



## Invalid skewed responses contributes to invalid factor solution in exploratory factor analysis: A validation approach using real-life data

Mohamad Adam Bujang<sup>1,2</sup>, Puzziawati Ab Ghani<sup>2</sup>, Shahrul Aiman Soelar<sup>3</sup>, Nor Aizura Zulkifli<sup>4</sup>, Evi Diana Omar<sup>5</sup>

<sup>1</sup>Faculty of Computer and Mathematical Sciences, Universiti Teknologi Mara, Shah Alam, Malaysia

<sup>2</sup>Clinical Research Centre, Sarawak General Hospital, Ministry of Health, Kuching, Malaysia

<sup>3</sup>Clinical Research Centre, Hospital Sultanah Bahiyah, Alor Setar, Malaysia

<sup>4</sup>Clinical Research Centre, Hospital Raja Permaisuri Bainun, Ipoh, Malaysia

<sup>5</sup>Institute for Clinical Research, Ministry of Health, Selangor, Malaysia

### ABSTRACT

**Background:** This study aimed to investigate the potential contribution of invalid skewed responses to invalid factor solution in the result from exploratory factor analysis. In the present study, “the invalid skewed response” is defined as when majority of the respondents consistently rate only at one side which will eventually change the real or valid pattern of overall responses.

**Methods:** A validation approach was conducted using a secondary data from a questionnaire validation study of an eight Likert scale that has a very stable and strong factor solution. Eight sub samples were retrieved from the data to represent multiple sets of analyses with sample size based on rule of thumbs from 3:1 until 10:1. From each sub sample, proportion of dummy response for the extreme left (scale of 0), the middle scale (scale of 4), and the extreme right (scale of 7) were assigned randomly at 10%, 20%, and 30%, respectively.

**Results:** The invalid consistent responses of a middle scale have very low impact toward the factor solution. The occurrence of the invalid skewed responses affected the factor solution. Majority of the factor solutions were still valid based on consistent responses with 10.0% only. However, the construct that was based on forcing into four-factor solution had helped to produce the valid factor solution though some resulted in cross-loadings. All Cronbach’s alphas and minimum corrected item to total correlation were relatively strong for all factor solutions although some of the solutions were invalid.

**Conclusions:** The skewed responses have the potential to change the ideal factor solution. Therefore, necessary steps need to be taken to avoid invalid skewed responses, especially in self-administered survey. Therefore, the recommended sample size guideline for exploratory factor analysis with justifications is proposed.

### ARTICLE HISTORY

Received June 28, 2019

Accepted December 18, 2019

Published December 30, 2019

### KEYWORDS

Exploratory factor analysis;  
invalid skewed responses;  
rule of thumb; sample size

## Introduction

Exploratory factor analysis (EFA) is a statistical tool used by researchers to identify the underlying relationships between measured variables [1]. This technique is widely used to develop or validate questionnaires, such as Children Depression Index, Depression Anxiety and Stress Scales, Quality of life, and The Summary of Diabetes Self-care Activities (SDSCA) [2–5]. There are certain criteria and assumptions need to be met before EFA can be used. The factor solution depends much on the

strength of the items or variables and the sufficient sample size [6–10]. Previous study has found that different measurement scales in Likert scale type of response also has a relation to sample size in producing correct factor solution [10].

Sample size guideline for EFA has been widely discussed in existing literatures [6,8–10]. However, these guidelines lack in addressing the effect from behavioral perspective of the respondents. The behavior of respondents in providing the responses when answering a questionnaire have to

**Contact** Mohamad Adam Bujang ✉ adam@crc.gov.my 📧 UiTM, Shah Alam, Malaysia.

be incorporated because these responses may be potentially influenced by the alteration of people's behaviors, including their actions, emotions, and thoughts. Therefore, the present study hypothesized that in any self-administered questionnaire study that require EFA, a behavioral effect in responding questions may contribute to invalid skewed response that eventually can produce invalid factor solution. This is because, the invalid skewed responses rated by respondents on the same Likert scale could change the overall true pattern of responses which lead to invalid factor solution.

In the present study, "the invalid skewed responses" is defined as when majority of respondents consistently rate only at either side which will eventually change the valid pattern of overall responses. Questionnaires studies that have been analyzed using EFA usually use ordinal scale where most Likert-type scales contain either five or seven response categories [11–13]. In the Likert scale of five, the invalid skewed responses may refer to respondents who consistently rate at either one extreme, such as "strongly agree" or "strongly disagree." or point of uncertainty "neither agree nor disagree."

