikmund *et al*., 2013). Meanwhile, the reliability usually based on Cronbach's alpha and exploratory factor analysis (UCLAIDRE, 2016). The problem with both of these methods is the different values of rule of thumb in deciding to maintain or remove the item/s. According to

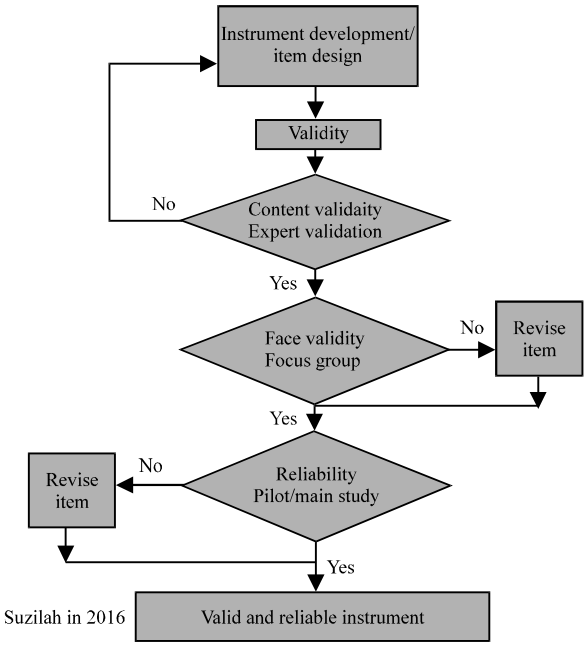


Fig. 1: Validity and reliability process of the instrument

Corresponding Author: Suzilah Ismail, School of Quantitative Sciences, Universiti Utara Malaysia, Kedah, Malaysia

DeVellis, ideally the Cronbach's alpha coefficient should be above 0.7 but Zikmund *et al.* (2013) used above 0.6. As for the exploratory factor analysis, the factor loading value is 0.3 (Coakes, 2013) and 0.5 (Hair *et al.*, 2010). This leads to the confusion of using the most appropriate rule of thumb especially for novice researchers. These values are solely technical where the process of assessing reliability become too robotic and skip the part of understanding.

In this study, we highlight the crucial role of EDA in assisting the reliability testing of survey instrument via. box plot to clearly understand the item/s.

MATERIALS AND METHODS

The instrument was adapted and adopted from well-established instrument of the National Survey of Student Engagement (TIU, 2011). Table 1 lists the 4 constructs, 13 dimensions and total of 59 items. The instrument measurement scale was modified from 4 Likert scale to 7 point numerical scale (Never to very often) because it takes into consideration of interval scale instead of ordinal which allowed more powerful statistical analysis to be implemented. The instrument passes through content validity (based on expert) and face validity which comprises of small focus group of full time final year students in Universiti Utara Malaysia (Fig. 1). Pilot and main study were conducted in assessing the instrument reliability that involves 163 and 493 randomly selected students, respectively.

In this study we propose to use Exploratory Data Analysis (EDA) prior to Cronbach's alpha and Exploratory Factor Analysis (EFA) in diagnosing the instrument reliability. EDA is an approach in analyzing data which uses variety of methods (mostly graphical) to maximize insight into a data set (Smith and Prentice, 1993; Keim *et al.*, 2006). The graphical chosen is box plot because we want to identify underlying patterns of the items in each dimension (Table 1) and understand the co-existence of the items in reflecting the dimension.

Table 1: Details of constructs, dimensions and number of items

Constructs	Dimensions/Abbreviation	No. of items
Adapted from NSSE		
Student engagement	Student-Faculty Interaction (SFI)	6
	Student-Student Interaction (SI)	5
	Diversity (DIV)	3
	Effort (EFF)	6
University experiences	Deep Learning (DL)	4
	Course Workload (CW)	5
	Opportunities (OPP)	5
	Viewpoints (VI)	3
Student growth	Personal Social Growth (PSG)	7
	Academic Growth (AG)	7
Student opinion	Quality of Relationship (QR)	5
	Social Climate (SC)	4
	Academic Climate (AC)	3

