

# A Step-by-Step Guide to **Questionnaire Validation Research**



**Dr. Mohamad Adam Bujang**  
**Hon Yoon Khee**  
**Dr. Lee Keng Yee**

---

**Editor: Prof. Dr. Shamsul Azhar Shah**

# **A Step-By-Step Guide to Questionnaire Validation Research**

**Authors**

**Mohamad Adam Bujang  
Hon Yoon Khee  
Lee Keng Yee**

**Editor**

**Prof. Dr. Shamsul Azhar Shah**

Institute for Clinical Research (ICR),  
National Institutes of Health, Blok B4, Aras 1, Jalan Setia Murni U13/52,  
Seksyen U13, Setia Alam, 40170 Shah Alam, Selangor, Malaysia  
[www.crc.gov.my](http://www.crc.gov.my) | Phone: 603-3362 7700

© INSTITUTE FOR CLINICAL RESEARCH (ICR), Mohamad Adam B., Hon YK, & Lee KY, 2022.

Cover, graphic design and typesetting service by Chew Cheng Hoon.  
All rights reserved.

The authors retain the sole copyright ownership of this book. Therefore, it is necessary to obtain prior permission from either the author or the Director of Institute for Clinical Research, Ministry of Health, Malaysia, in order to reuse the content(s) of this book and/or to reproduce the content(s) of this book in any form. However, the author wishes for this book to serve as a useful and not-for-sale reference for managing and conducting medical research consultations.

## **A Step-By-Step Guide to Questionnaire Validation Research**

First Edition, 2022.

Cite as: Mohamad Adam B., Hon YK, & Lee KY. (2022). **A STEP-BY-STEP GUIDE TO QUESTIONNAIRE VALIDATION RESEARCH**. Kuala Lumpur, Malaysia: Institute for Clinical Research, NIH MY.

ISBN: 9789671694060 (MY)

DOI: 10.5281/zenodo.6801209

### **Authors**

Dr. Mohamad Adam Bujang  
Clinical Research Centre, Sarawak General Hospital  
Institute for Clinical Research  
Ministry of Health Malaysia

Hon Yoon Khee  
Institute for Clinical Research,  
National Institutes of Health,  
Ministry of Health Malaysia

Dr. Lee Keng Yee  
Sector of Ethics and Research Surveillance,  
National Institutes of Health,  
Ministry of Health Malaysia

# Acknowledgement

The author would like to thank the Director General of Health Malaysia for his permission to publish this book. The appreciation also goes to Dato' Dr. Goh Pik Pin (former Director of Institute for Clinical Research), Dr Kalaiarasu Peariasamy (The Director of Institute for Clinical Research) and Dr Alan Fong Yean Yip for their support and guidance. A special appreciation also goes to Prof. Shamsul Azhar Shah for his review and constructive comment on this book, and also to Dr Chew Cheng Hoon for helping us in the process to publish this book. Finally, we would also like to take this opportunity to thank our family and colleagues who are always supporting us in this effort. Thank you so much!

# Synopsis

The original version of most validated questionnaires that are widely used throughout the world will usually be developed in the non-native language. If it is necessary to adopt such a questionnaire for use within a native speaking population, then it will first have to undergo a proper validation process. This book aims to provide a useful guide for describing all the necessary steps of validating a questionnaire. Since different researchers may adopt slightly different approaches for validating a translated questionnaire; hence, it is helpful for us to standardize a list of common requirements for validating the translated questionnaires. In this book, we shall focus on developing a standard checklist which contains 17 items that are recommended to serve as a useful guide for researchers conducting studies involving

questionnaire validation. Therefore, it is recommended for the researchers to make use of this checklist to remind them of all the necessary steps for them to take when they validate a translated questionnaire.

# Abbreviations

ACS: Adolescent Coping Scale

CDI: Children Depression Inventory

DQOL: Diabetes Quality of Life

SF-36: The Short Form Health Survey

QoL: Quality of Life

DASS: Depression Anxiety Stress Scale

SME: subject-matter experts (SME)

SEM: Structural Equation Modeling

SDSCA: Summary Diabetes Self-Care Activities

# Table of Content

Acknowledgement

Synopsis

Abbreviations

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Why validation of an existing questionnaire is of a higher priority than the original development of a new questionnaire?

1.2 Four different types of questionnaires

1.2.1 Questionnaire Type 1: Questionnaire to collect facts or information

1.2.2 Questionnaire Type 2: Questionnaire to assess knowledge level

1.2.3 Questionnaire Type 3: Questionnaire for screening or for a diagnostic assessment

1.2.4 Questionnaire Type 4: Questionnaire to measure latent variables

1.3 A Summary of defining features and characteristics of different types of questionnaires

1.4 How to select the right type of questionnaire to validate?



## Chapter 2: Process for Conducting Questionnaire Validation Studies

### 2.1 Content validity

### 2.2 Translation process

### 2.3 Face Validity

### 2.4 Pilot studies

#### 2.4.1 The sampling procedure and sample selection for a pilot study

### 2.5 Field-testing of a questionnaire

## Chapter 3: Establishment of Reliability & Validity

### 3.1 Reliability testing

### 3.2 Validity testing

#### 3.2.1 Concurrent validity

#### 3.2.2 Predictive validity

#### 3.2.3 Diagnostic assessment

#### 3.2.4 Known-groups comparison

#### 3.2.5 Construct validity

### 3.3 A summary of requirements for conducting reliability and validity tests for a questionnaire validation study

## Chapter 4: Other Issues Pertaining to Reliability and Validity Testing for a Questionnaire Validation Study

### 4.1 Copyright protection

4.2 The practice of combining more than one existing questionnaire to form a new questionnaire

4.3 The practice of combining a validation study and a research study together by using an unvalidated questionnaire that is awaiting validation for a research purpose

4.4 Sample size requirements

4.5 In what circumstances will validation of the questionnaire be deemed unnecessary?

4.6 Types of questionnaires that are probably not suitable for undergoing validation

## Chapter 5: Conclusion

## References

### About the Authors

Dr. Mohamad Adam Bujang

Mr. Hon Yoon Khee

Dr. Lee Keng Yee

### About the Publisher

# List of Tables

Table 3.1: Recommended statistical techniques to be deployed for testing reliability and validity in four different types of questionnaires

Table 5.1: A brief guide of planning and conducting validation for a questionnaire

# List of Figures

Figure 1.1: A conceptual framework for developing the job satisfaction questionnaire (JSQ) derived from Ahmad et al., (2020)

Figure 1.2: Level of difficulty for validating four different types of questionnaires

Figure 2.1: A simplified guide for the validation process of a questionnaire

Figure 2.2: The schematic flow diagram showing the recommended step-by-step procedures for an initial translation and a subsequent validation of a questionnaire

Figure 3.1: Illustration of an actual example of testing for validity based on an assessment of concurrent validity - Results obtained from Jacobson et. al., (1994)

Figure 3.2: Illustration of an actual example of testing for validity based on an assessment of predictive validity (Results obtained from Levy et. al., (1999))

Figure 3.3: The calculation for sensitivity and specificity of a diagnostic test

Figure 3.4: Illustration of an actual example of testing for validity based on known-groups comparison (Part of the results obtained from Tiong et. al., (2018))

Figure 3.5: A visual representation of how an exploratory factor analysis works

Figure 3.6: A visual concept of reliability and validity

# Chapter 1: Introduction

A questionnaire is a universal tool that has commonly been used to collect relevant and intended information from the respondents of a survey. The researchers often use the questionnaire to collect the relevant information from the respondents in order to measure each of the variables or items constituting its domain of interest, and also for addressing the objectives of a research study. Assume that a validated questionnaire was created in a non-native language, such as English, because it was intended for English-speaking respondents. If the original English version of these questionnaires must be administered to a group of non-English-speaking respondents, we must first ensure that the original

questionnaire is subjected to a proper initial translation and subsequent validation process.

It is commonly understood that the validation of a translated questionnaire is a scientific process which involves a translation of a questionnaire from a foreign language into a local (or native) language (i.e. intended for the local population) and will then be followed by demonstrating that the translated questionnaire has an adequate level of both reliability and validity by administering it to a representative sample of the local (or native) population. Therefore, it is not scientifically valid to simply administer the translated version of a questionnaire to the local (or native) respondents because both its validity and reliability have not yet been established. Hence, such a practise should be condoned and therefore must be discouraged.

It is commonly found that the content of the original questionnaire in the foreign language does not suit the local population because many factors can potentially contribute to differing perceptions or understanding of the context. Differences in understanding some contexts may be caused by differences in individual religious beliefs, cultural heritage, lifestyle, social and living environments, and so on. It is commonly found that people from the west will have different family values compared to people from the east, especially those from countries where the majority are Muslim (Bujang et., al., 2010; Omar et. al., 2010). Therefore, it is possible that different people may perceive or understand some words differently.

In other words, it is often possible for incorrect responses to be elicited from the respondents by merely translating the questionnaire and then administering it to a group of respondents who are conversing in a different



language from that of the original questionnaire. This means that it is wasteful to conduct a research study by using a questionnaire that has merely been translated into a different language but has not yet been properly validated because the responses elicited from the respondents will not be valid, which have arisen from the fact that the translated version of the questionnaire has not undergone the proper validation process.

### **1.1 Why validation of an existing questionnaire is of a higher priority than the original development of a new questionnaire?**

Researchers may be interested in comparing their findings across different ethnic groups or countries. Say, for example, a study aims to compare the quality of life among kidney failure patients across different countries. Ideally, it is necessary to administer the same

questionnaire such as The Short Form Health Survey (SF-36) for measuring the quality of life of patients from different countries in order to ensure that the findings obtained from this research study are valid. However, since different countries are usually using different languages, the same SF-36 questionnaire must first be translated into different languages before it is administered to the respondents from all the different countries (Fukuhara et., al., 2003).

When making a comparison between different countries (for example, in a 'cross-country' type of study), it is not recommended for the different studies conducted in different countries to each use a different set of questionnaires that is measuring the same latent variable such as quality of life because it will be difficult to ensure that there is a fair comparison of the quality of life measures. If a majority of existing studies have already

been using the same questionnaire such as the SF36 to measure the quality of life, then it is preferable to validate the same questionnaire SF36 as compared to developing a completely new questionnaire to achieve the same purpose.

Therefore, the decision to validate an existing questionnaire rather than develop a completely new one that measures the same latent variable is ultimately based on the fact that the original version of an existing questionnaire (i) has a sufficiently high level of both reliability and validity, (ii) has been widely used in research, and (iii) has been found to be relevant for the current application or to suit the specific requirements of the study objective(s). Otherwise, the development of a completely new questionnaire may prove to be a more viable option.

In addition, it is often much easier to validate an existing questionnaire than to develop a completely new questionnaire. This is because one of the most difficult steps for conducting a study involving the development of a new questionnaire is to establish a valid new set of questions and then confirm that these questions have a sufficient level of both reliability and validity. To do so, the researchers must have strong background knowledge of the subject matter in order to identify and select an appropriate set of questions when developing a new questionnaire from scratch. On the other hand, if the researchers choose to merely validate the translated version of an existing questionnaire, then the researchers will merely have to translate all the questions from the original questionnaire into a local (or native) language, which will then be followed by the process of determining both the reliability and validity of the translated

questionnaire by administering it to a sample of the local (or native) population (Yunus et., al., 2013; Bujang et., al., 2016).

## **1.2 Four different types of questionnaires**

This book has categorized four major types of questionnaires to further elaborate on the above-mentioned point. It is important to recognise the existence of numerous types of questionnaires because each type of questionnaire may require a different approach to determining its reliability and validity (Bujang & Tan-Hui, 2022).

### ***1.2.1 Questionnaire Type 1: Questionnaire to collect facts or information***

The first type of questionnaire is designed to collect facts or information from the respondents such as patients

or healthcare professionals, which are usually quantifiable facts about an individual or subject. Each item or question will be represented as an independent variable rather than being part of the collection of items or questions relating to a specific latent variable. Examples of these items include specific information related to the patient's socio-demographic characteristics such as gender, ethnicity, comorbidities, smoking history, treatment outcomes, etc. Researchers will usually use a straightforward approach to seek this type of information without engaging in an interrogative tone. This type of questionnaire is usually meant for record-keeping purposes which can also serve as future reference. However, the data obtained can also be used for research purposes whenever necessary (Bujang & Tan-Hui, 2022).

### ***1.2.2 Questionnaire Type 2: Questionnaire to assess knowledge level***

The second type of questionnaire is designed to assess the knowledge level of an individual respondent. An example of this type of questionnaire is best illustrated by an examination paper which contains a set of questions to assess the knowledge level of a specific study respondent population on a specific subject. All the questions in this type of questionnaire are directly related to facts about a particular subject matter. Usually, all the questions are worded by using an interrogative approach such as “what”, “when”, “who”, “why” and “how”. The questionnaire will then have a specific formula to compute an overall score which reflects an assessment of the knowledge level of an individual. Unlike the first type of questionnaire, the answers provided by the respondents are usually not for record-keeping purposes but merely for assessing the

knowledge level of the respondents in order to determine the rank of each respondent based on his/her knowledge score, and such ordinal data can subsequently be used for research purposes (Bujang & Tan-Hui, 2022).

### ***1.2.3 Questionnaire Type 3: Questionnaire for screening or for a diagnostic assessment***

The third type of questionnaire is designed for screening or diagnostic assessment. This type of questionnaire is also related to specific facts or knowledge about a subject because the parameter is often defined by two distinct criteria, often in a mutually exclusive or binary form such as ‘with disease’ or ‘without disease’. All the questions in this type of questionnaire will directly correspond to the outcome, and hence this type of questionnaire will aim to screen for a probable or likely association between the variable and the outcome. Hence,



all the questions in the questionnaire will have to be based on the findings obtained from the textbooks and existing literature. The responses elicited from the respondents by using this type of questionnaire will be transformed into a score and a pre-determined cut-off score is used to distinguish between the state of having the disease (i.e. 'with disease') or not having the disease (i.e. 'without disease') (Bujang & Tan-Hui, 2022).

The process of screening patients for a disease or performing diagnostic procedures on patients occur frequently in clinical medicine. These patients can often be diagnosed by using methods like structured interviews such as those conducted by psychiatrists for diagnosing depression. While in other instances, the diagnosis can also be confirmed by using a specialized machine. However, it is often too time-consuming and/or costly to employ such diagnostic procedures as a structured

interview or performing a diagnostic test by using a specialized machine, especially when there are already too many patients waiting to be screened.

Therefore, a questionnaire that is specifically developed for diagnostic research can often be used for screening purposes, even though it may not be robust enough to confirm a diagnosis. This can potentially save the cost of healthcare for the patients since only those patients who exhibit a higher cut-off score as determined by the screening questionnaire will be required to undergo further diagnostic tests for confirmation. Hence, this type of questionnaire serves an important function by providing a necessary complement to the existing diagnostic procedures (Yunus et. Al., 2013).