We expected few scenarios of how the behavioral effect may contribute to the invalid skewed responses. Respondents may have tendencies to rate at either one extreme of the scale due to time constraint. This could be due to either when they respond to questionnaire in a hurry or when there are too many questions or items to be answered within limited time. Other than that, respondents might also be afraid or reluctant to think hard to give in their true opinions. In such situation, they might feel more convenient to rate at the middle scale that represents neutral and safe opinion [14]. Thus, the invalid skewed responses could happen in any observational studies which use an ordinal scale in self-administered questionnaires [15].

The other possible reason of invalid skewed responses could be due to skewed sampling where respondents who only rate scale 1 or 5 for Likert scale of 5 are selected. This situation commonly occurs among extreme respondents whereby their opinions are only extreme at one side (totally disagree or agree) or when their condition could also be too good or too poor. The selected respondents could happen by chance though researchers usually aim to recruit respondents with various characteristics to ensure the response will reflect the general construct of a specific measurement in the population. The skewed responses from these types

of respondents can influence the solution for the construct; hence, the sample size estimation for EFA is necessary to be revisited.

Although ample of guidelines for sample size estimation for EFA are currently available, however, these guidelines have not taken into consideration of the behavioral effect that leads to the invalid skewed responses. As invalid skewed responses could affect the validity of factor solution, a comprehensive study that focuses on the effect of the invalid skewed responses toward the valid factor solution is warranted. The aim of this study was to evaluate and explain the effect of the invalid skewed responses toward obtaining the valid factor solution for EFA. This is particularly important to increase the awareness among researchers to avoid having such responses by taking the necessary steps. With the findings, we propose a revised guideline for sample size determination of EFA.

## Methods

This study used data from a previously published study. It is a validation study of a questionnaire, "SDSCA" in Malay version among patients with type II diabetes mellitus (T2DM) [16]. The SDSCA uses eight Likert-scale. It has four main domains, i.e., diet, exercise, blood sugar monitoring, and foot care with total of 10 main items where each domain has two items except for diet domain with four items. The study recruited 225 patients and results found that the SDSCA Malay version is a reliable and valid instrument to be used among T2DM patients [16].

The first 100 patients with complete responses were used for the analysis. Eight subsets of samples were extracted to represent the set of analysis with different sample size of 30, 40, 50, 60, 70, 80, 90, and 100 to represent the sample size based on rule of thumbs from 3:1 (3 respondents for 1 item) to 10:1 (10 respondents for 1 item). For each subset, the proportion of dummy response for scale of 0 (represents extreme at left side), scale of 4 (represents middle scale), and scale of 7 (represents extreme at right side) were assigned randomly at 10%, 20%, and 30%, respectively.

These dummy responses are representing the possibility of the invalid skewed responses rated by respondents in a self-administered survey. In other words, the invalid skewed responses represent the possibility of variation in behavioral effect of the respondents in rating the scale. The objective of these settings is to determine to what extent the volume of consistent dummy responses and ratings

could contribute to invalid factor solution. In this study, the valid factor solution is defined when the construct yields 4 main domains such as diet (items 1,2,3, and 5A), exercise (items 5 and 6), monitoring sugar control (items 7 and 8), and foot care (items 9 and 10) [5]. The items that fall into the respective domains should also report the highest loadings for the targeted domains.

EFA was conducted using the Principle Component Analysis extraction method with varimax rotation. The number of factor solutions produced by the analysis was based on eigenvalues of more than one and also was based on setting the four-factor solutions [5]. Cronbach’s alphas with the minimum corrected items to total correlation (CITC) in each domain were reported from some parts of the overall results. The purpose is to prove that although statistics such as Cronbach’s alpha with more than 0.70 and corrected item to total correlation with more than 0.30, but the factor solution may not necessary be valid. All analyses were carried out using SPSS version 20.0 (IBM Corp. Released 2011. IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.).

## Results

Based on the minimum guideline with 3:1 ratio (sample size of 30), result yielded valid factor solution with minimum factor loading of 0.534 and there was no cross loading that had occurred (Table 1).