RESULTS AND DISCUSSION

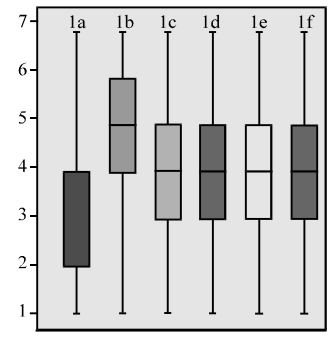
Table 2 presents pilot study results for the first dimension, i.e., student-faculty interaction which comprises of six items together with Exploratory Factor Analysis (EFA) factor loading, Cronbach's alpha value and Cronbach's alpha value if an item is deleted. Overall Cronbach's alpha for the 6 items is 0.798. Item 1a (Used email to communicate with lecturer) has the lowest EFA factor loading 0.335 and highest Cronbach's alpha 0.817 if this item is removed from the student-faculty interaction dimension. Followed by item 1b (discussed grades or assignments with lecturer) which has 0.489 EFA factor loading but if this item is deleted the Cronbach alpha is 0.784 which is lower than overall Cronbach's alpha value 0.798. Usually, based on these two values (EFA factor loading and Cronbach's alpha), we remove item 1a from the instrument but retain item 1b before we proceed with main study. This process of removal or maintaining items solely depend on these values becomes a robotic process and confusion arises when there are different rule of thumbs regarding those values (Zikmund *et al.* 2013; Coakes, 2013; Hair *et al.*, 2010). Therefore, understanding why item 1a is contributing to low factor loading and Cronbach's alpha is important because when we understand the reasons, it is more meaningful for us to remove or retain the item to be in the instrument and overcome the issues of different rule of thumbs. One of the ways to understand is via. Exploratory Data Analysis (EDA) that is using box plot.

Figure 2 displays the box plot of the six items. Box plots item 1a and b are showing different patterns as compared to the other four items (1c-f). Box plot item 1a indicates many students use less email to communicate with lecturer. As for box plot item 1b reveals many students often discussed grades or assignment with lecturers. Further, investigations by interviewing the students confirms that they preferred to have face to face communications rather than e-Mail and commonly the communications are regarding assignments and grades. This is due to the fact they are full time students who have twice a week face to face classes and 2 h of face to face consultation. Thus, they have many opportunities to communicate with the lecturers via. face to face rather than through email. Perhaps if this study was conducted on part time students, the results may be different. The EDA results align with EFA factor loading and Cronbach's alpha as in Table 1. However, the advantage of using EDA make us understand better regarding the EFA and Cronbach alpha values. Those values are calculated based on correlation, therefore, similar pattern of items will contribute to higher values of EFA and Cronbach's alpha because in reliability these values measures internal consistency of the items to represent the dimension. Although, we have understood and obtain

Table 2: Student faculty interaction: EFA and Cronbach alpha (Pilot study, n = 163)

	Pilot	
Dimension: student faculty interaction (items)	EFA factor loading	Cronbach alpha if item deleted
1a. Used e-Mail to communicate with lecturer	0.335	0.817
1b. Discussed grades or assignments with lecturer	0.489	0.784
1c. Talked about career plans with a lecturer or faculty advisor	0.671	0.732
1d. Discussed ideas from readings or classes with lecturer	0.718	0.731
1e. Received prompt written or oral feedback regarding academic performance from lecturer	0.645	0.758
1f. Worked with lecturer in other activities	0.580	0.768
Cronbach alpha: 0.798		

Table 3: EDA (Box plot), EFA factor loading and cronbach alpha of student faculty interaction (Main study)

			Cronbach alpha
Items	EDA (Boxplot)	EFA factor loading	Cronbach alpha if item deleted
1a		0.366	0.795
1b		0.481	0.762
1c		0.669	0.725
1d		0.727	0.722
1e		0.629	0.762
1f		0.584	0.758
Cronbach alpha: 0.787			