#### *1.2.4 Questionnaire Type 4: Questionnaire to measure latent variables*

The fourth and final type of questionnaire is designed to measure one or more latent variables. Since these latent variables cannot be measured or observed directly (such as by using any machine or measuring device). Therefore, they will have to be inferred from other variables that can be measured or observed directly. In the medical field, a few examples of such questionnaires that fall under this category are Medical Outcome Study Short-form36 (MOS SF36), Depression Anxiety Stress Scale (DASS) and Summary Diabetes Self-care Activities (SDSCA) (Ware & Sherbourne, 1992; Lovibond & Lovibond, 1995; Toobert & Glasgow, 1994) which are widely used for measuring latent variables. An example is shown in Figure 1.1 where the latent variable of “Job Satisfaction” is measured by seven other latent variables

namely “Leadership”, “Reward & Recognition”, “Empowerment”, “Training & Development”, “Working Hours”, “Communication” and “Working Conditions” (Ahmad et. al., 2020).

A latent variable is a relatively pure measure of a construct that is extracted from the observed variables, by capturing the common or shared variance among multiple observable variables or indicators of a construct. Since a latent variable is usually created by factor analytic methods; hence, it will generally represent “shared” variance, or the degree to which variables “move” together. The common factors model has postulated that only those variables without any correlation do not result in the formation of a latent construct, which means that the construct validity of a latent variable shall depend on whether (or not) it has the capacity to represent “shared”

variance, or the degree to which variables “move” together (Tabachnick & Fidell, 2001).

Measures of quality of life (QoL), depression and satisfaction are some common examples of latent variables. Furthermore, a broad domain such as QoL can be subdivided into several sub-domains that measure specific aspects of QoL, such as physical aspects of QoL, emotional aspects of QoL, family aspects of QoL, and so on, all of which are considered latent variables. At least two questions or items are needed to adequately measure a domain of interest and some research studies may require at least ten (or more) questions to measure a specific domain of interest (The DCCT Research Group, 1988; Toobert & Glasgow, 1994).

In order to measure a specific latent variable by using a questionnaire, the researcher will need to develop a study instrument or a questionnaire to measure the

observed variable and then infer mathematically the unseen or latent variable from the observed variable. It is well-known that the questionnaire which measures the latent variable(s) is often the most difficult to develop and validate because unlike the first and second types of questionnaires, it is more difficult to measure the latent variable(s) since they are interpreted in a purely subjective way, i.e. different respondents can possibly perceive the latent variable(s) differently (Bujang & Tan-Hui, 2022).



Figure 1.1: A conceptual framework for developing the job satisfaction questionnaire (JSQ) derived from Ahmad et al., (2020)

### 1.3 A Summary of defining features and characteristics of different types of questionnaires

From the above, it is now apparent that since each of the four types of questionnaires serves a different purpose. Hence, the approach to validating each of these

four types of questionnaires will be slightly different. The summary characteristics of the four different types of questionnaires and their relationship with both their reliability and validity are visually represented in Figure 1.2.

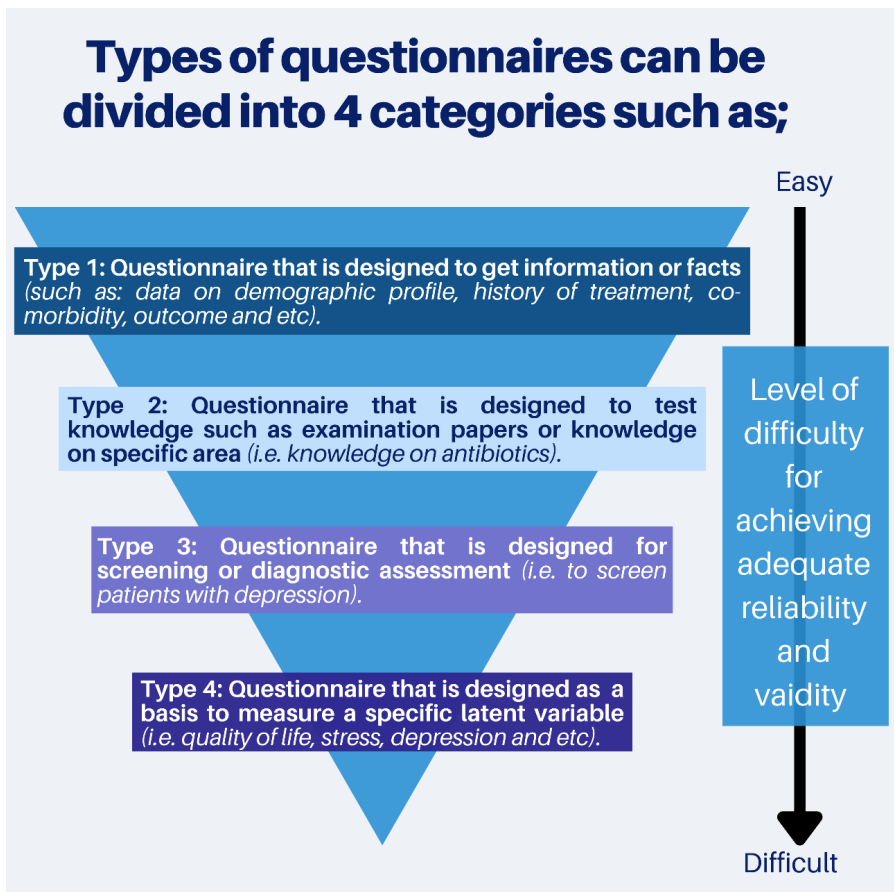




Figure 1.2: Level of difficulty for validating four different types of questionnaires

#### **1.4 How to select the right type of questionnaire to validate?**

A researcher will frequently use a previously validated and published questionnaire as a basis of reference for adapting it to the specific purpose of a new study by tailoring it to fulfill the specific objectives of the new study in order to save time and resources. To achieve this, a researcher will have to validate a previously validated questionnaire by ensuring that it also serves as a valid measure for the specific group of respondents who are being examined in the new study. However, a researcher may very often have to choose the most suitable questionnaire to validate if he/she has found that there are more than one existing standardized and validated

questionnaire available for adaptation to a specific purpose of this new study. The choice of the most suitable questionnaire to validate is often made based on whether (or not) it has been designed to address the specific objective(s) of the new study, which can be determined by a preliminary feasibility study.

Apart from the above, a researcher may also need to conduct a preliminary literature review to gain familiarity with the research area in order to determine what type of data will be collected by the questionnaire and also what new information will be generated by this new study. A review of the relevant literature shall also enable a researcher to gather all relevant information from different review articles on the subject, which will enable him/her to make a proper comparison between all the different validated and published questionnaires that are currently available. This can in turn serve as a useful guide

for a researcher in the selection of the most suitable existing questionnaire to undergo the subsequent validation process, in order to adapt it for a specific purpose of this new study.

Many factors should be considered when selecting the most suitable questionnaire to validate for adapting it for a specific purpose of the new study. These include:

- (i) whether it is a new or an old questionnaire,
- (ii) cost of acquisition of the questionnaire (i.e. whether it is available for free of charge or not),
- (iii) the number of items or questions in the questionnaire, and
- (iv) the total citation counts of the questionnaire (i.e. whether the questionnaire has been widely cited or not).

- (v) overall feedback pertaining to the applicability and versatility of a questionnaire by the fraternity of professionals and scholars

However, it is often up to the individual researcher to choose the most suitable questionnaire to validate because this choice will very much depend on each individual researcher's list of priorities. In general, the most suitable questionnaire to be chosen for validation is the one containing the highest number of appropriately-worded items along with having high levels of both validity and reliability. This means that it has already been developed in such a way that it will measure what it intends to measure all the variables within the same context as that of the intended respondents (such as the time, country of origin and its cultural setting). This can usually be determined by locating the source of publication of the questionnaire; in that if it comes from a

peer-reviewed journal, then it is more likely for it to fulfil all those requirements which are mentioned above.

Another reason why a questionnaire with high levels of both reliability and validity is most suited for undergoing the validation process is that it is often much easier to validate a well-developed questionnaire because the methodological basis for a questionnaire validation study will usually be directly adopted from the core methodologic approach for the initial development of the original questionnaire. Therefore, if the original questionnaire has been shown to have an adequate level of both reliability and validity, then it is more likely for it to also be reliable and valid when it has been administered to the other respondent populations provided that the translation into another language has been accurately done (Toobert & Glasgow, 1994; Bujang et. al., 2016).

Besides that, a researcher may also be concerned about the relevance and suitability of the questionnaire to be administered to the respondents in the future. This is because how the current respondents perceive the situation can be profoundly different from how the former respondents had perceived the same situation in the past. For example, the establishment of a wide existing social networking service has made it possible for most of the current respondents to perceive interpersonal or social relationships differently from the former respondents in the past. For example, in the past, people were used to keeping contact with family members by giving them a phone call, visiting them at their homes, etc. as a way to foster a strong family relationship.

On the other hand, in this new age of digitalization, it has become a norm that nowadays there are so many modern social networking platforms and apps to enable

family members to keep in touch with each other. These include communication via a digital platform such as sending and receiving messages by using emails and/or social networking services. Hence, the advances in digitalization can have a major impact on how people in the past had perceived differently the various ways of maintaining interpersonal or social relationships with each other (Kraut et. al., 1988).

In addition, the cost of the acquisition of a validated and published questionnaire can also be an important consideration, especially for researchers from developing countries. The purpose of a questionnaire validation study is to determine whether (or not) a validated questionnaire can be re-used in a local respondent population within a local setting. Since many researchers may have difficulty in paying for the access to a validated and published questionnaire if they do not receive any financial support,

therefore those questionnaires which are available free-of-charge such as Depression Anxiety Stress Scale (DASS) have now become one of the most popular questionnaires for use widely throughout the world and until now (i.e. by the end of year 2019), it has already been cited more than 8000 times by the end of year 2021 (Lovibond & Lovibond, 1995).

In summary, there are several important considerations that should be considered when choosing the right questionnaire to be validated. Based on these considerations, it is therefore no surprise that the most suitable questionnaires to undergo validation will be the ones that are measuring important latent variables which also shall desirably exhibit evidence of excellent validity and reliability.



# **Chapter 2: Process for Conducting Questionnaire Validation Studies**

The methodological approach to validate a questionnaire shall be the same or almost the same as that for the development of the original questionnaire because a researcher will need to show that the newly-validated questionnaire has a comparable level of reliability and validity as that of the original questionnaire. Therefore, it is compulsory for the researchers to obtain a copy of the questionnaire construction manual which describes how the original questionnaire had been developed. This manual is sometimes published in a technical report, or more often in a published article. However, a researcher can also opt to furnish further evidence to demonstrate

both the reliability and validity of the original questionnaire. Hence, as illustrated in Figure 2.1, this present study aims to propose a useful guide for conducting the validation process of a questionnaire (specifically for the purpose of validating a translated version of an original questionnaire).

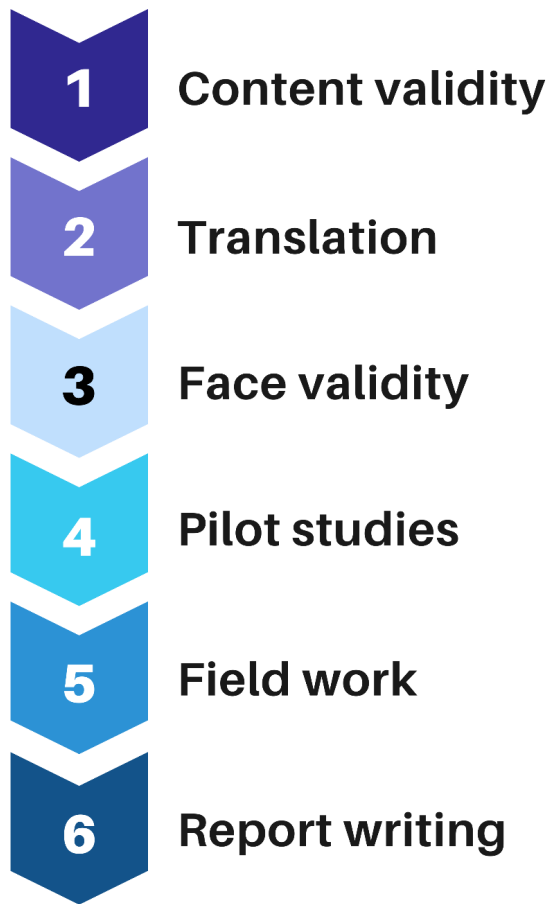


Figure 2.1: A simplified guide for the validation process of a questionnaire

## **2.1 Content validity**

Content validity is an assessment of the degree to which each item or question is able to measure what it is supposed to measure. Therefore, the content validity of a questionnaire should be assessed by a subject matter expert within a particular field and so it will be evaluated by using a qualitative approach (Lawshe, 1975). For example, if the questionnaire is related to quality of life for diabetic patients, then an endocrinologist or family medicine specialist will be a suitable subject matter expert to determine its content validity. It is commonly accepted that by the time the questionnaire has been accepted for publication, the content of all the items in the questionnaire shall have already been determined to be reasonably accurate in measuring what they are supposed to measure.

Moreover, it is often advisable to re-assess the content validity of a questionnaire to determine whether (or not) the questionnaire is specifically tailored for the current objective(s) of the study. Additional precautions should also be taken when re-assessing the content validity of a questionnaire if it contains specific questions which are related to perceptions, feelings, cultural beliefs, lifestyle choices or any other aspects of human behavior because different humans from different language backgrounds often perceive things differently. Such differences can often be due to many possible causes such as the differences in their cultural background, environment and lifestyle (Nisbett, 2003). Therefore, if an original questionnaire has been translated into another local (or native) language, it is therefore recommended to re-assess its content validity to ensure that the questions are specifically tailored for the intended respondent

population who come from a different language background.

Upon re-assessing the content validity of the translated version of a questionnaire, a researcher may make one of the following three decisions:

- the researcher may decide all items in the original questionnaire are relevant and so it will be adequate to retain the original version of the questionnaire. This means that the original version of the questionnaire is exactly the same as its translated version (Yunus et. al., 2013; Bujang et. al., 2016).
- the researcher may decide to add and/or omit and/or one or more question(s). This means that the translated version of the questionnaire will be different from its original version (Bujang et. al., 2018).

- the researcher may decide that most of the items are relevant except for a few items that should be rephrased. This means that the content of the translated version of the questionnaire will be slightly different from its original version (Jalaludin et. al., 2013).

Nevertheless, it is often strongly recommended for a researcher to always try as much as possible to retain the original version of the questionnaire because by doing so, the researcher will be able to retain the original scoring method (if available) so that the responses elicited by the original questionnaire are made directly comparable with those of other studies which have used the same questionnaire, albeit in a different language (since the scoring method found in the original questionnaire will have already been standardized and validated). For

example, SF-36 has been validated in many different languages and the results are made directly comparable between different ethnic groups and countries (Fukuhara et. al., 2003). This is only possible because the original version of the questionnaire had been retained when it had been translated into many different languages. This is very important because if the original SF-36 questionnaire were not used for the translation process, which means that a different version of the same questionnaire had been used for the translation; then it would not be possible to ensure that the responses elicited by the original questionnaire are made directly comparable with those of its translated version (in all the other languages).

The following example provides an excellent illustration of how the content validity of a questionnaire should be assessed.