Based on the effect from the invalid consistent responses of a middle scale (scale of 4 for Likert

scale of 0 to 7), the valid factor solutions were still produced although with the responses of the middle scale up to 30%. This shows that the invalid consistent responses of a middle scale have very low impact toward the factor solution. The lowest rule of thumb such as 3:1 and 4:1 also produced the correct factor solution, however, with the occurrence of minimal cross loadings (Table 2).

Based on the effect of the extreme left and right scale (scale of 0 or 7 for Likert scale of 0 to 7), the occurrence of the invalid skewed responses does affect the factor solution. Majority of the factor solutions were still valid based on consistent responses with 10.0% only. However, majority of the factor solutions were invalid based on the invalid skewed responses up until 30.0%. However, the construct that based on forced into four factor solution had helped to produce the valid factor solution although some of the results yielded cross loadings especially for the invalid skewed responses with scale of 7 (Tables 3 and 4).

All Cronbach’s alphas and minimum corrected item to total correlation were relatively strong for all factor solutions although some of the solutions were invalid. For example, with sample size of 40 (based on rule of thumb 4:1) and the invalid skewed responses of scale 1 for 30%, result still yielded high Cronbach’s alpha (more than 0.70) and CITC (more than 0.30) although the solution was invalid (Table 5). Result should yield 4-factor solution as showed in Table 1 instead of with three-factor solution as shown in Table 5. The recommended sample size guideline for EFA with justifications is presented in Table 6.

**Table 1.** Factor solution of SDSCA with sample data n=30 without skewed responses.

Items	Diet	Blood glucose monitoring	Exercise	Foot care
Q1	0.915			
Q2	0.857			
Q5A	0.592			
Q3	0.534			
Q8		0.959		
Q7		0.958		
Q6			0.901	
Q5			0.899	
Q10				0.887
Q9				0.845
Cronbach's alpha	0.749	0.944	0.802	0.775
Minimum corrected item-to-total correlation	0.451	0.894	0.671	0.633

## Discussion

The present study has shown that the invalid skewed responses do affect the valid or an ideal factor solution. The invalid skewed responses refer to a situation where majority of respondents invalidly and consistently rated their responses only at one side from the Likert scale in which it changes the real or true pattern of overall responses. Besides that, the solution that is obtained either based on eigenvalue of more than one or by forcing to a specific number of factor approach are also matter. When the approach of eigenvalue more than one is applied as the basis to determine the appropriate number of distinct factors, this study found that the factor solution was likely invalid if the invalid skewed responses were appeared in the data.

EFA is commonly used to determine the number of latent variables that are required to explain

**Table 2.** Summary of results for EFA with skewed responses of scale 1 (from 0 to 7 Likert scale) based on various sample sizes.

Rule of thumb	number (%) of skewed responses	Eigen value more than 1			Forced to 4 factor solution		
		Correct factor solution	Minimum factor loading >0.4	Cross loading	Correct factor solution	Minimum factor loading >0.4	Cross loading
3:1	3 (10%)	Yes	Yes	No	Yes	Yes	No
	6 (20%)	Yes	Yes	No	Yes	Yes	No
	9 (30%)	No			Yes	Yes	Yes
4:1	4 (10%)	Yes	Yes	No	Yes	Yes	No
	8 (20%)	No			Yes	Yes	No
	12 (30%)	No			Yes	Yes	No
5:1	5 (10%)	Yes	Yes	No	Yes	Yes	No
	10 (20%)	Yes	Yes	No	Yes	Yes	No
	15 (30%)	No			Yes	Yes	No
6:1	6 (10%)	Yes	Yes	No	Yes	Yes	No
	12 (20%)	Yes	Yes	No	Yes	Yes	No
	18 (30%)	No			Yes	Yes	No
7:1	7 (10%)	Yes	Yes	No	Yes	Yes	No
	14 (20%)	No			Yes	Yes	No
	21 (30%)	No			Yes	Yes	No
8:1	8 (10%)	Yes	Yes	No	Yes	Yes	No
	16 (20%)	No			Yes	Yes	No
	24 (30%)	No			Yes	Yes	No
9:1	9 (10%)	Yes	Yes	No	Yes	Yes	No
	18 (20%)	Yes	Yes	No	Yes	Yes	No
	27 (30%)	No			Yes	Yes	No
10:1	10 (10%)	Yes	Yes	No	Yes	Yes	No
	20 (20%)	No			Yes	Yes	No
	30 (30%)	No			Yes	Yes	Yes

the correlations among a set of observed variables. Commonly, EFA is used for questionnaire development studies or questionnaire validation studies. The present study used secondary data of a validation study of SDSCA questionnaire [16]. The construct of SDSCA in this paper was sufficiently strong with excellent factor loadings and the result had produced valid factor solution although a minimum sample size based on 3:1 ratio was used. Therefore, this dataset is suitable to be used to evaluate the impact from the invalid skewed responses.