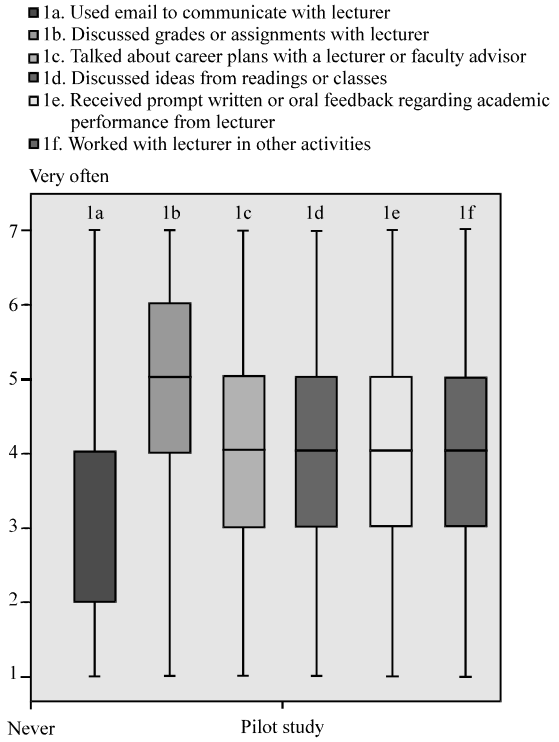


Fig. 2: Box plot of student faculty interaction (Pilot study)

genuine reasons to remove item 1a that is student really use less e-Mail to communicate with lecturer but since, this study is a pilot which involve small sample size, i.e., 163 students, we decided to maintain item 1a in the student-faculty interaction dimension and conducted main study using the same instrument.

Table 3 presents main study results for the first dimension, i.e., student-faculty interaction which comprises of 6 items together with EDA (via. box plot), EFA factor loading, Cronbach's alpha value and Cronbach's alpha value if item deleted. The box plot shows similar pattern as in pilot study where many students use less email to communicate with lecturer. This is also reflected by the lowest factor loading 0.366 and highest Cronbach's alpha 0.795 if item 1a is removed from the student-faculty interaction dimension. This proof that item 1a is not suitable to be in the instrument because using small (pilot study) or large sample size (main study), both reveals the students use less email to communicate with lecturer. Therefore, item 1a should be removed from the instrument. Similar approach was also applied on the rest of the 12 dimensions (Table 1) in assessing reliability of the instrument.

CONCLUSION

This study highlight the crucial role of EDA in assessing survey instrument reliability where in depth understanding regarding the item helps to overcome the problem of different rule of thumb of EFA factor loading and Cronbach's alpha values. The EDA (via. box plot) should be used together with EFA factor loading and Cronbach's alpha values in order to maximize insight into

a data set and avoiding confusion based on too robotic reliability testing process. Thus by accompanying EDA in the reliability testing can provide better guidelines, especially to novice researchers in decision making regarding the items either to retain or remove it from the instrument.

REFERENCES

Coakes, S.J., 2013. SPSS Version 20.0 for Windows: Analysis without Anguish. 20th Edn., John Wiley and Sons, Australia, ISBN: 978-1-118-33776-9, Pages: 282.

Hair, J.F., W.C. Black, B.J. Babin and R.E. Anderson, 2010. Multivariate Data Analysis. Pearson, London, UK.,.

Keim, D.A., F. Mansmann, J. Schneidewind and H. Ziegler, 2006. Challenges in visual data analysis. Proceedings of the 10th International Conference on Information Visualization, July 5-7, 2006, IEEE, London, England, UK., ISBN:0-7695-2602-0, pp: 9-16.

Smith, A.F. and D.A. Prentice, 1993. Exploratory Data Analysis. In: A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues, Keren, G.E. and C.E. Lewis (Eds.). Lawrence Erlbaum Associates, New Jersey, USA., ISBN:9780805810363, pp: 349-390.

TIU., 2011. National survey of student engagement. Trustees of Indiana University, Indianapolis, Indiana. www.nsse.iub.edu.

Tukey, J.W., 1977. Exploratory Data Analysis. Addison-Wesley, Boston, Massachusetts, USA., ISBN:9780201076165, Pages: 688.

UCLAIDRE., 2016. SPSS FAQ: What does cronbach's alpha mean?. UCLA Institute for Digital Research and Education, Los Angeles, California. http://www.ats.ucla.edu/stat/spss/faq/alpha.html.

Zikmund, W.G., B.J. Babin, J.C. Carr and M. Griffin, 2013. Business Research Methods. Cengage Learning, Boston, Massachusetts, USA., ISBN: 13-978-1-111-82694-9, Pages: 669.