Let's say we are going to design a new questionnaire to be filled by children and adolescents whose ages are between 15 to 18 years. This questionnaire aims to measure to what extent their parents emphasize the importance of imparting the value of knowledge acquisition to their children. Several of the items in this questionnaire are as follows:

- Question 1: My parents will always bring me to the library.
- Question 2: My parents will allocate a specific time within a day or at least several days within a week for supervising me to read.
- Question 3: My parents will send me to attend tuition classes.

At first glance, although it seems that the above three questions will serve as an indicator of whether (or not) the parents have emphasized the importance of

imparting the value of knowledge acquisition to their children; however, there could also be other possible reasons which explain why some of these questions can no longer be relevant for indicating whether (or not) the parents have emphasized the importance of knowledge acquisition among their children. For example, it is possible for logistical problems such as the long distance between home and the library to affect the parents' decision of whether (or not) to send their children to the library. Similarly, it is also possible for only those parents who can afford tuition fees to be able to send their children to attend tuition classes. This shows that whether (or not) the parents are sending children to the library or to attend tuition classes does not always provide a fair and accurate assessment of the parents' emphasis on the importance of imparting the value of knowledge acquisition to their children.

The above proposition also illustrates the point that Question 2 is an example of a good question in a questionnaire. This is because, irrespective of logistical problems or whether (or not) the parents who can afford tuition fees, spending a specific time within a day for the parents to supervise their children to read will always be doable. From these examples, it is now clear that certain questions for a domain subscale in a questionnaire may not be applicable to a subset of a local respondent population because they have different baseline characteristics. Therefore, a researcher will need to first examine the differing baseline characteristics that may exist within the local respondent population in order to make allowances for them when he/she develops the items in a questionnaire. This can be achieved by rephrasing or modifying the original item in the questionnaire.

However, since a researcher usually does not know whether (or not) those new items which have been rephrased/modified are actually better than the original ones; therefore, the best way is to obviate the need to substitute the original items with the rephrased/modified ones. Instead, it would be far better to initially conduct the survey by combining these original items from the original questionnaire with a list of newly revised questions (which are formed by rephrasing and/or modifying the original questions). Then, a researcher shall have to decide whether (or not) it is desirable or necessary to substitute the original items in the original questionnaire with the rephrased/modified ones by performing the necessary statistical tests (Bujang et., al., 2016).

This brings us to the next step, in which a researcher must decide whether to remove the item(s) from the questionnaire, add more item(s), or re-insert the

original item(s). Such a decision will have to be made during the final phase of assessing the validity of a questionnaire. A major limitation of using a modified version of the questionnaire is that the results of a study using a modified questionnaire may not be directly comparable to those of other studies using the original questionnaire. This is because the modified questionnaire may contain a list of different items or domains from the original questionnaire, and each of the items will have a different scoring method for the designation of its domains in the modified questionnaire.

Therefore, it is often recommended for researchers to retain the same number of items and/or constructs within the domain subscales as far as possible. If this does not prove feasible, then the last resort would be to derive a modified version of the original questionnaire (Omar et. al., 2011). However, if it is found that there are too many

items that require modification upon performing an initial evaluation of the content validity of the original questionnaire; then it will usually be recommended for the researchers to choose another suitable questionnaire if available. Otherwise, the only viable option remains for a researcher might be to develop a new questionnaire from scratch.

## **2.2 Translation process**

Translation of the original questionnaire is an important step for establishing the validity of a questionnaire. This process involves initially translating the questionnaire from the foreign language into a local language and at the same time matching the semantic feature of each item or question as closely as possible with that of the original questionnaire (Wild et al., 2005). Past research work in this area revealed that there were many

different approaches to undertake the translation of study instruments such as a questionnaire. One of the best approaches is to have two arms working independently for performing an initial forward translation and then a backward translation (Omar et al., 2011; Jalaludin et al., 2013; Yunus et al., 2013; Bujang et al., 2016).

The flowchart of the translation process is illustrated in Figure 2.2. The first arm consists of the subject matter experts (SME) from a particular field. For example, if the questionnaire is related to diabetes mellitus, then the subject matter expert will be either an endocrinologist or a family medicine specialist. Apart from having the expertise in the field of the subject matter, he/she should also be able to perform translation from the original language into a local language, meaning they shall be proficient with both languages.

The second arm will usually be represented by several language experts, particularly those who understand at least both the foreign language of the original questionnaire and also the local language of the translated version of the original questionnaire. Let's assume that the translated questionnaire to be validated had initially been developed in English. So, the ideal candidate would be a language expert such as a certified English linguistic officer or a certified English teacher (Omar et al., 2011). However, if the English language is a second language in the country, then the other professionals such as lecturers or medical doctors, etc. can also be regarded as an acceptable translator for this purpose as long as they are well versed in both English and the local language (Jalaludin et al., 2013).

There are no strict criteria for selecting the right candidates for the second arm because the individuals will



only have to be well versed in both languages, which means that the selection of the panel of subject matter experts (SMEs) for the first arm of the translation process will be the core determinant of success/failure in the translation of the questionnaire. Hence, it is crucial to select the most suitable candidate for the first arm of the translation process to ensure a valid and accurate translation of the core subject content of the questionnaire from the original language into a local (or native) language has been done.

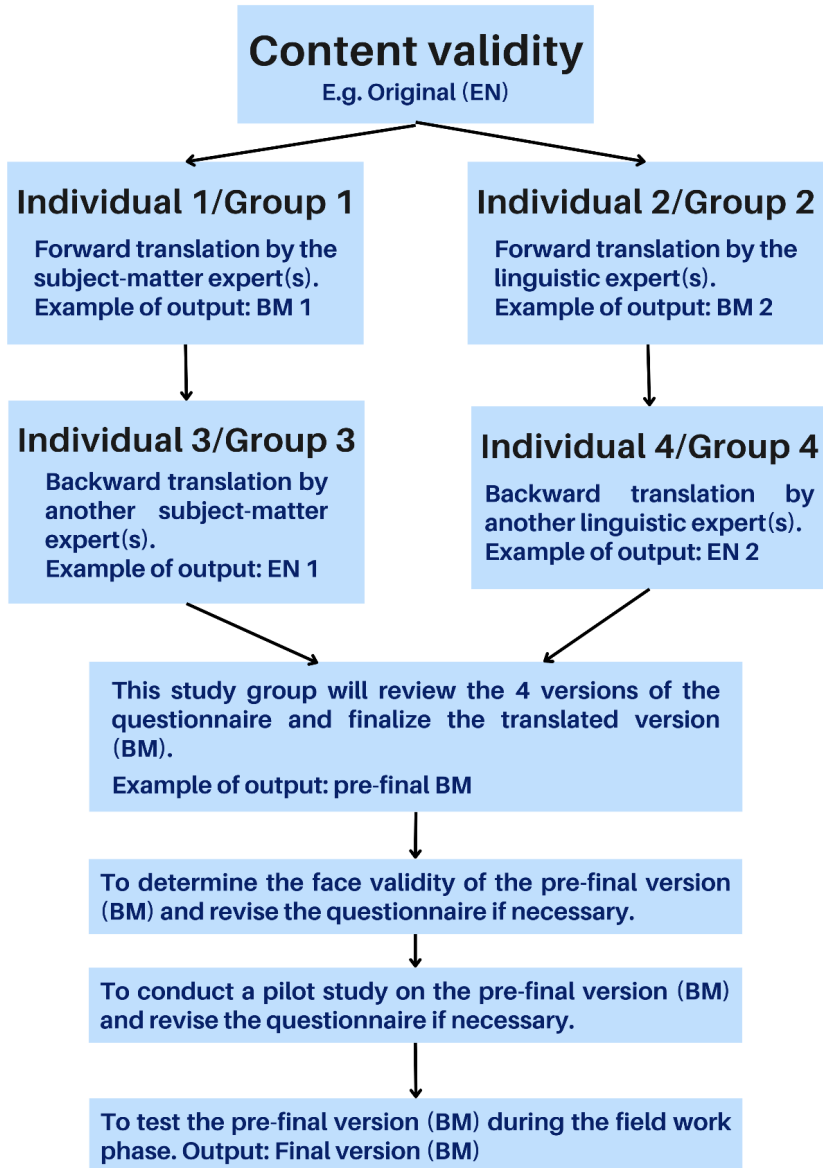


Figure 2.2: The schematic flow diagram showing the recommended step-by-step procedures for an initial

## translation and a subsequent validation of a translated questionnaire

The first step in the translation process is a forward translation, which involves translating the original questionnaire from the foreign language into a local (or native) language. This initial step of the translation process is very important because the translators will also have to take into account any cross-cultural differences between the two languages (Guillemin et al., 1993). The process can be done by an individual or in a group preferably involving 2-3 persons. The expected outcome is to elicit two versions of forward translation from the two arms (i.e. a panel of subject matter experts and a group of linguistic professionals). Then, a backward translation has to be independently conducted also by the same two arms (i.e. a panel of subject matter experts and a group of linguistic professionals). To minimize the chance of inadvertently

introducing bias into the backward translation process (which may occur due to the same people performing both the forward and backward translation), an additional and independent individual or group of individuals from both arms who are blinded from the original questionnaire who shall then be assigned to translate the questionnaire [which had already been translated into a local (or native) language] back to the foreign language.

Hence, a researcher will receive a total of four translated versions of questionnaires at the end of the both forward and backward translation processes. This is because the subject matter experts (i.e. SMEs) have produced two different versions of the translated questionnaire, one obtained from forward translation and another obtained from backward translation. Likewise, the linguistic experts have also produced two different versions of the translated questionnaire (i.e. one obtained

from forward translation and another obtained from backward translation).

The purpose of obtaining two different translated versions of the same questionnaire (for both forward and backward translations) is to elicit a balanced view of the content of the questionnaire by the two groups of translators each from a distinct background. This is because while the subject matter experts are able to accurately translate the questionnaire by adopting appropriate subject-specific terminology; however, it is also possible for them to overly use technical jargon in the translated version of the questionnaire, which makes it more difficult for a layperson to understand.

On the other hand, it will also not be adequate to merely obtain a translation by the linguistic experts because they do not fully understand the context in which the subject matter of the questionnaire has been derived.

As a result, the linguistic experts by themselves can possibly have misconstrued the original questionnaire and then yielded a translated version which does not reflect an accurate representation of the original questionnaire. Then, the researcher will carefully review each of the four sets of questionnaires (e.g. BM1, BM2, BI1 and BI2) in order to select a pre-final version of the translated questionnaire (e.g. for BM version) to be used for subsequent research purposes, which can realistically provide a closest match to the intended meaning of the original questionnaire.

The process of selecting the most ideal items to be included in the pre-final version usually involves a qualitative process. Those items which will be included in this pre-final version are usually generated from the BM1 version (a forward translation by SME) only, or from the BM2 version (a forward translation by a linguistic expert)

only, or from a combination between BM1 and BM2 versions. To expedite this selection process, the principal investigator or the leader in the study group will carefully review all the items in the two (BM1 and BM2) different translated versions of questionnaires and then decide which items should be included in this pre-final version. Thus, the translation process produces a pre-final version of the translated questionnaire (e.g., a pre-final BM version). This pre-final version of the translated questionnaire will then be re-assessed in terms of its face validity and also to undergo a pilot study before finalizing it through an assessment of its validity during the fieldwork phase.

The existing literature in this field has described a variety of translation approaches for translating a questionnaire. Some questionnaire studies only rely on a single arm for translation, which means that the translation

was done usually by the subject matter experts and not by the linguistic experts (Abdullah et al., 2006; Ting et al., 2007; Lim et al., 2018). It is rare for a validation study involving a questionnaire to use only forward translation because most researchers are well aware of the standard two-way translation process for a questionnaire validation study. Nevertheless, it is still possible for a single forward translation process which has been performed by any professionals who are able to understand bilingual languages to yield an accurately translated version of the original questionnaire.

The difficulties commonly encountered in the translation process are highly dependent on the level of language competence acquired by the translators in order for them to understand the true semantic representation of the original items in the questionnaire. Therefore, a single forward translation process which has been



performed by only the subject matter experts (i.e. SMEs) can very often be adequate for most of the circumstances, particularly for a questionnaire which consists of very straightforward questions; for example, a questionnaire that is designed to collect a list of simple facts from the respondents. Nevertheless, the incorporation of both forward and backward translation processes for the validation of a questionnaire has now become a conventionally-accepted standard practice for most questionnaire validation studies.

### **2.3 Face Validity**

The term face validity refers to the extent to which a test appears to measure what it claims to measure based on face value (i.e. the apparent worth or implication of something). In questionnaire validation studies, face validity is referring to the extent to which a questionnaire is subjectively viewed by potential respondents to be

adequate for them to understand the question and then to provide the correct feedback in response to the question because it is measuring the intended domain of interest (Frederick & Forzano, 2012). As the assessment of the face validity of a questionnaire often involves a qualitative process; hence, it is important to elicit the correct feedback from the respondents. On the other hand, there are some questionnaire validation studies which attempt to determine the face validity of a questionnaire by quantifying the feedback obtained from respondents, such as the design of Likert scale survey questions in order for the respondents to provide their feedback via a Likert scale (Bujang et. al., 2016; Bujang et. al., 2017), which are designed to measure a respondent's opinion or attitude towards a given topic.

For the purpose of assessment of face validity, a cognitive debriefing process is deemed very important.

Cognitive debriefing is a process in which a questionnaire or other assessment will be tested among representatives of the target population and target language group. The questionnaire is administered to the participating respondents. Afterwards, they are personally interviewed about the scope of the study, instructions, the response format itself and all the instrument items. Participants are asked questions such as what they thought the item was asking if they can reformulate the item in their own words, how they arrived at their answer, what they associated with particular words if an item was clear or unclear and if there were any words they did not understand or found unacceptable. If there are alternative word choices, participants can provide suggestions so they can understand the questions exactly from what is intended by the researchers.

Findings are compiled together for a final report, which includes relevant details about the participants, as well as final recommendations on improving the instrument. The process will ensure that the translated version of the original questionnaire will ideally reflect the meaning of the original questionnaire itself. Thus, cognitive debriefing is a very important step in linguistic validation.

When validating a questionnaire, it is not sufficient to merely elicit the necessary feedback from the subject matter expert(s) alone because the questionnaire will very likely be answered by a respondent who is not a subject matter expert. So in face validity, it is necessary to ensure that all potential respondents of a questionnaire are able to understand the questions and then provide correct and valid feedback to these questions. Nonetheless, it might also not be necessary to make all the amendments

recommended by the respondents during an assessment of the face validity of a questionnaire because face validity is usually assessed by less than 10 respondents in a small-scale study. This is because the feedback provided by a small number of respondents might not mimic the feedback provided by the majority of the respondents. Therefore, the researchers will first need to analyze the feedback provided by all the respondents during an assessment of the face validity of a questionnaire, and then to discuss with the study group for reaching a general consensus in deciding whether it shall be recommended to revise an item in the questionnaire or not (Jalaludin et. al., 2013).

In other instances, a researcher may enlist the assistance of a bilingual respondent who is able to understand both the original language of the questionnaire and also the local (or native) language of the translated

questionnaire (Jalaludin et. al., 2013). This is to enable the respondent to answer the questionnaire in both languages at two different time-points so that the researcher will be able to assess the level of agreement of the responses provided by the bilingual respondent at two different time-points. By doing so, it is then possible to determine whether (or not) the translated questionnaire (i.e. which is translated into a local or native language) has produced a semantic content which matches exactly with that of the original questionnaire (in the foreign language). This can be preplanned if the study group has intended to incorporate it into the study plan. However, some researchers may feel that this step is not necessary because a proper translation process had already been performed earlier.