From these results, this study summarized three things. First, although the minimum requirement of sample size for EFA is based on three cases per variable (3:1 ratio) for at least Likert scale of four as suggested by previous study, but researchers should consider increasing the sample size to overcome the effect of the invalid skewed responses [10]. To incorporate invalid skewed responses,

the present study suggests that the least appropriate rule of thumb is based on five cases per one variable (5:1 ratio). Such rule of thumb also has been recommended by other studies [8,17].

However, increasing sample size alone might not help to get valid factor solution when the invalid skewed responses are not being managed properly. Therefore, the invalid skewed responses have to be kept to the minimum level at best. The invalid skewed responses are potentially occurring from two main sources. The first source is due to attitude of the respondents. Few recommendations to address this would be limiting the number of questions, ensuring sufficient time for respondents to fill up the questionnaires, and providing incentives whenever appropriate.

The other source is due to skewed sample where the study recruited majority of respondents that have some extreme characteristics. From the design

**Table 3.** Summary of results for EFA with extreme middle responses of scale 4 (from 0 to 7 Likert scale) based on various sample sizes.

Rule of thumb	number (%) of skewed responses	Eigen value more than 1			Forced to 4 factor solution		
		Correct factor solution	Minimum factor loading >0.4	Cross loading	Correct factor solution	Minimum factor loading >0.4	Cross loading
3:01 <i>n</i> = 30	3 (10%)	Yes	Yes	Yes	Yes	Yes	Yes
	6 (20%)	Yes	Yes	Yes	Yes	Yes	Yes
	9 (30%)	Yes	Yes	Yes	Yes	Yes	Yes
4:01 <i>n</i> = 40	4 (10%)	Yes	Yes	Yes	Yes	Yes	Yes
	8 (20%)	Yes	Yes	No	Yes	Yes	No
	12 (30%)	Yes	Yes	Yes	Yes	Yes	Yes
5:01 <i>n</i> = 50	5 (10%)	Yes	Yes	No	Yes	Yes	No
	10 (20%)	Yes	Yes	No	Yes	Yes	No
	15 (30%)	Yes	Yes	No	Yes	Yes	No
6:01 <i>n</i> = 60	6 (10%)	Yes	Yes	No	Yes	Yes	No
	12 (20%)	Yes	Yes	No	Yes	Yes	No
	18 (30%)	Yes	Yes	No	Yes	Yes	No
7:01 <i>n</i> = 70	7 (10%)	Yes	Yes	No	Yes	Yes	No
	14 (20%)	Yes	Yes	No	Yes	Yes	No
	21 (30%)	Yes	Yes	No	Yes	Yes	No
8:01 <i>n</i> = 80	8 (10%)	Yes	Yes	No	Yes	Yes	No
	16 (20%)	Yes	Yes	No	Yes	Yes	No
	24 (30%)	Yes	Yes	No	Yes	Yes	No
9:01 <i>n</i> = 90	9 (10%)	Yes	Yes	No	Yes	Yes	No
	18 (20%)	Yes	Yes	No	Yes	Yes	No
	27 (30%)	Yes	Yes	No	Yes	Yes	No
10:01 <i>n</i> = 100	10 (10%)	Yes	Yes	No	Yes	Yes	No
	20 (20%)	Yes	Yes	No	Yes	Yes	No
	30 (30%)	Yes	Yes	No	Yes	Yes	No

of the study itself at the very beginning, researchers must ensure that the questionnaire they develop suit the targeted population knowing their extreme characteristics. The differences of population characteristics could yield different factor solutions [2,18]. For questionnaire development studies or questionnaire validation studies that measuring latent variables, the possible characteristics of the population has to be first identified. Then, the researchers need to carefully select the respondents who represent almost fair variation of the characteristics in the particular population.

Second, this study also recommends researchers to consider using force by a specific number of factors solutions for validation studies since the number of specific domains are known [19]. However, in a situation where the specific domains are uncertain, especially during exploratory work, researchers need to conduct EFA by using larger

samples and to conduct the EFA based on the eigenvalue more than one approach.