## **2.4 Pilot studies**

The first-step towards conducting a primary research study is usually to conduct an initial pilot study, which is a small-scale preliminary study. This aims to evaluate feasibility of the study in answering the study objectives before deciding to conduct the actual full-scale study. For a questionnaire validation study, a researcher will conduct a pilot study to assess the preliminary feasibility of a questionnaire for determining whether (or not) it has been designed to collect all relevant information that are originally intended to be collected by the researchers (Hulley, 2007).

Unlike face validity, a pilot study for a questionnaire is considered as one type of quantitative study which involves a quantitative analysis of the feedback elicited by the respondents. This explains why a pilot study for a questionnaire will usually require a larger sample of at

least about 20 to 50 respondents as compared to that for the assessment of face validity of a questionnaire (which typically requires only 10 respondents or less).

A validated questionnaire should have adequate levels of both reliability and validity. In a questionnaire validation study, both the properties of reliability and validity are interrelated because a low level of reliability will place a limit on the overall validity of a test, and vice versa (which means that a low level of validity will also manifest itself as unreliable responses that can be impacted considerably by many other random or systematic factors to which they should be robust) (Murphy & Davidshofer, 2005). The reliability of the questionnaire will usually be pre-examined during the pilot study although it will be assessed again in the fieldwork phase. An appropriate statistical test should also be selected before conducting a pilot study for a



questionnaire, which is particularly important for the determination of the reliability of the questionnaire.

It is usually not possible to assess the validity of the questionnaire by performing a statistical test for a pilot study because its small sample size will most likely not be adequate for this purpose. Therefore, the validity of the questionnaire will usually be assessed in the fieldwork phase. It is noteworthy to mention that having a high level of reliability cannot guarantee that the validity is also high. However, one thing for sure is that a questionnaire must have a high level of reliability in order for it to be valid. This means that if a pilot study has found that a questionnaire has low reliability, then it will be necessary to first revise the content of a questionnaire before subjecting it to a field-test for its validation. Therefore, the findings obtained from statistical tests for the determination of reliability of a questionnaire during the

pilot study shall serve to inform a researcher whether it is necessary to revise the content of a questionnaire first, before conducting a fieldwork by using the same questionnaire in order to determine its validity.

#### ***2.4.1 The sampling procedure and sample selection for a pilot study***

Despite having a small sample size, the pilot study for a questionnaire validation study should ideally be conducted by using a representative sample of the respondents with a similar distribution of the socio-demographic characteristics as that of the intended respondent population. For example, let's say a study aims to evaluate the quality of life (QoL) among patients with type 2 diabetes mellitus. From a total of at least 30 respondents for a pilot study on a questionnaire, a researcher may have to ensure that the sample shall have

an equal distribution of the disease characteristics of all the respondents such as (i) newly-diagnosed type 2 diabetes mellitus, (ii) long-standing type 2 diabetes mellitus, (iii) poorly-controlled type 2 diabetes mellitus, (iv) type 2 diabetes mellitus with long-term complications, etc. From a wide variety of respondents who exhibit a broad spectrum of widely-differing disease characteristics, a researcher should then be able to determine whether (or not) an objective and accurate pattern of observations have emerged from the data collected using the sample by performing the necessary statistical tests.

Nonetheless, although the sample size of a pilot study is small, it should ideally still be able to generate the data that can accurately depict the true nature of the relationship between ‘effect’ and ‘outcome’; for example, the healthy respondents will generally show a higher QoL score than the unhealthy respondents. This poses a major

concern to a researcher if he/she is using a questionnaire to collect data from only the healthy respondents for a pilot study, since it will then not be possible for this questionnaire to distinguish between the differing responses provided by the respondents due to the variations that may exist in their health conditions.

In this situation (such as the different patterns of QoL which are reported by patients with different characteristics), it is often not necessary to report the p-value obtained from this pilot study for the determination of the association between 'effect' and 'outcome'. Instead, it is more important to report the results obtained from performing statistical tests (such as the p-value) when assessing the reliability of this questionnaire because the purpose of a pilot study is to determine the reliability of a questionnaire, and not to infer the results from a pilot study and then extrapolate

them to the whole population, as in the determination of the association between 'effect' and 'outcome'. This is because the purpose of a pilot study is not to infer the results from a pilot study and then extrapolate them to the whole population; instead, the aim of conducting a pilot study in this instance is to determine the reliability of this questionnaire.

A non-probability sampling method (especially convenience sampling) is usually the preferred sampling method for use in a pilot study, which aims to ensure that a small group of respondents who exhibit a list of target characteristics can be recruited for a pilot study (since the pilot study will usually have a limited sample size). However, a probability sampling method (such as simple random sampling or stratified sampling method) is also possible if a researcher has the capacity to recruit potential respondents by using this sampling method.

Nevertheless, a probability sampling method is often more popular because it can minimize any potential sources of bias that may occur in the sampling process. This is because a non-probability sampling method such as convenience sampling will occasionally fail to ensure that an even distribution of all the respondents' baseline characteristics will be sampled, since these characteristics may not be evident during the initial sampling process.

The above description provides a brief summary of the pros and cons of the selection of a suitable sampling method to obtain a representative sample for a pilot study. An important point to take note of is that it is always necessary to recruit a sample of respondents for a pilot study who shall exhibit baseline characteristics which are representative of the entire population; and also to ensure that this sample size is sufficiently large of at least 20 to 50

subjects to be feasible for conducting the subsequent statistical analysis.

*Key messages:*

- *The determination of reliability of a questionnaire will initially be performed during a pilot study by conducting the required statistical tests using a small sample size.*
- *Subsequently, the determination of reliability of a questionnaire will be performed again in the fieldwork by conducting the same statistical tests by using a larger sample size.*

## **2.5 Field-testing of a questionnaire**

After the translated version of the questionnaire has been finalized, it shall have to be subjected to a further

assessment of its reliability and validity on a larger scale. This can be done by conducting a fieldwork by using the translated questionnaire. However, it is necessary to bear in mind that the validation of a translated questionnaire during the fieldwork phase will have to adopt the same research design and methodologic approach as that for the development of the original questionnaire, including:

- (i) the determination of inclusion and exclusion criteria for recruiting the respondents,
- (ii) the research design and methodology approach (i.e. selection of a sampling method and determination of the minimum sample size), and
- (iii) the proposal of its statistical analysis plan.

Therefore, the researcher should have read the questionnaire construction manual of the original questionnaire in order to ensure that



the design features and methodologic approach of the translated questionnaire are sufficiently similar to those of the original questionnaire, before conducting a fieldwork to determine the validity of the translated questionnaire. This is because fieldwork can often be very costly and labor-intensive, and so a researcher will usually do it only once.

Any differences in the research design and/or research methodology between the original questionnaire and the translated questionnaire will usually result in different results obtained for the assessment of validity of both questionnaires. For example, although exploratory factor analysis which is used to test for construct validity of a questionnaire may produce similar results for both the original questionnaire and its translated version; however,

this same exploratory factor analysis can also produce different results for the assessment of construct validity of the two questionnaires if the original questionnaire has adopted a totally different research methodology from its translated version. For example, Children Depression Inventory is a questionnaire with five major domains of interest which had previously been validated by a sample of respondents who were recruited from children attending a secondary school (Kovacs, 1992). However, another study which had recruited a sample of respondents who were recruited from the child's psychiatric clinic had identified several other new domains in the same questionnaire (Tan et. al., 2013).

Despite all the above restrictions, there are certain modifications which are still permitted to be made during the development of a research methodology for a questionnaire. For example, a researcher may modify an

existing questionnaire to further substantiate the evidence of its reliability and validity. An example is best illustrated using the Diabetes Quality of Life (DQOL) questionnaire, which contains three domains that were previously validated through an assessment of both its content validity and concurrent validity by determining the association between the three domains. In order to provide further evidence of its validity, the researchers have subsequently attempted to establish an additional degree of association of the three domains in the questionnaire with another related item such as the diabetic complications.

Meanwhile, several other researchers have also conducted a second round of assessment of concurrent validity by using a different quality of life questionnaire such as SF36 (Jacobson et. al., 1994), in order to ensure that validation results obtained from the new

questionnaire will actually correspond to those of a previously-validated questionnaire. From the above, we can see that researchers are allowed to make the necessary modifications in the research methodology for the development and validation of a questionnaire as long as they can provide a strong and valid justification for them.

*Key messages:*

- *In the fieldwork phase, a researcher will re-evaluate the reliability of the translated questionnaire and then test for its validity.*
- *To prepare for the fieldwork phase, a researcher will need to calculate or estimate the minimum sample size required for performing the required statistical analysis to fulfill the evidence of validity of the questionnaire.*
- *It is important to note that the types of statistical analysis required for the fieldwork phase are usually stated in the published article or report of the original development of the questionnaire itself.*

# Chapter 3: Establishment of Reliability & Validity

## 3.1 Reliability testing

A weighing machine is said to be reliable when it consistently shows the same results of measurement. It is also termed as “stability” and “repeatability”. For questionnaire validation studies, the common and standard test to assess the reliability of a questionnaire will either be using (i) the Cohen’s kappa test (to test for the level of an agreement of a response that is expressed in a categorical manner via two different time points), or (ii) the intra-class correlation test (to assess the level of agreement of a domain score at two different time points), or (iii) the Cronbach’s alpha test (to test for an internal consistency of a domain).

A kappa agreement test is suitable for reporting the reliability of all types of questionnaires since it measures to what extent each individual item or question demonstrates agreement. On the other hand, the intraclass correlation test is suitable for reporting the reliability of a questionnaire which consists of several domain scores. Both these analyses require a pre- and post-type of study design where a researcher will have to administer the questionnaire twice within a time lag of one to two weeks (Omar et. al., 2011; Tan et. al., 2013). The overarching idea is to ensure that the response which is elicited at time two (for example: one week after a questionnaire administered to the same respondents at time one) to match as close as possible as the response which is elicited at time one, meaning that the respondents have already reached an agreement between the responses elicited at time one and at time two. This is to indicate that the respondents have

understood the questions well and so their responses will remain the same over time. In this test, a researcher will assume that those factors that can potentially affect the respondents' responses within that period of time (such as the social and living environment, the respondents' health condition, etc.) shall remain constant.

The third type of reliability test is the Cronbach's alpha test, which is suitable for use in a questionnaire that has at least one domain and the measurement scale of the domain(s) is consistent and standardized (for example: consistently using either four- or five-point Likert-type scales) and hence, it will be a suitable reliability test for use in a questionnaire that measures latent variables (Bujang et. al., 2016; Bujang et. al., 2017).

In summary, different types of questionnaires may require different statistical analyses to test for their reliability depending on the types of measurement scales



of the items. In other words, a researcher will need to carefully select the appropriate statistical test for conducting the analysis for its reliability testing. The manual which outlines the development of the original questionnaire will usually contain a statement which indicates the type of statistical analysis that is used to measure the reliability of the questionnaire. Thus, researchers can adopt the same statistical techniques to test for the reliability of the translated questionnaire. Researchers may also conduct further reliability tests for a particular translated questionnaire whenever necessary, as long as it shall not in any way violate the underlying terms and assumptions of a statistical test.

A researcher should take the first step by determining the reliability of the questionnaire by conducting a pilot study using a small sample. It is especially important to carefully screen through all the

questions at this juncture because a well-designed questionnaire which produces reliable results at the pilot stage will also produce reliable results in a full-scale study. In addition, if the results obtained from a pilot study are found to be reliable, then they are useful for convincing the researchers to pursue the actual full-scale fieldwork or survey, which is the next step for determining the validity and reliability of a questionnaire.

However, we need to be cautioned that there is no guarantee that by merely having excellent reliability, the questionnaire will also have a high level of validity (Omar et. al., 2011; Tan et. al., 2013). Likewise, it also follows naturally that a researcher can realistically predict that a questionnaire will not have a high level of validity if its reliability is found to be low during the initial pilot study (Bujang et. al., 2010; Omar et. al., 2010). Therefore, the findings obtained from the initial reliability testing of a

questionnaire during the pilot study of a questionnaire are very important because they serve to guide the researcher on what should be the next step for the validation of a questionnaire.

Depending on the findings obtained from the initial pilot study of a questionnaire, a researcher may either (i) modify the content of the questionnaire before conducting a fieldwork to assess its validity, or (ii) proceed directly to assess the validity of the questionnaire in the fieldwork phase by using a larger sample of respondents (if the results obtained from the pilot study are found to be satisfactory). If the reliability of a questionnaire is found to be too low, then it is necessary for a researcher to carefully revisit all the questions in the questionnaire again to identify the actual reason for its poor reliability (i.e. whether it is due to the inaccurate translation of the original questionnaire or to the inadequate tailoring of the

actual content of the original questionnaire to the specific group of local respondents).

Similarly, this also applies to a questionnaire with a moderate to low level of content validity, in that it is recommended for a researcher to rephrase the items in the questionnaire in order to improve its content validity. However, to avoid the possibility of inadvertently changing the true meaning of each item or construct in the original questionnaire, it is often recommended for a researcher to add more items to the original questionnaire rather than omitting or changing some of the items in the original questionnaire. This is because it is usually advisable for a researcher to replace the old items in the original questionnaire with some new ones only after conducting a fieldwork to perform an extensive assessment of both validity and reliability of these original items in the questionnaire (Bujang et. al., 2016; Jalaludin et. al., 2013).

Since a pilot study is conducted by using a small sample, therefore a researcher will need to be very careful when deciding to drop or to modify any questions in the original questionnaire before conducting a fieldwork survey using the questionnaire. It is often recommended for a researcher to add more questions or items to the original questionnaire, instead of deleting a few questions (or items) from the original questionnaire in order to avoid introducing errors which may arise from making inadvertent omission of these questions (or items) found in the original questionnaire. This is because the reliability of a questionnaire will always be amenable to further improvement when it is being tested in a larger sample of respondents, which also explains why it is always recommended for all the original items in the questionnaire to be retained for any subsequent testing and/or refinement.

In addition, although it may be tempting for a researcher to omit or revise some of the items in the original questionnaire based on their content alone; however, it is often preferable for such amendments to be made to the original questionnaire only after the results have been obtained by conducting the appropriate statistical tests. By doing so, it will enable a researcher to justify his/her omission, or addition of, or amendment of an item in the original questionnaire based on the results of the appropriate statistical tests.

*Key messages:*

- *As the fieldwork phase shall only be conducted on the translated version of a questionnaire after a researcher is satisfied with the reliability of the questionnaire which has been determined during a pilot study, hence it is recommended to make all*

*necessary corrections on the translated questionnaire before subjecting it to the fieldwork phase.*

- *Data collected in the fieldwork phase will also be used for a re-assessment of both reliability and validity of the questionnaire by using a larger sample size.*
- *Therefore, a pilot study can serve as a useful preliminary study that shall enable a researcher to determine whether or not the translated questionnaire will require any further amendment before conducting a fieldwork using the questionnaire to further improve its reliability and validity on a larger scale.*

### **3.2 Validity testing**

A weighing machine is said to be valid when it provides an accurate measurement. In other words, validity measures the degree of accuracy of the questionnaire which depicts how accurate a questionnaire measures what it is supposed to measure (Murphy & Davidshofer, 2005). Hence, the validity of a questionnaire is referring to the level of accuracy of a questionnaire or study instrument in deriving a correct and valid feedback from the respondents by posing the right questions.

It is important to bear in mind that a questionnaire with a very high level of reliability does not necessarily mean it will also have a very high level of validity. This means that although a weighing machine is repeatedly showing the same measurement for the weight of a particular person, it does not mean the measurement of weight as shown by the weighing machine is correct.



Therefore, in order to confirm that this weighing machine is giving an accurate measurement of weight, a researcher will have to validate the performance of this weighing machine by comparing its performance against that of a gold standard (i.e. a previously-calibrated weighing machine that is now regarded as a gold standard for benchmarking purposes).

In questionnaire validation studies, the validation process of a questionnaire which has been translated into a local (or native) language should be exactly the same as that of the original questionnaire, as it has already been validated in a foreign language. For example, one of the most commonly-used questionnaires is The Summary Diabetes Self-Care Activities (SDSCA), which is an excellent example of a questionnaire that was previously validated through an assessment of its construct validity (Toobert & Glasgow, 1994). Therefore, the same validation

process should also be applied for validating this questionnaire in all the other languages (Reviriego et. al., 1996; Yildirim et. al., 2007; Sato et. al., 2014; Bujang et. al., 2016).

For example, since exploratory factor analysis is a commonly-used statistical technique for testing the construct validity of a questionnaire; therefore, the same technique should also be used to assess the construct validity of the same questionnaire in all the other languages. In addition, it is worthwhile to remember that not only the same statistical technique should be used to assess the construct validity of a questionnaire in all the other languages; but the detailed statistical computations for conducting the exploratory factor analysis should also exactly the same, such as the choice of the factor extraction method and rotation method in exploratory

factor analysis (Reviriego et. al., 1996; Yildirim et. al., 2007; Sato et. al., 2014; Bujang et. al., 2016).

The validity of a questionnaire can often be assessed by using more than one validation techniques. The greater number of validation assessments which have been performed for a questionnaire, the higher level of evidence can be accrued to its validity. Therefore, some researchers have attempted to undertake further validation assessments by conducting Structural Equation Modeling (SEM) or RASCH analysis, apart from the conventional exploratory factor analysis (Haley et. al., 1994; Keller et. al., 1998). Since the beginning of this book has already provided a description of the qualitative validation assessments of a questionnaire such as the determination of its content and face validity, therefore the following section shall provide a description of some of the quantitative validation assessments of a

questionnaire (including concurrent validity, predictive validity, diagnostic assessment, known-groups comparison and construct validity).

### ***3.2.1 Concurrent validity***

Concurrent validity is a measure of validity in which another previously-validated study instrument or questionnaire is used as a yardstick for assessing the validity of a new questionnaire; and such an assessment shall also amount to the accruing evidence for affirming the validity of the questionnaire that is awaiting validation (Murphy & Davidshofer, 2005). In order to assess the concurrent validity of a new questionnaire, it is necessary to have validated and published an existing questionnaire which is widely used for the same purpose. This validated questionnaire shall then serve as a base of reference for the validation of the new questionnaire.

For example, a quality of life questionnaire using SF-36 is a widely-used questionnaire that is used to validate a new questionnaire called Diabetes Quality of Life (DQoL) (Jacobson et. al., 1994). A common statistical method to test for concurrent validity is through the determination of correlation or association where a researcher attempts to calculate a score which assesses the degree of correlation or association between the same domains from two different questionnaires. From Figure 3.1 below, it is clear that the result has shown that DQOL domains are associated with some of their corresponding domains from the other QoL instrument which is SF-36.

Table 2—Pearson correlations of DQOL and SF-36 scales for type I and type II diabetic patients

	SF-36									
	Type I diabetes					Type II diabetes				
	Physical functioning	Social functioning	Role physical functioning	Pain score	General health score	Physical functioning	Social functioning	Role physical functioning	Pain score	General health score
DQOL										
Total	0.38*	0.56*	0.51*	0.33†	0.60*	0.35*	0.34*	0.40*	0.38*	0.43*
Impact	0.37*	0.59*	0.49*	0.30†	0.58*	0.35*	0.32†	0.34*	0.39*	0.41*
Satisfaction	0.30†	0.43*	0.44*	0.28†	0.50*	0.33†	0.37*	0.42*	0.36*	0.42*
Diabetes worry	0.12	0.34†	0.26†	0.16	0.44*	0.08	0.19	0.26†	0.19	0.23
Social worry	0.21	0.46†	0.31	0.13	0.31	0.001	0.05	0.17	-0.003	0.17

\*  $P \leq 0.0001$ . †  $P < 0.01$ .

Figure 3.1: Illustration of an actual example of testing for validity based on an assessment of concurrent validity - Results obtained from Jacobson et. al., (1994)

### ***3.2.2 Predictive validity***

This is a measure of validity which involves the determination of the ability of a measurement score on a scale or test to predict a particular criterion measure. Let's say the test score obtained from the questionnaire is taken at time 1. To determine predictive validity of this questionnaire, it is necessary to test whether the score taken at time 1 will be able to predict another event in the future. Therefore, the best way to assess the predictive validity of a questionnaire is to determine the likelihood of an initial test score obtained from this questionnaire to accurately forecast either a better or a worse outcome in the future.

For example, a study done by Levi and colleagues (1999) had reported that some of the domains in the questionnaire (such as the composite functional measure)

were validated after two years to determine whether or not the negative outcome indicators such as psychiatric hospitalization and illicit drug use are possibly associated with personality disorders that were measured at baseline (Levy et. al., 1999). Based on Figure 3.2 below, the result clearly shows that participants who were screened to have personality disorders at baseline have a higher likelihood to be admitted for further psychiatric hospitalization and also to be involved in illicit drug use.



**TABLE 2. Functional Interview Composite Scores at 2-Year Follow-Up of 141 Adolescent Inpatients With and Without Personality Disorders**

Composite Functional Measure	With Personality Disorders (N=61)		Without Personality Disorders (N=40)		ANOVA (df=1, 100)	
	Mean	SD	Mean	SD	F	p
Employment level	1.57	0.63	1.41	0.55	1.49	n.s.
Further psychiatric hospitalization	1.64	0.78	1.29	0.60	0.86	<0.05
Drug use	1.69	0.90	1.38	0.71	3.43	<0.05
Alcohol abuse	1.72	0.84	1.88	0.94	0.74	n.s.
Legal difficulties	1.48	0.59	1.64	0.65	1.51	n.s.
Psychiatric symptoms	1.55	0.57	1.56	0.51	0.01	n.s.
Social/romantic relationships	1.44	0.41	1.44	0.45	0.00	n.s.
Family relationships	1.77	0.46	1.92	0.40	2.25	n.s.

Figure 3.2: Illustration of an actual example of testing for validity based on an assessment of predictive validity (Results obtained from Levy et. al., (1999))

Since the researchers will require a longer period of time to detect any changes in the concrete endpoints or outcomes when they assess the predictive validity of a questionnaire, it is therefore a less popular form of validity measurement. Nevertheless, predictive validity remains to be a very useful measure of assessing how likely it is for

the test scores in a questionnaire to predict the same outcomes in the future.

### ***3.2.3 Diagnostic assessment***

In medical research, a questionnaire is commonly used for screening and/or diagnostic purposes because it is cheap and quick when compared to the actual screening or diagnosing by using a gold standard measure. For example, a newly-developed questionnaire is said to be valid if it is able to discriminate between the ‘diseased’ and ‘non-diseased’ groups. Diagnostic assessment can be analyzed by using sensitivity and specificity test where the calculation is shown in Figure 3.3.

For example, both Berlin OSA and STOP-Bang questionnaires were successfully developed to screen patients with obstructive sleep apnoea (OSA) (Netzer et al., 1999; Farney et. al., 2011). In this case, if it is necessary to

validate the Berlin OSA and STOP-Bang questionnaires that have been translated into the other languages, then a similar analysis such as a sensitivity and specificity test would need to be applied to furnish their evidence of validity.

		Disease		Predictive Value	
		(+)	(-)		
Test	(+)	<b>A</b> True Positive (TP)	<b>B</b> False Positive (FP)	<b>Positive Predictive Value (PPV)</b> $\frac{TP}{TP+FP} = \frac{A}{A+B}$	<b>Total Positive Results</b> (A+B)
	(-)	<b>C</b> False Negative (FN)	<b>D</b> True Negative (TN)	<b>Negative Predictive Value (NPV)</b> $\frac{TN}{FN+TN} = \frac{D}{C+D}$	<b>Total Negative Results</b> (C+D)
Sensitivity & Specificity		<b>Sensitivity</b> $\frac{TP}{TP+FN} = \frac{A}{A+C}$	<b>Specificity</b> $\frac{TN}{FP+TN} = \frac{D}{B+D}$		
		<b>All diseased patients</b> (A+C)	<b>All non-diseased patients</b> (B+D)		

Figure 3.3: The calculation for sensitivity and specificity of a diagnostic test

For example, the Malay version of Berlin OSA (Berlin-M) was reported to have a sensitivity and a specificity of 92.0% and 17.0% respectively. These findings have shown that Berlin-M is a reliable and valid screening tool for OSA. However, its low specificity renders it unsuitable as a confirmatory tool to diagnose OSA. Thus, Berlin-M cannot be used to confirm the diagnosis of OSA. However, it can still serve as a useful screening tool for patients who may have OSA. For instance, Berlin-M can be used to screen patients in primary care who require referral to the tertiary center to confirm the diagnosis of OSA as well as for its further management (Yunus et. Al., 2013).

### ***3.2.4 Known-groups comparison***

Known-groups validity is being demonstrated whenever a test or questionnaire can realistically discriminate between two groups that are known to differ on the variable of interest (Davidson, 2014). A known-groups comparison can be performed to validate an individual item or a set of questions (or otherwise termed as a domain). A question or domain is considered valid if the two populations of specific and distinctly different characteristics appear to have significant differences in their mean or median score. An example is shown in Figure 3.4. The result shows that the two different groups namely “Healthy individuals” and “Patients receiving treatment in a psychiatric clinic” reported different mean scores for distress which was measured by using the K10 questionnaire. (Tiong et al., 2018). Thus, the K10 questionnaire is useful for

differentiating between the two extreme groups (which in this case, are distressed and non-distressed groups).

Table 4. A comparison of scores obtained by the Malaysian version of K10 between healthy individuals and those who are currently receiving treatment in psychiatric clinic

Known-groups (or extreme-groups) comparison	Median	IQR	p-value
● Healthy individuals	13.0	IQR (11.0, 17.0)	<0.001
● Individuals who are currently receiving treatment in psychiatric clinic	26.5	IQR (18.5, 32.0)	

Figure 3.4: Illustration of an actual example of testing for validity based on the known-groups comparison (Part of the results obtained from Tiong et. al., (2018))

**3.2.5 Construct validity**

The construct validity is often considered to be one of the preferred tests for both assessment and determination of the validity of latent variables that are measured by some of the questionnaires. As we have

alluded to earlier, a latent variable such as quality of life, compliance and depression cannot be objectively measured by using any machine or device. The best way to measure these latent variables is to develop a reliable and valid questionnaire for measuring them. Some examples of the commonly-used questionnaires that measure latent variables include CDI (measure depression symptoms among children), SF-36 (measure quality of life), SDSCA (measure diabetes self-care activities), DASS (measure psychological symptoms) etc. and all these questionnaires have already been validated through an assessment of their construct validity (Kovacs, 1992; Ware & Sherbourne, 1992; Toobert & Glasgow, 1994; Lovibond & Lovibond, 1995).

The construct validity measures both convergent and discriminant validity at the same time. Convergent validity refers to the degree to which the scores obtained

from a test correlates with (or are related to) scores obtained from the other tests that are designed to assess the same construct. Meanwhile, discriminant validity aims to furnish evidence of the extent to which the measures of a construct that theoretically are not highly related to each other are, in fact, found not to be highly correlated with each other (Bujang & Tan-Hui, 2022). For questionnaires that measure latent variables, convergent validity and discriminant validity have to be fulfilled based on exploratory factor analysis and/or confirmatory factor analysis. For this type of questionnaire, stating a questionnaire is valid merely depends on convergent validity or discriminative validity alone is not sufficient.

As shown in Figure 3.5, there are eighteen items used to measure three different latent variables. A good construct consists of items which represent their own domain only, which will enable it to accurately identify



common factors that can explain the order or structure among the measured latent variables. For example, based on Figure 3.5, “8 items coloured with green” measures Factor 1, “5 items coloured with purple” measures Factor 2 and “6 items coloured with red” measures Factor 3. On the other hand, a problematic construct consists of items which overlap with the other domains, which will preclude it from identifying only those factors that can explain the order or structure among the measured latent variables.

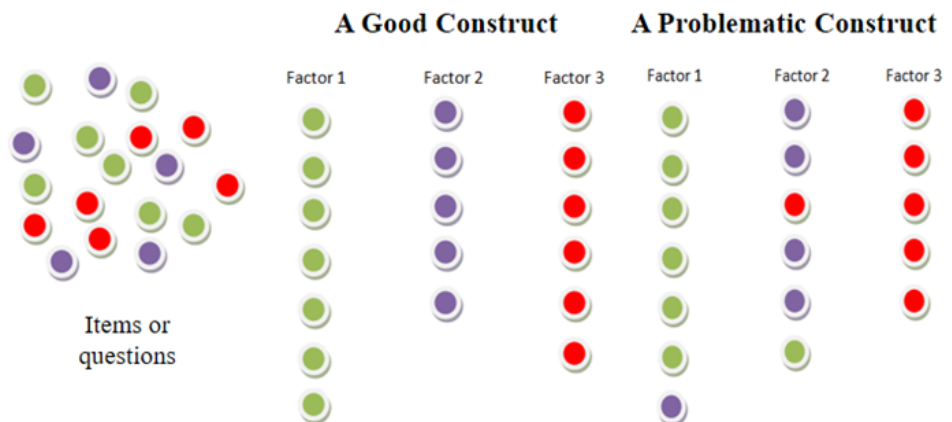


Figure 3.5: A visual representation of how an exploratory factor analysis works

The theoretical basis which underpins the concept of construct validity is that if a group of items in a questionnaire is said to be measuring a single domain of interest, then these items can be grouped together as a single construct. In addition, if all the items for a questionnaire are said to be measuring more than one domain of interest, then the proposed statistical analysis should ideally be able to discriminate between the items and segregate them into different domains. For example, the short version of the DASS questionnaire consists of three main domains of interest (such as depression, anxiety and stress) with seven items for each domain (Lovibond & Lovibond, 1995). Therefore, any validation studies to be performed for determining the construct validity of this questionnaire which has been translated to

all the other languages should also be able to derive the three main domains of interest, with each domain containing the same number of items as those in the original questionnaire (Patricia et al., 2002; Musa et al., 2007).