Third, the high values of Cronbach’s alpha (more than 0.70) and the respective CITC (more than 0.30) do not confirm the ideal factor solution. The finding is supported by previous study [10]. Although the factor solutions were wrong when based on eigenvalue more than one approach, the Cronbach’s alpha and the respective minimum CITC could still be possible to be high [20,21]. These indicate the importance of scientific relevance rather than just relying on statistical point of view. This finding shows that the statistics such as Cronbach’s alpha and CITC can only be the supporting evidence to indicate the strength of factor solution after the ideal and valid factor solution is produced [10,22].

From these findings, a guideline of sufficient sample size for EFA is proposed. The summary of sample size guideline is presented in Table 6 where

**Table 4.** Summary of results for EFA with skewed responses of scale 7 (from 0 to 7 Likert scale) based on various sample sizes.

Rule of thumb	number (%) of skewed responses	Eigen value more than 1			Forced to 4 factor solution		
		Correct factor solution	Minimum factor loading >0.4	Cross loading	Correct factor solution	Minimum factor loading >0.4	Cross loading
3:01	3 (10%)	Yes	Yes	Yes	Yes	Yes	Yes
	6 (20%)	No			Yes	Yes	Yes
	9 (30%)	No			Yes	Yes	Yes
4:01	4 (10%)	No			Yes	Yes	Yes
	8 (20%)	No			Yes	Yes	Yes
	12 (30%)	No			Yes	Yes	Yes
5:01	5 (10%)	No			Yes	Yes	No
	10 (20%)	No			Yes	Yes	Yes
	15 (30%)	No			Yes	Yes	Yes
6:01	6 (10%)	Yes	Yes	No	Yes	Yes	No
	12 (20%)	No			Yes	Yes	Yes
	18 (30%)	No			Yes	Yes	Yes
7:01	7 (10%)	Yes	Yes	No	Yes	Yes	No
	14 (20%)	No			Yes	Yes	No
	21 (30%)	No			Yes	Yes	Yes
8:01	8 (10%)	Yes	Yes	Yes	Yes	Yes	Yes
	16 (20%)	No			Yes	Yes	Yes
	24 (30%)	No			Yes	Yes	Yes
9:01	9 (10%)	Yes	Yes	Yes	Yes	Yes	Yes
	18 (20%)	No			Yes	Yes	Yes
	27 (30%)	No			Yes	Yes	Yes
10:01	10 (10%)	No			Yes	Yes	No
	20 (20%)	No			Yes	Yes	Yes
	30 (30%)	No			Yes	Yes	Yes

**Table 5.** Factor solution of SDSCA with sample size  $n = 40$  and skewed sample of 30.0% with scale 1.

Items	Factor 1	Factor 2	Factor 3
Q1	0.906		
Q5A	0.862		
Q2	0.853		
Q3	0.836	0.401	
Q9	0.537		0.437
Q6		0.932	
Q5		0.820	
Q10	0.520	0.581	
Q8			0.933
Q7			0.913
Cronbach's alpha	0.847	0.758	0.876
Minimum corrected item-to-total correlation	0.482	0.548	0.787

this summary of sample size guideline for EFA. Previously, the least minimum sample size guideline for EFA is 2:1 ratio, but this guideline is only restricted for variables in numerical form such as ratio and interval type of measurement. To yield an ideal factor solution, the construct of latent variables has to be formulated based on scientifically and biologically reasons [10].

For studies that are based on self-administered survey, the recommended minimum sample size guideline is 3:1 ratio. In this present study, such small sample size is still sufficient to yield a correct factor solution (Table 1). However, considering the effect of the invalid skewed responses, researchers are advised to recruit more respondents. Therefore, this study recommends the ideal sample size guideline for EFA in self-administered survey is 5:1 ratio. This rule of thumb is recommended after