For these types of questionnaires, the exploratory factor analysis (EFA) is one of the common statistical methods used to test for the construct validity of a questionnaire and some other studies will then further analyze the findings obtained from the confirmatory factor analysis (CFA) by using a structural equation modeling technique to analyze the structural relationship between measured variables and latent constructs (Sato et al., 2014; Bujang et al., 2018).

A problem commonly encountered when testing for the construct validity of a questionnaire is that the constructs that are derived from the questionnaire

validation study are not similar to those found in the original questionnaire. Assuming that both the research design and research methodology of the questionnaire validation study (such as the socio-demographic characteristics of the study population, the overall translation process, the sampling frame, the choice of sampling methods and the sample size chosen) are exactly the same as those of the development of original questionnaire; then such differences could possibly be due to the perceived differences in the conceptual context between the respondents from the validation study and those from the development of the original questionnaire. This can arise from any differences that may exist between the way by which the local (or native) population and that by which the foreign population has perceived each item in the questionnaire undergoing validation and in the original questionnaire, respectively.

If such a problematic construct occurs, then it is necessary to take a remedial step to overcome this problem by minimizing such perceived differences in the conceptual context of a construct. One of the ways to achieve this is by identifying and defining a new domain, which can be done if the validation study has identified a new construct with a reasonably strong factor loading (Omar et al., 2011). However, it is also possible for some other validation studies to reach the conclusion that there is insufficient evidence to refute the original construct of the questionnaire, and hence, all the items in the original questionnaire should be retained to only reflect the overall domain of interest (Tan et al., 2013). This can occur when a researcher fails to identify a suitable new construct in the questionnaire to replace the problematic construct after completing the validation study—even if they are aware that the construct of the questionnaire undergoing

validation is slightly different from the original questionnaire.

On the other hand, it is frequently advised that a researcher not change the original questionnaire's construct even if the questionnaire validation study yielded a slightly different questionnaire construct (Musa et al., 2007). This can be deemed acceptable if the construct of the original questionnaire has a reasonably sizable factor loading in the original domain as compared to the other domain(s). For example, the item or question 1 should report a higher factor loading in domain "A" but it has actually reported a higher factor loading in a different domain such as domain "B". In this case, even if the factor loading of item 1 in domain A is just slightly lower than that in domain "B"; a researcher can still regard item 1 as belonging to domain "A" because of its higher content validity.

This is because although the post-tests in the validation procedure, which are usually based on factor-analytic techniques such as EFA and CFA, are more scientific and rigorous, the preliminary steps for conducting a content analysis on the scales are still mandatory because whenever a new scale or questionnaire needs to be developed, following a structured method will enable us to develop a quality instrument, since any mistakes committed in the initial stages of scale development will often lead to problems in the later stages. This also means that any decisions which arise from the preliminary content analysis on the scales shall be allowed to override those that are derived from post-tests in the validation procedure.

Hence, it must be emphasized that many problems can arise if a researcher decides to introduce a totally different construct into the original questionnaire, or not

to adopt its original construct in the questionnaire which is awaiting validation. This is because it will be very difficult to compare the findings across different study respondent populations with different languages if a new construct is being introduced into the original questionnaire. On the contrary, some researchers may opt to omit a few items from the original questionnaire with the intention of developing a better construct of the questionnaire which is awaiting validation.

However, omitting several items from the original questionnaire can lead to devastating consequences because not only the domains are now changed, but the total number of items in each domain is also different now, which will necessitate a totally different approach for the calculation of the total score for each domain (Omar et al., 2011). Hence, this modified version of a validated



questionnaire is no longer the same as the original questionnaire.

In summary, there is often no standard approach for a researcher to decide whether (or not) it is acceptable to modify the original construct of the questionnaire that is awaiting validation. Under all circumstances, a researcher should always attempt to consider both the ‘pros’ and ‘cons’ of each approach he/she chooses to take when dealing with any differences in the constructs of the questionnaire which may have arisen during the validation process, in order to derive a valid conclusion(s) from the study.

*Key messages;*

- *Before embarking on a pilot study, it is necessary to determine the content validity and the face validity of a questionnaire.*

- *Next, it is then necessary to conduct a fieldwork study and measure the validity of the questionnaire by using an appropriate statistical technique. Although there is a myriad of different types of validity testing, a researcher will need to select one or more suitable test(s) and does not necessarily have to conduct all the various types of validity tests in order to validate a questionnaire.*
- *Therefore, it is recommended for a researcher to obtain the manual which contains the details of the development of the original version of the questionnaire that is awaiting validation. The manual usually is published as a scientific article or book or report.*
- *The specific methods employed to validate the original questionnaire which is provided by the manual can often be applied to the questionnaire*

*awaiting validation. If necessary, a researcher can supplement these methods with other additional means to validate the questionnaire, so long as it is justifiable and appropriate to do so.*

### 3.3 A summary of requirements for conducting reliability and validity tests for a questionnaire validation study

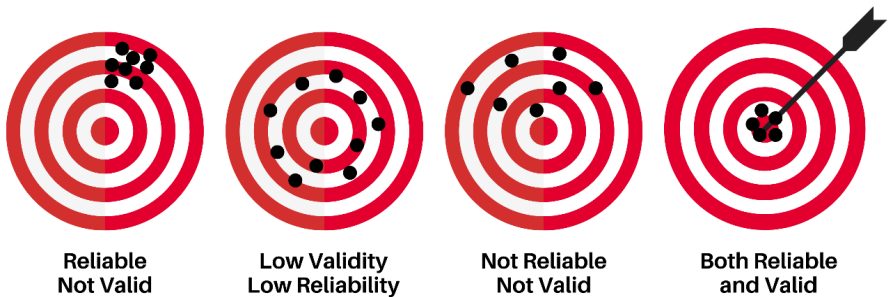


Figure 3.6: A visual concept of reliability and validity

A good questionnaire should have both traits of being reliable and valid. Figure 3.6 shows some examples

of what it means a concept of reliability and validity. Based on this figure, both reliability and validity are manifested when all the dots are consistently plotted at the centre or the main target. Similarly, a reliable and valid questionnaire should fulfill these two elements. Basically, one of the criteria to say a questionnaire is reliable is when the respondents can give a consistent response to the same questions. And a questionnaire is said to be valid when the questionnaire measures a particular variable or domain correctly.

Although there are so many different approaches to testing for the reliability and validity of a questionnaire, it is not necessary for a researcher to apply all these tests when they validate a questionnaire. This is because, although a good questionnaire must be both reliable and valid, the correct approach for a researcher to take for determining the reliability and validity of a questionnaire

will vary according to the type of questionnaire. Therefore, each type of questionnaire should have its own recommended approach for determining both its reliability and validity.

This point has already been highlighted in a previous section of this paper, where the readers are being introduced to the first two types of questionnaires, namely *Questionnaire Type 1* (questionnaire that is designed to collect information) and *Questionnaire Type 2* (questionnaire that is designed to assess the knowledge level) questionnaires. The *Questionnaire Type 1* questionnaire is one that aims to collect facts or information about a respondent and does not measure any latent variables. A common example of this type of questionnaire is one which collects specific information pertaining to an individual respondent, such as the socio-demographic characteristics of a patient, etc. Since

these data are not measured by using a Likert scale, so it will not be necessary to assess its reliability using Cronbach's alpha (which measures its internal consistency).

However, it is still necessary to ensure that an acceptable degree of agreement has been reached by determining its test-retest reliability using the Cohen's kappa or intra-class correlation coefficient. Apart from the above, it is also necessary to assess its content validity among a few subject matter experts (i.e. SMEs), and its face validity among a few potential subjects in order to confirm its validity. In other words, it is the authors' opinion that it is usually not necessary to assess the validity of *Questionnaire Type 1* questionnaires by using a statistical test(s).

*Questionnaire Type 2* questionnaire is one that is designed to assess the knowledge level of the respondents,

which is regarded as a single latent variable. A common example of this type of questionnaire is an examination paper. It is compulsory for this type of questionnaire to have an adequate level of content validity; however, it is often not necessary to assess its face validity because the robustness of this type of questionnaire has already been built-in from its content validity. Therefore, it is often sufficient to confirm its content validity alone, which is done by administering it to the subject matter experts in the relevant field.

One of the reasons for obviating the need to assess the face validity of *Questionnaire Type 2* questionnaire is that, for example, it will be unethical to expose the examination paper to the students. Likewise, it will also not be necessary to determine the face validity of a questionnaire which assesses the knowledge level of respondents (such as knowledge of washing hands,

awareness about antibiotic use, etc.) as long as a researcher (who is also the subject matter expert in this field) has knowledge about the scope, breadth and depth of knowledge to be measured by the questionnaire and therefore will be able to incorporate a list of suitably-designed questions that aim to elicit the right responses from the respondents.

Notwithstanding the above, the content validity of *Questionnaire Type 2* questionnaires must always be assessed, which can be done by conducting a specific statistical analysis to determine whether the set of questions in the examination paper actually matches the students' ability. These statistical analyses will be conducted after all the respondents have completely answered all the examination questions. For example, the RASCH analysis is a useful statistical technique to determine the suitability of the level of difficulty of the



questions that are set during an examination to assess whether the differing levels of difficulty in the examination questions have directly corresponded to the various levels of students' capability (Liu & Boone, 2006).

Finally, the *Questionnaire Type 3* questionnaire (for diagnostic and screening purposes) and *Questionnaire Type 4* questionnaire (for measurement of specific latent variables) are also heavily reliant on the appropriate statistical analyses to confirm their validity, apart from the usual assessment of both content and face validity (Yunus et al., 2013; Bujang et al., 2016). To summarize, Table 3.1 shows a list of recommended statistical techniques that can be applied to test for reliability and validity of four different types of questionnaires. Conventionally, researchers should perform assessments of both the validity and reliability of a questionnaire by following the same statistical methods which were used for testing the

validity and reliability of the original questionnaire during its initial development by the author. However, researchers may opt to furnish further evidence to support and strengthen the claim of both its validity and reliability. Table 3.1 can serve as a guideline for researchers to choose what types of statistical techniques would be appropriate to be deployed for validating a questionnaire.

Table 3.1: Recommended statistical techniques to be deployed for testing reliability and validity in four different types of questionnaires

Types of reliability and validity testing	Questionnaire designed to			
	collect facts or information	measure knowledge level	measure latent variables	screen for or diagnose a disease in a subject
Test-retest reliability	X	X	X	X
Internal consistency			X	X
Content validity	X	X	X	X
Face validity	X	X	X	X
Known-groups comparison		X	X	X
Concurrent validity		X	X	X
Predictive validity		X	X	X
Construct validity			X	
Diagnostic assessment				X

Note:

“X” indicates the possible type of statistical technique(s) to be deployed for validity and reliability testing

# **Chapter 4: Other Issues Pertaining to Reliability and Validity Testing for a Questionnaire Validation Study**

## **4.1 Copyright protection**

Although there were many previous successful attempts at validating a questionnaire which has been translated into many different languages, it must be emphasized that the original author of these questionnaire remains as the sole copyright owner of this questionnaire because he/she will most probably have been granted the exclusive copyright of this questionnaire (Ware & Sherbourne, 1992). Hence, before using an existing questionnaire for any research purposes, it is first

necessary to obtain prior written permission from the copyright owner of this questionnaire unless the author has given specific written instructions that the questionnaire will be made available for public access (and it can also be freely available from the website) (Lovibond & Lovibond, 1995). In this case, a researcher should then acknowledge the contribution of the author(s) and/or institution(s) that were involved in both the development and validation of the questionnaire, by giving appropriate credit to the original author(s) by citing their work in the reference list. However, the original authors of some questionnaires may require a researcher to pay a nominal fee before granting him/her access to them (Ware & Sherbourne, 1992; Kovacs, 1992).

## **4.2 The practice of combining more than one existing questionnaire to form a new questionnaire**

Is it viable for the researchers to combine a few validated questionnaires into a single questionnaire? The answer to this question will depend on the situation; however, in general, it is not recommended for the researchers to do so because there are many legal repercussions for combining more than one questionnaire into a single questionnaire. Firstly, the copyright owner of the questionnaire may not allow other researchers to re-use his/her questionnaire for the purpose of creating a new questionnaire. Secondly, different questionnaires may adopt different approaches in the way they ask questions and they may also use a different measurement scale for measuring the same domain. Hence, it may not be viable to combine the different questionnaires into one because this will pose difficulties for a researcher when he/she will

have to subject this new questionnaire to various other statistical tests for validating it later on.

One way to obviate the need to merge several different validated questionnaires into a single questionnaire is to simultaneously use several different validated questionnaires for a particular research study (Bujang et al., 2013; Bujang, et al., 2015). For example, it is a common practice to use two different validated questionnaires such as SDSCA and Diabetes Quality of Life (DQoL) for the same survey which aims to determine the association between diabetes' self-care activities and the quality of life among patients with type 2 diabetes mellitus (Bujang et al., 2013). By doing so, the researchers have decided not to modify the questions in the two original questionnaires but instead, use the original questions from both the questionnaires for a particular survey. Thus, the originality of both questionnaires will be

retained. However, if it becomes inevitable for certain modifications to be made on either of the two questionnaires, then such modifications along with their accompanying justifications should be pre-specified in the methodology section of the paper.

A special case occurs when a researcher has been granted permission by the publishers or authors to re-use their questions from several different validated questionnaires, and the researcher has found there are sufficient grounds for validating a new questionnaire which shall now consist of questions obtained from several different previously-validated questionnaires. This type of work has very rarely been reported in the literature, because although it seems an easy task to combine several questionnaires into one, it is usually very difficult to do so since many potential problems can crop up. Therefore, if a researcher has found that a single



questionnaire is not available to fulfill the objectives of the study (for example, to measure the latent variables), then the next best option will be to develop a completely new questionnaire for the purpose of fulfilling the research objectives of the study.

### **4.3 The practice of combining a validation study and a research study together by using an unvalidated questionnaire that is awaiting validation for a research purpose**

It is sometimes necessary for some questionnaire validation studies to adopt a different approach for its methodology as compared to that of the original survey research (in which the initial questionnaire will be used for a research purpose). For example, during the validation of a questionnaire, a researcher may have to recruit samples from different population characteristics as compared to

the population from an actual study. For example, a questionnaire may require a healthy and normal population for validation purposes but, a researcher may intend to use the same questionnaire for a non-healthy population to answer his research question. In this case, the study must be divided into two parts: one for questionnaire validation and one for using the validated questionnaire for research purposes (in a research study or survey research).

However, if the methodology especially regarding the populations' inclusion and exclusion criteria for the validation study is exactly the same as that of the research study, then a researcher may choose to combine the two studies into one (Bujang et al., 2016; Mohd-Hatta et al., 2018). Meaning to say, researchers recruit respondents for a survey but will use some of the respondents for validation purposes. However, it must be emphasized that

this approach can be taken by a researcher if and only if the questionnaire has already been determined to be reliable and valid. Therefore, a researcher will have to first establish the validity of a questionnaire by publishing a paper for an initial questionnaire validation study before proceeding to publish the research study in which the same questionnaire has been used for a research purpose. This approach can be risky especially when the questionnaire is found to be invalid. Then data collection for a whole survey can be a waste.

A review of the existing literature in this subject has identified several previously published papers which had combined the validation study and research study together by using an unvalidated questionnaire that is awaiting validation for a research purpose. One of these studies involves the assessment of knowledge, attitude and practices (KAP) in a group of study respondents (Askarian

et al., 2007; Khaled et al., 2008). Such an approach can easily be adopted for both *Questionnaire Type 1* and *Questionnaire Type 2* types of questionnaires because there are only a few specific requirements for their validation as compared to the validation of *Questionnaire Type 3* and *Questionnaire Type 4*, which can often be much more complicated.

Therefore, it is usually not recommended for a single paper to combine a questionnaire validation study with a research study by using either a *Questionnaire Type 3* and *Questionnaire Type 4* in which an unvalidated questionnaire that is awaiting validation will be used for a research purpose (Note: see the classification of the four categories of the questionnaire, section: Types of questionnaires). This is because the validation process of these types of questionnaires are very complicated, which will make the paper appear cluttered with too many tables

and figures, and most peer-reviewed journals will only allow a maximum of 5 to 6 tables or figures for a single paper.

#### **4.4 Sample size requirements**

To perform an assessment of a questionnaire, a researcher will usually evaluate its reliability such as by using Cronbach's alpha test, kappa agreement test or intra-class correlation test. These statistics will usually require fairly small sample sizes (Bujang & Baharum, 2017a; Bujang & Baharum 2017b, Bujang et al., 2018). However, there are other statistical tests for assessing its validity which will usually require larger sample sizes, such as the exploratory factor analysis which is commonly used to test for construct validity and the sensitivity specificity analysis that is commonly used to test for discriminant

validity (Osborne & Costello, 2004; Bujang & Adnan, 2016; Bujang et al., 2012; Bujang et al., 2019).

A pilot study will typically require a small sample size of less than 30 subjects for assessing both the internal consistency and the degree of inter-rater agreement for a questionnaire. However, in order to conduct a fieldwork study for the purpose of questionnaire validation, it will be necessary to determine the minimum sample size required, which is based on the specific statistical analysis(es) to be deployed for the measurement of validity of a questionnaire. For example, if a researcher is conducting a validation study for *questionnaire Type 3* which is developed to measure some latent variables, therefore an appropriate validity test is Exploratory Factor Analysis (EFA). Therefore, the minimum sample size required for this purpose shall be estimated by using the 'rule of thumb' to determine sample size requirements for

EFA such as 10 respondents are required for one item. In other words, a minimum of 360 respondents are required to test for 36 items (Bujang et al., 2012; Bujang et al., 2019; Bujang, 2021).

#### **4.5 In what circumstances will validation of the questionnaire be deemed unnecessary?**

It may occasionally be possible for both the translation and the subsequent validation process of a translated questionnaire to be deemed unnecessary. For example, the original questionnaire has been developed in the English language since the survey was initially conducted for an English-speaking population. Then, a researcher intends to use the same questionnaire for various other non-English speaking populations, such as in Malaysia. This is possible as long as the following three conditions have been fulfilled:

- a. The population has the same level of ability to understand spoken English as the English-speaking population.
- b. All the questions fit their knowledge, subject matter expertise, personal experience and situation well.
- c. None of the questions can possibly elicit any responses that may deviate from the norm due to any differences in understanding of the question that might be arisen due to differences in (i) the terminologies used, (ii) the cultural factors, (iii) the lifestyle choices, (iv) the living and social environment or (v) the religious beliefs and/or practices.

For example, an original questionnaire has been developed in English to assess the level of perception among medical doctors regarding the use of antibiotics medications. In most countries, the graduate program for



the medical degree is taught in English, especially in Malaysia where English has become the second language for the country. In this situation, the ability of our medical doctor in understanding the English language for instance is believed to be as good as that among the medical doctors in either the United Kingdom or the United States. Although the level of language proficiency might be not exactly the same; however, our medical doctors can still speak, read and write excellent English and that is why they can graduate.

The next question is whether or not our medical doctors are capable of providing valid responses towards all the questions. Our medical doctors can provide an appropriate response to the question if it relates to their own personal data such as their demographic profiles, or to the basic subject matter regarding common medical knowledge, or medical practice, etc. However, if our

medical doctors have a differing professional experience (such as the choice of treatment options offered for a particular medical condition) as compared to those from Western countries due to advancements in the use of technology and medication per se, perhaps the questions might not be suitable for our medical doctors. They might not be able to provide a valid response since they have a different level of professional experience in terms of their choice of treatment modality for a particular medical condition.

The third condition refers to differences in lifestyle choices, cultural factors, religious beliefs and/or practices and many other issues that might create differences in their understanding of the content of the question. Take an example about the cultural difference for instance, how Western people perceive certain things might be different from how Asian people perceive the same things. For

example, walking into the house with a pair of shoes is commonly practiced by Western people but it is usually prohibited by most Asian people. Thus, the practice of entering a house with a pair of shoes may be regarded as impolite among Asians but not among Westerners. So, this example illustrates how two people may possibly perceive the same question differently when both of them come from a different background.

The authors have previously highlighted the three reasons where the translation and a whole validation process for the translated version is not necessary for a particular questionnaire. Therefore, it is necessary for researchers to at least assess both content and face validity of a questionnaire that has been developed in a foreign language. Besides the language issue, the content validity shall also be determined in order to ensure all the questions in the translated questionnaire are relevant to

the local population and also commensurate with the scope of research. Then, face validity will ensure the local respondents are now able to provide their valid responses when they fill in the foreign questionnaire (i.e. questionnaire that is developed in English) and they understand the content of the questionnaire in exactly the same way as the understanding among the English-speaking population.

Thus, there is always a possibility for a foreign questionnaire not to require translation and hence the necessity to undergo the whole process of validation. However, it is still mandatory for a foreign questionnaire to undergo a careful assessment of both its content and face validity before the researcher can foretell that the foreign questionnaire will not require a proper translation and validation process. Under a majority of the circumstances, however, a proper translation and a whole

validation process is still necessary to avoid the possibility of introducing bias in the feedback provided by the respondents which might occur due to a variety of reasons.

#### **4.6 Types of questionnaires that are probably not suitable for undergoing validation**

It should be borne in mind that not all questionnaires published in the literature, books, internet or reports are deemed suitable for validation studies. This is because some of these questionnaires had been accepted for publication without going through a proper questionnaire development process. This occurs when a questionnaire was published without providing sufficient evidence of its validity and reliability. As a common practice, the researcher will usually develop a questionnaire and then will publish it alongside a questionnaire development on

the same questionnaire. The work can be published as a scientific paper or book or report.

This situation is true for questionnaire Type 3 (questionnaire to measure latent variable) and questionnaire Type 4 (questionnaire for screening and diagnostic purposes). Questionnaire Type 1 (questionnaire to derive facts and information) and Questionnaire Type 2 (questionnaire to measure knowledge) will probably not require a separate publication to describe its entire development process since the development process can always be stated in the same scientific article alongside the findings of the survey itself. According to the authors' experience, publishing a single scientific paper to specifically describe a questionnaire development process for Questionnaire Type 1 (questionnaire to derive facts and information) and Questionnaire Type 2 (questionnaire to

measure knowledge) is rarely necessary, though it is still possible.

The main take-home message is that there is no guarantee that any questionnaires which have been published in the existing literature, books, internet or full-text reports can be truly reliable and valid. For a questionnaire to be deemed valid and reliable, it is necessary for the researchers who are involved in developing the questionnaire to garner sufficient evidence of both its reliability and validity, which can often be found in the citing literature. By simply assuming that a questionnaire is reliable and valid when it is actually not, the researcher may subsequently encounter problems when he/she is conducting the validation study on a questionnaire if it turns out to have failed to produce sufficient evidence of both its reliability and validity. Nevertheless, nobody can forbid a researcher to conduct a

validation study on a questionnaire that did not undergo a proper validation process during its initial development.

For example, the Diabetes Quality of Life (DQoL) questionnaire was initially developed by the Diabetes Control and Complications Trial (DCCT) Research Group which used it for a clinical trial project (The DCCT Research Group, 1988). The DQoL instrument contained 46 items and was used to measure health-related quality of life among diabetes patients by basing on three main domains, namely, “satisfaction,” “impact,” and “worry.” However, there is no publication on the initial development and validation of the DQoL questionnaire. It is possible that this questionnaire went through some sort of questionnaire development and validation process; however, such information is not available in the existing published literature. Another plausible explanation for this is that a validated questionnaire for the determination of



the quality of life among diabetic patients does not exist and hence, the DQoL was merely designed and developed for use within a specific clinical trial project by the DCCT Research Group, 1988.

After the study findings of the clinical trial project were published in 1988, the DQoL instrument had sparked interest among many researchers from all over the world to use the same questionnaire. Some of these researchers had assisted to perform the questionnaire validation for DQoL in different languages such as Spanish, Turkish, Japanese, Chinese, Taiwanese, Iranian, and Malaysian (Reviriego et al., 1996; Cheng et al., 1999; Yildirim et al., 2007; Huang et al., 2008; Pakpour et al., 2012; Sato et al., 2014; Bujang et al., 2017). Most of these studies reported that DQoL have demonstrated evidence of excellent reliability but only some evidence of validity. However, it must be emphasized that the evidence for its validity is not

satisfactory since the construct of the three domains namely “satisfaction”, “impact” and “worry” had not originally been derived from an assessment of its construct validity.

Some efforts have been made for developing a shorter version of DQoL instrument, and it was considered as a brief version of DQoL instrument (Thomas et al., 2004). However, the validation of the three main domains was not covered. A recent publication on the overall validation of DQoL had incorporated the function of the three domains. Basically, the researchers reduced the number of items and then validated the shorter version of DQoL by using Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA) and RASCH analysis (Bujang et al., 2018). Therefore, a reliable and valid shorter version of DQoL has been successfully developed.

In summary, it is not recommended to conduct a validation study for a questionnaire which does not have a clear history of proper questionnaire development process. If a researcher still intends to conduct a validation study for that particular questionnaire, then it is necessary for him/her to set up the appropriate reliability and validity testing for that questionnaire. Some tips on identifying which test to be used for validation and reliability-testing of a questionnaire can be obtained by referring to Table 3.1, Chapter 3. Nonetheless, it must be borne in mind that prior to the actual validation study, it is necessary for a researcher to conduct an initial phase of assessment by performing a proper translation process and then evaluating both the face validity and content validity of its translated version, in order to determine whether or not the subsequent validation study would be feasible and worthwhile.

# Chapter 5: Conclusion

A researcher will need to be cognizant of the fact that a suitable questionnaire for use in a validation study is one which has already been published and validated previously. As a result, a researcher must have access to the original questionnaire's development manual or questionnaire construction, which is the primary source of reference that provides specific hands-on instructions on both the development and validation of the original questionnaire. During the validation of a questionnaire, a researcher will have to follow almost all the same procedures as those when the questionnaire was originally developed; therefore, it is very important for a researcher to ensure that the right research methodology (for the

development and validation of questionnaire) has been adopted right from the very beginning, including the selection of the research design and the proposed plan for statistical analysis. However, it is heartening to know that a researcher shall always have the liberty to furnish any further evidence of both the reliability and validity of a questionnaire whenever necessary, in an effort to further confirm its validity and versatility.

The final concluding remarks the authors can offer here is that this book is providing a simple checklist (Table 5.1) to serve as a useful guide for a researcher to conduct a questionnaire validation study. A total of 17 checklists are available for a researcher to use to conduct a questionnaire validation study, right from the conception of study design through to the derivation of its conclusion. The same checklist can also be used as a tool for performing the critical appraisal of a scientific paper that

describes a research study involving questionnaire validation. As there are no standard guidelines which are currently available for stipulating all the minimum requirements for a questionnaire validation study, therefore the above recommendations provided by this book shall serve an important function of facilitating a researcher in conducting a questionnaire validation study by providing a useful checklist of the minimum standards and requirements for a questionnaire validation study.

Table 5.1: A brief guide of planning and conducting validation for a questionnaire

No	A detailed list of step-by-step procedures for a careful evaluation and appraisal of a questionnaire validation study	Remark(s)*
1	The title should emphasize on "Questionnaire validation study..."	Yes/No/Not applicable
2	The selection process for choosing a suitable questionnaire for the purpose of validation has been adequately described.	
3	The process for the assessment of its content validity has been adequately described. <i>(To state who makes up the panel for the assessment of its content validity.)</i>	
4	The result(s) obtained from the assessment of its content validity is/are found to be satisfactory. <i>(To decide if a further amendment to the items or questions will still be required.)</i>	
5	The translation process has been appropriately and adequately conducted. <i>(One or two arms? Forward and backward translations or just forward translation only?)</i>	
6	The process of assessment of its face validity has been appropriately conducted. <i>(How has its face validity been assessed and who are the respondents involved in this assessment process?)</i>	
7	The result(s) obtained from the assessment of its face validity is/are found to be satisfactory.	

	<i>(To decide if a further amendment to the items or questions will still be required.)</i>	
8	Describe how the pilot study has been conducted. <i>(How many respondents and who are the respondents involved in the pilot study?)</i>	
9	The result(s) obtained from the pilot study is/are found to be satisfactory. <i>(To decide if further amendments to the items or questions will still be required.)</i>	
10	The methodology (which includes: population, sampling technique, sample size and statistical analysis) of the validation study is about similar to that of the development of the original questionnaire. <i>(Please describe and provide valid justifications for any variation(s) that may exist between the methodological approach of the validation study and that of the development of the original questionnaire.)</i>	
11	How is the assessment of its reliability being performed?	
12	Are the results obtained from the assessment of its reliability found to be satisfactory? <i>(To describe the results.)</i>	
13	How is the assessment of its validity being performed?	
14	Are the results obtained from the assessment of its validity found to be satisfactory?	



	<i>(To describe the results.)</i>	
15	Is it still necessary to make further amendments to the translated version of the questionnaire in order to tailor it for use by the local population? <i>(If “no”, then proceed directly to question 16.)</i>	
16	To confirm that the translated version of the questionnaire is suitable for use by the local population. <i>(To describe its versatility for use by the local population in terms of its validity and reliability, and to list down any of its inherent limitations and/or weaknesses.)</i>	
17	State the limitation(s) of this study if necessary.	

\*Note: To answer “yes” or “no” or “not applicable” and then to provide the valid justifications.

# References

Abdullah, A., Naing, L., & Nizar, A.J. (2006). Validation of the Malay Brief Pain Inventory Questionnaire to Measure Cancer Pain. *J Pain Symptom Manage*, 31(1), 13–21.

Askarian, M., McLaws, M.L., & Meylan, M. (2007). Knowledge, attitude, and practices related to standard precautions of surgeons and physicians in university-affiliated hospitals of Shiraz, Iran. *Int J Infect Dis*, 11, 213–219.

Bujang, M. A. (2021). A step-by-step process on sample size determination for medical research. *Malaysian Journal*

of Medical Sciences, 28(2), 15–27. <https://doi.org/10.21315/mjms2021.28.2.2>

Bujang, M.A., Ab-Ghani, P., Soelar, S.A., & Zulkifli, N.A. (2012). *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). Langkawi, Malaysia, Sept 10–12. doi:10.1109/ICSSBE.2012.6396605.

Bujang, M.A., & Adnan, T.H. (2016). Requirements for Minimum Sample Size for Sensitivity and Specificity Analysis. *J Clin Diagn Res*, 10(10), YE01–YE06.