**Table 6.** The proposed rules of thumb for EFA and their remarks.

Rules of thumb	Remarks
2:1	Minimum guideline for EFA. This guideline is sufficient for an excellent construct and with excellent factor loadings. So far, it was proven works only for variables in numerical form (not in a Likert scale type of response) and the variables are biologically or scientifically correlated instead of measuring variables that are based on responses or feedback from respondents [10].
3:1	Recommended as minimum guideline for EFA for self-administered survey. This guideline is sufficient for an excellent construct with excellent factor loadings. The effect of skewed responses has to be kept in minimum. The guideline is not recommended for survey that uses Likert scale below than five. The application of forced into specific number of factor may help to produce a relevant construct. Not encourage for studies related to questionnaire development.
5:1	It is an ideal guideline for EFA for an observational studies based on self-administered survey. Such rule of thumb is reasonably sufficient to incorporate skewed responses. Recommended for research that related to validation studies. However, such sample size is also sufficient for questionnaire development studies provided the items are excellent in producing a good construct.
>5:1 to 10:1	Is necessary if larger sample size is needed to capture a fair variation of subjects from the intended population especially when the population number is large. The guideline is also necessary for lower type of ordinal scale. Recommended for research that to explore and develop a relevant construct of latent variables such as in questionnaire development studies.
>10:1	Researchers have the flexibility to recruit more subjects or respondents in their studies. But, taking subjects or respondents more than 20:1 mainly for EFA could waste a resources. Unless the larger sample size is necessary to support other statistical inferences such as to determine a certain incidence of a particular variable in a population with very high accuracy or for the purpose of disease modeling in which usually require multivariate analysis.

incorporating the effect from the invalid skewed responses. To reduce the impact of the invalid skewed responses, researchers can apply force the factor solution to a specific number of factors solution. The guideline of 5:1 ratio is also recommended for questionnaire development studies provided the items are excellent in producing a good construct.

In general, larger sample size is necessary (e.g., sample size from more than 5:1 ratio up to 10:1 ratio) for questionnaire development studies. In this type of study, the factor solution usually will be restricted to be based on eigenvalue more than one approach instead of setting up the specific number of factor solution. The larger sample size provides an advantage to include a fair variation of respondents in the population. Besides that, the larger sample size is also necessary for questionnaire with ordinal scale such as Likert scale of three [10]. However, irrespective of sample size, the effect of the invalid skewed responses can still contribute to invalid factor solution. Therefore, it is still necessary to control the impact of the invalid skewed responses.

Taking sample size more than 10:1 ratio for EFA might not be necessary as the excellent construct does not need large sample size. Thus, taking respondents or respondents such as more than 20:1 mainly for EFA could be a waste of resources. However, the larger sample size may be necessary if a particular study is also aimed to support other statistical inferences such as to determine a certain

proportion (%) of a particular variable in a population with an excellent accuracy or if the study has another important objective to be fulfilled such as for multivariate analysis [23–25]. Basically, EFA does not need large sample size because the sample size for EFA is highly dependent on the number of items to be investigated and to what extent the items can produce reliable and valid factor solutions. Thus, this is mainly determined by the choice and the strength of individual questions or items [9,10].

The present study has few limitations, especially on the selection of an extraction and rotation methods. Since the aim was to study the effect of the invalid skewed responses toward the changes in factor solution; therefore, this study limited the analysis using specific extraction and rotation method. In addition, this study used secondary data from study that used questionnaire with a very strong and stable construct. Hence, other data from studies with less stable construct may give different results. Besides that, future studies may use simulation analysis to assess the effect of invalid skewed responses in EFA.

## Conclusions

This study concluded that in EFA, the behavioral effect is potential to give invalid skewed responses that usually occur in self-administered surveys and eventually could affect the ideal factor solution.

Appropriate plan in recruiting the respondents must be executed to avoid the invalid skewed responses that could change the ideal pattern of overall responses. In addition, it is also important to assess the data quality to determine whether or not the data have outliers due to the invalid skewed responses prior to taking necessary actions to clean the data. The revised sample size guideline for EFA is proposed to ease researchers in their sample size planning that require EFA. The rule of thumb with minimum sample size guidelines for different scenarios is necessary, especially in a situation where respondents are difficult to get such as in clinical setting. Forcing the solution to a specific number of factors is a reasonable technique to be considered, especially when the number of factor solution is known.

### Acknowledgement

We would like to thank the Director General of Health Malaysia for his kind permission to publish this article and Dr Lee Keng Yee for her effort to conduct proofread of this article.

### Abbreviations

EFA	Exploratory Factor Analysis
SDSCA	Summary of Diabetes Self-care Activities
T2DM	Type II Diabetes Mellitus
CITC	Corrected item to total correlation
SPSS	Statistical Package for Social Science

### Conflict of interests

The authors declared that they have no conflict of interests.

### Funding

None.

### References

- [1] Norris M, Lecavalier L. Evaluating the use of exploratory factor analysis in developmental disability psychological research. *J Autism Dev Disord* 2009; 40(1):8–20.
- [2] Kovacs M. *Children's depression inventory*. Multi-Health System, North Tonawanda, NY, 1992.
- [3] Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behav Res Ther* 1995; 33:335–43.
- [4] McHorney CA, Ware JE, Lu JR, Sherbourne CD. The MOS 36- Item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care* 1994; 32:40–66.
- [5] Toobert DJ, Glasgow RE. Assessing diabetes self-management: the summary of diabetes self-care activities questionnaire. In: Bradley C (ed.). *Handbook of psychology and diabetes*. Harwood Academic, Switzerland, pp 351–75, 1994.
- [6] Barrett PT, Kline P. The observation to variable ratio in factor analysis. *Personality Study Group Behav* 1981; 1:23–33.
- [7] Cattell RB. *The scientific use of factor analysis in behavioral and life sciences*. Plenum Press, New York, 1978.
- [8] Gorsuch RL. *Factor analysis*. 2nd edition, Erlbaum, Hillsdale, NJ, 1983.
- [9] MacCallum RC, Widaman KF, Zhang S, Hong S. Sample size in factor analysis. *Psychol Methods* 1999; 4:84–99.
- [10] Bujang MA, Ghani PA, Soelar SA, Zulkifli NA. Sample size guideline for exploratory factor analysis when using small sample: taking into considerations of different measurement scales. *Statistics in Science, Business, and Engineering (ICSSBE), International Conference IEEE, 2012*. doi:10.1109/ICSSBE.2012.6396605
- [11]. Bearden WO, Netmeyer RG, Mobley MF. *Handbook of marketing scales: multi-item measures for marketing and consumer behavior research*. Sage, Newbury Park, CA, 1993.
- [12] Peter JP. Reliability: a review of psychometric basics and recent marketing practices. *J Market Res* 1979; 16:6–17.
- [13] Shaw ME, Wright JM. *Scales for the measurement of attitudes*. McGraw-Hill, New York, 1967.
- [14] Kamoen N, Holleman B, Mak P, Sanders T, Van-Den-Bergh H. Agree or disagree? Cognitive process in answering contrastive survey questions. *Discourse Process* 2011; 48:314–32.
- [15] Friedman HH. Rating the rating scales. *J Mark Manage* 1999; 9:114–23.
- [16] Bujang MA, Ismail M, Mohd-Hatta NKB, Baharum N, Othman SH, Mat-Lazim SS, et al. Validation of the summary of diabetes self-care activities (SDSCA) in Malay language for Malaysian adults. *Malaysian J Public Health Med* 2016; 16(3):227–34.
- [17] Hatcher L. *A step-by-step approach to using the SAS® System for factor analysis and structural equation modeling*. SAS Institute Inc., Cary, NC, 1994.
- [18] Tan SM, Loh SF, Bujang MA, Haniff J, Rahman A, Nazri F, et al. Validation of the Malay version of children's depression inventory. *Int Med J* 2013; 20(2):189–92.
- [19] Jalaludin MY, Fuziah MZ, Hong JYH, Mohamad-Adam B, Jamaiyah H. Reliability and validity of the revised summary of diabetes self-care activities (SDSCA) for Malaysian children and adolescents. *Malays Fam Physician* 2012; 7:10–20.

- [20] Nunally JC. Psychometric theory. 2nd edition, McGraw-Hill, New York, 1978.
- [21] Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 2nd edition, Oxford University Press, Oxford, UK, 1995.
- [22] Sijtsma K. On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika* 2009; 74:107–20.
- [23] Bujang MA, Sa'at N, Joys AR, Mohamad-Ali M. An audit of the statistics and the comparison with the parameter in the population. AIP Conference Proceedings. 1682:050019, 2015. doi:10.1063/1.4932510
- [24] Bujang MA, Sa'at N, Tg-Abu-Bakar-Sidik TMI. Determination of minimum sample size requirement for multiple linear regression and analysis of covariance based on experimental and non-experimental studies. *Epidemiol Biostat Public Health* 2017; 14(3):e12117(1–9).
- [25] Bujang MA, Sa'at N, Tg-Abu-Bakar-Sidik TMI, Lim CJ. Sample size guidelines for logistic regression from observational studies with large population: emphasis on the accuracy between statistics and parameters based on real life clinical data. *Malays J Med Sci* 2018; 25(4):122–30.