Bujang, M.A., Adnan, T.H., Mohd Hatta, N.K.B., Ismail, M., Lim, C.J. (2018). A revised version of Diabetes Quality of

Life Instrument maintaining domains for satisfaction, impact, and worry. *J Diabetes Res*, 22, 5804687-5804610

Bujang, M.A., & Baharum, N. (2017a). Guidelines of the minimum sample size requirements for Kappa agreement test. *Epidemiol Biostat Public Heal*, 14(2). doi:10.2427/12267.

Bujang, M.A., & Baharum, N. (2017b). A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review. *Arch Orof Sci*, 12(1), 1-11.

Bujang, M. A., Ghani, P. A., Soelar, S. A., Zulkifli, N. A., & Omar, E. D. (2019). Invalid Skewed Responses Contributes to Invalid Factor Solution in Exploratory Factor Analysis:

A Validation Approach Using Real-Life Data.*J Behav Health, 8*(4), 152–160.

Bujang, M.A., Ismail, M., Mohd-Hatta, N.K.B., Baharum, N., Othman, S.H., Mat-Lazim, S.S., & Shah, S.A. (2016). Validation of the summary diabetes self-care activities (SDSCA) in Malay language for Malaysian adults. *Malaysian Journal of Public Health Medicine, 16*(3), 227–234.

Bujang, M.A., Ismail, M., Mohd-Hatta, N.K.B., Othman, S.H., Baharum, N., & Mat-Lazim, S.S. (2017). Validation of the Malay version of Diabetes Quality of Life (DQOL) Questionnaire for Adult Population with Type 2 Diabetes Mellitus. *Malays J Med Sci, 24*(4), 86–96.

Bujang, A., Musa, R., Haniff, J., Muhammad, N.A., & Omar, K. (2010). Comparison of Family Environmental Scale (FES) subscales between Malaysian setting with the original dimension of FES. *Malaysian Journal of Psychiatry*, 19(1), 1–5.

Bujang, M.A., Musa, R., Liu, W.J., Chew, T.F., Lim, C.T.S., & Morad, Z. (2015). Depression, anxiety and stress among patients with dialysis and the association with quality of life. *Asian J Psychiatr*, 18, 49–52.

Bujang, M.A., Omar, E.D., & Baharum, N.A. (2018). A review on sample size determination for Cronbach's alpha test: a simple guide for researchers. *Malays J Med Sci*, 25(6), 85–99.

Bujang, M.A., Supramaniam, P., Haniff, J., Muninathan, P., Jalaludin, M.Y., Zain, F., et al. (2013). The Association between Self-Care Practices, HbA1c and Quality of Life for Type 1 DM in Children and Adolescents. *International Medical Journal*, 20(3), 337–339.

Burroughs, T.E., Desikan, R., Brian, M., et al. (2004). Development and validation of the Diabetes Quality of Life Brief Clinical Inventory. *Diabetes Spectrum*, 17, 41–49.

Cheng, A.Y., Tsui, E.Y., Hanley, A.J., & Zinman, B. (1999). Developing a quality of life measure for Chinese patients with diabetes. *Diabetes Research and Clinical Practice*, 46(3), 259–267.

Frederick, J.G., & Forzano, L.A.B. (2012). Research Methods for the Behavioral Sciences (4th ed.). Calif. Wadsworth, Belmont.

Fukuhara, S., Lopes, A.A., Bragg-Gresham, J.L., et al. (2003). Health-related quality of life among dialysis patients on three continents: The Dialysis Outcomes and Practice Patterns Study. *Kidney Int*, 64, 1903–1910.

Guillemin, F, Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol*, 46, 1417–1432.

Haley, S.M., McHorney, C.A., & Ware, Jr. J.E. (1994). Evaluation of the MOS SF-36 physical functioning scale (PF-10): I. Unidimensionality and reproducibility of the Rasch item scale. *J Clin Epidemiol*, 47(6), 671–684.

Huang, I.C., Liu, J.H., Wu, A.W., Wu, M.Y., Leite, W., & Hwang, C.C. (2008). Evaluating the reliability, validity and



minimally important difference of the Taiwanese version of the diabetes quality of life (DQOL) measurement. *Health and Quality of Life Outcomes*, 6(1), 87–90.

Hulley, S.B. (2007). *Designing Clinical Research*. Baltimore, Lippincott Williams & Wilkins.

Jacobson, A.M., Degrott, M., & Samson, J.A. (1994). The evaluation of two measures of quality of life in patients with type I and II diabetes. *Diabetes Care*, 16, 267–274.

Jalaludin, M.Y., Fuziah, M.Z., Hadhrami, M.Z., Hong, J.Y.H., Mohamad-Adam, B., & Jamaiah, H. (2013). Reliability and Validity of the Malay Translated Version of Diabetes Quality of Life for Youth Questionnaire. *Malays Fam Physician*, 8(1), 12–19.

Keller, S.D., Ware, J.E., Bentler, P.M., Aaronson, N.K., Alonso, J., Apolone, G., et al. Use of structural equation modeling to test the construct validity of the SF-36 Health Survey in ten countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol*, 51, 1179–1188.

Khaled, M., Elaziz, A., & Bakr, I.M. (2008). Assessment of knowledge, attitude and practice of hand washing among health care workers in Ain Shams University hospitals in Cairo. *The Egyptian Journal of Community Medicine*, 26(2), 1–12.

Kovacs, M. (1992). Children's Depression Inventory (CDI) manual 1<sup>st</sup>(Eds.). Multi-Health. North Tonawanda, NY, Systems Inc.

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukophadhyay, T., & Scherlis, W. (1988). Internet paradox: A social technology that reduces social involvement and psychological well-being? *Am Psychol*, 53(9), 1017–1031.

Lawshe, C.H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28, 563–575.

Levy, K.N., Becker, D.F., Grilo, C.M., Mattanah, J.J.F., Garnet, K.E., Quinlan, D.M., et al. (1999). Concurrent and predictive validity of the personality disorder diagnosis in adolescent inpatients. *Am J Psychiatry*, 156(10), 1522–1528.

Lim, C.J., Ab-Hamid, S.A., Yaacob, N.M., Mohd-Hairon, S., Kueh, Y.C., & Bujang, M.A. (2018) Validation of Malay version of body self-image questionnaire-short form

among Malaysian young adults. *Malays J Med Sci*, 25(4), 131–141.

Liu, X., & Boone, W.J. (2006). Applications of Rasch Measurement in Science Education (Eds.). Maple Grove USA, AM press.

Lovibond, S.H, & Lovibond, P.F. Manual for the Depression Anxiety Stress Scales. 2nd ed. Psychology Foundation; Sydney: 1995.

Murphy, K.R., & Davidshofer, C.O. (2005). Psychological testing: principles and applications, 6<sup>th</sup>(Eds.). N.J. Upper Saddle River, Pearson/Prentice Hall

Musa, R., Fadzil, M.A., & Zain, Z. (2007). Translation, validation and psychometric properties of Bahasa Malaysia

version of the Depression, Anxiety and Stress Scales (DASS). *ASEAN J Psychiatr*, 8(2), 82–89.

Nisbett, R.E. (2003). *The Geography of Thought: How Asians and Westerners Think Differently...and Why*. New York, The Free Press.

Nur-Khairul-Bariyyah, M.H., Bujang, M.A., Baharum, N., Mastura, I., & Shah, S.A. (2018). Self-Care Activities among Diabetic Patients and Factors Affecting Glycaemic Control in Primary Health Care, Malaysia. *J Diabetes Clin Prac* 1, 103.

Omar, K., Musa, R., Hanif, J., & Azimah, N. (2010). Reliability of Bahasa Malaysia version of family environment scale and its measurement issues. *ASEAN Journal of Psychiatry*, 11, 1.

Omar, K., Bujang, M.A., Mohd-Daud, T.I., Abdul-Rahman, F.N., Loh, S.F., Haniff, J., et al. (2011). Validation of the Malay Version of Adolescent Coping Scale. *Int Med J*, 18(4), 288–292.

Patricia, D., Diane, M.N., Stanley, M.A., & Patricia, A. (2002). The Depression Anxiety Stress Scale-21: Spanish translation and validation with a Hispanic sample. *J Psychopathol Behav Assess*, 24, 195–205.

Osborne, J.W., & Costello, A.B. (2004). Sample size and subject to item ratio in principal components analysis. *Pract Assess Res Eval*, 9, 1–15.

Pakpour, A.H., Saffari, M., & Burri, A. (2012). Translation and validation of an Iranian version of the diabetes quality

of life measure. *Journal of diabetes investigation*, 3(5), 471–478.

Reviriego, J., Millan, M.D., & Millan, M. (1996). Evaluation of the diabetes quality-of-life questionnaire in a Spanish population. An experience of translation and reliability. *PharmacoEconomics*, 10(6), 614–622.

Sato, F., Mita, T., Yamamoto, R., Hirose, T., Ito, C., Tamura, Y., et al. (2014). Reliability and validity of the Japanese version of the Diabetes Quality-Of-Life questionnaire for Japanese patients with type 2 diabetes mellitus. *Diabetol Int*, 5(1), 21–29.

Tabachnick, B.G., & Fidell, L.S. (2001). Using Multivariate Analysis. Boston, Allyn and Bacon.

Tan, S.M.K., Loh, S.F., Bujang, M.A., Haniff, J., Abd-Rahman, F.N., Ismail, F. et. al. (2013). Validation of the Malay Version of children's depression inventory. *Int Med J*, 20(2), 188–191.

The DCCT Research Group. (1988). Reliability and Validity of a Diabetes Quality-of-Life Measure for the Diabetes Control and Complications Trial (DCCT). *Diabetes Care*, 11(9), 725–732.

Ting, J.C., Tarn, Y.H., Ching, L.H., Liou, W.S., Shaw, J.W., & Chiou, X.G. (2007). Taiwanese version of the EQ-5D: Validation in a representative sample of the Taiwanese population. *J Formos Med Assoc*, 106(12), 1023–1031.

Toobert, D.J., & Glasgow, R.E. (1994). Assessing diabetes self-management: the summary of diabetes self-care



activities questionnaire. In: Bradley, C. (ed.). Handbook of Psychology and Diabetes (pp. 351–375). Switzerland, Harwood Academic.

Ware, J.E., & Sherbourne, C.D. (1992). The MOS 36-item short-form health survey (SF-36) 1: conceptual framework and item selection. *Med Care*, 30, 473–483.

Wild, D., Grove, A., Martin, M., et al. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) Measures: report of the ISPOR task force for translation and cultural adaptation. *Value Health*, 8, 94–104.

Yildirim, A., Akinci, F., Gozu, H., Sargin, H., Orbay, E., & Sargin, M. (2007). Translation, cultural adaptation,

cross-validation of the Turkish diabetes quality-of-life (DQOL) measure. *Quality of Life Research*, 16(5), 873–879.

Yunus, A., Seet, W., Mohamad-Adam, B., & Haniff, J. (2013). Validation of the Malay version of Berlin questionnaire to identify Malaysian patients for obstructive sleep apnea. *Malays Fam Physician*, 8(1), 3–9.

# About the Authors

## **Dr. Mohamad Adam Bujang**

Dr. Mohamad Adam Bujang is a Research Officer at the Institute for Clinical Research (ICR), Ministry of Health, Malaysia. Having received his Bachelor's degree in Statistics from MARA University of Technology in 2004, Dr. Mohamad Adam undertook his postgraduate study, receiving his Master's degree (MBA) in 2008, and then at the same university, receiving his PhD (Information Technology and Quantitative Sciences) in 2017. Currently, he is working as a biostatistician and researcher in the Clinical Research Centre, Sarawak General Hospital. His ongoing research interests involve the research methodology such as questionnaire development, questionnaire validation, sampling techniques, sample size planning, and statistical analyses. He has completed many research projects with local and international publications. He is actively involved in research training, a reviewer for a number of peer-reviewed journals and also one of the former and active members of the Malaysia Institute of Statistics (ISM) and Association of Clinical Registries Malaysia (ACRM).

### *Contact details:*

Dr Mohamad Adam Bujang  
Research Officer,

Clinical Research Centre, Sarawak General Hospital

Email address: [adam@crc.gov.my](mailto:adam@crc.gov.my)

Tel: 082-276820 & Fax: 082-276823

## **Mr. Hon Yoon Khee**

Mr. Hon Yoon Khee is a pharmacist working in the Institute for Clinical Research (ICR), Ministry of Health, Malaysia since May 2008 until now. He graduated with a Bachelor of Pharmacy degree from Otago University in New Zealand in December 1998, and had since been practising in community pharmacy and hospital pharmacy for several years before joining the Institute for Clinical Research, Ministry of Health in 2008. He also studied part-time clinical pharmacy from September 2007 until August 2012, and was awarded a Graduate Diploma in Clinical Pharmacy by the University of South Australia in August 2012.

### *Contact details:*

Mr. Hon Yoon Khee  
National Institutes of Health (NIH)  
Ministry of Health Malaysia  
Email address: yoonkheehon@gmail.com  
Tel: +603-33628407

## **Dr. Lee Keng Yee**

Dr. Lee Keng Yee is a Research Medical Officer at the NIH Manager office, National Institutes of Health (NIH), Ministry of Health, Malaysia. She received a Medical Degree from University Putra Malaysia in 2012. Dr. Lee worked in the National Clinical Research Centre (currently known as Institute for Clinical Research) from 2016 to 2018. Dr. Lee received training under the Children's Mercy Certificate Program in Pediatric Bioethics in 2016 and was appointed as the Secretary of Medical Research & Ethics Committee (MREC), Ministry of Health Malaysia in 2018. She is currently working in the Sector for Ethics and Research Surveillance of NIH, Malaysia. Her ongoing research interests include research methodology and bioethics. She is actively involved in research ethics training such as Good Clinical Practice (GCP) and Good Research Practice (GRP).

### *Contact details:*

Dr Lee Keng Yee

Sector of Ethics and Research Surveillance

NIH Manager Office

National Institutes of Health (NIH)

Ministry of Health Malaysia

Email address: leekengyee@crc.moh.gov.my

Tel: +603-33628407

# About the Publisher



## **Institute For Clinical Research,**

National Institutes of Health (NIH), Ministry of Health  
Malaysia, Block B4, No.1, Jalan Setia Murni U13/52,  
Seksyen U13, 40170 Shah Alam, Selangor Darul Ehsan,  
Malaysia.

Phone: 603-3362 7700 | Fax: 603-3362 7701

Email: [contact@crc.gov.my](mailto:contact@crc.gov.my)





Most validated questionnaires that are widely used around the world are created in a non-native language. If such a questionnaire is to be used in a native-speaking population, it must first be validated. This book aims to be a useful guide for describing all the steps required for validating a questionnaire. This book also provided a standard checklist with 17 items that are recommended to serve as a useful guide for researchers conducting questionnaire validation research.

Institute For Clinical Research  
National Institutes of Health (NIH)  
Ministry of Health Malaysia  
Block B4, No.1,  
Jalan Setia Murni U13/52,  
Seksyen U13, 40170 Shah Alam,  
Selangor, Malaysia  
Phone: 603-3362 7700  
Fax: 603-3362 7701  
Email: [contact@crc.gov.my](mailto:contact@crc.gov.my)