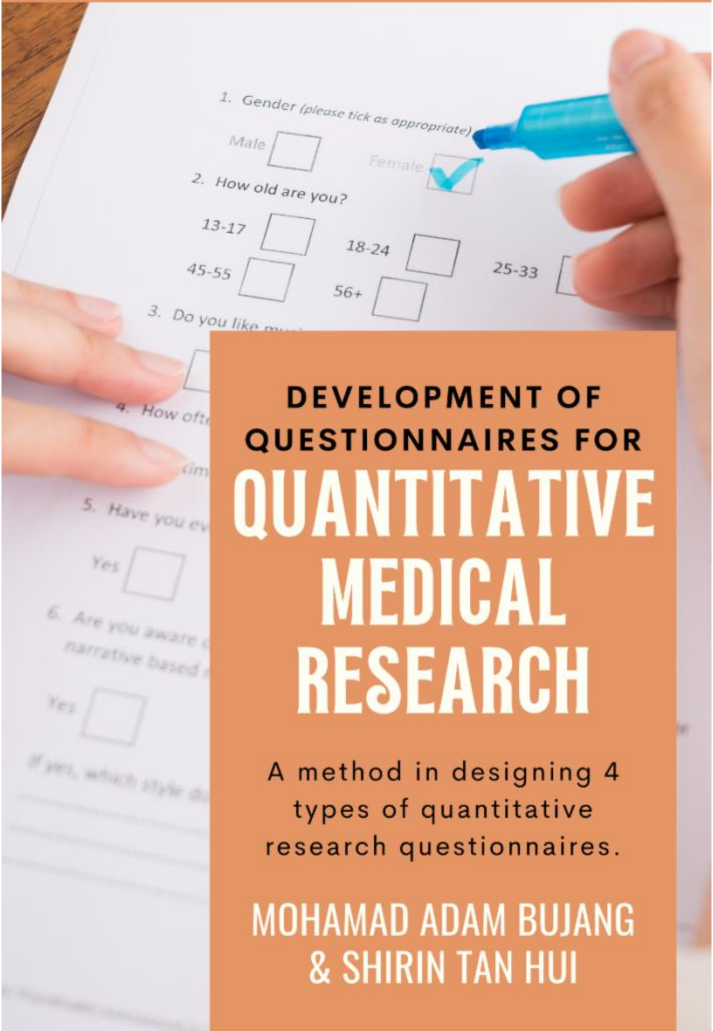
INSTITUTE FOR CLINICAL RESEARCH



DEVELOPMENT OF QUESTIONNAIRES FOR QUANTITATIVE MEDICAL RESEARCH

A method in designing 4 types of quantitative research questionnaires.

MOHAMAD ADAM BUJANG & SHIRIN TAN HUI

Development of Questionnaires for Quantitative Medical Research

Authors
Mohamad Adam Bujang
Shirin Tan Hui

Editors
Prof. Shamsul Azhar Shah
Prof. Jamaluddin Ab Rahman

Development of Questionnaires for Quantitative Medical Research

DOI: doi.org/10.5281/zenodo.6137926
ISBN: 9789671694053(MY)

Published by
Institute for Clinical Research,
National Institutes of Health, Ministry of Health
Malaysia

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

Madam Shirin Tan Hui
Clinical Research Centre, Sarawak General Hospital
Institute for Clinical Research
Ministry of Health Malaysia

Editors
Prof. Dr. Shamsul Azhar Shah
Department of Community Health,
Faculty of Medicine,
Universiti Kebangsaan Malaysia

Prof. Dr. Jamalludin Ab Rahman
Department of Community Medicine
Kulliyyah (Faculty) of Medicine
International Islamic University Malaysia

Preface

A questionnaire is an important tool to collect information in research. The quality of the questionnaire influences the success in data collection and eventually help to answer the study objective(s). This present book discusses the overall approach in developing four types of questionnaires for quantitative research namely (a) Questionnaire that is designed to collect facts or information, (b) Questionnaire that is designed to measure knowledge, (c) Questionnaire that is designed to measure latent variables, and (d) Questionnaire that is designed to screen or diagnose a subject. The process starts from understanding of the objective(s) of study, content and scope of study, knowledge exploration, item development and ultimately testing the reliability and validity of the questionnaire. This book aims to serve as a quick guide for researchers in developing a quality questionnaire for their survey research.

Adam & Shirin

November 2021

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 Mr Hon Yoon Khee for his proofreading of the book and Dr Chew Cheng Hoon for helping in the process to publish this book.

Abbreviations

No.	Abbreviations	Definition
1	5W1H	Five interrogative words such as what, where, when, why and how
2	AGFI	Adjusted goodness-of-fit test
3	AHI	Apnoea-Hypopnea Index
4	BP	Blood pressure
5	CFA	Confirmatory factor analysis
6	CFI	Comparative fit index
7	DASS	Depression, anxiety, stress scale
8	DQOL	Diabetes Quality of Life

9	EFA	Exploratory factor analysis
10	GFI	Goodness-of-fit index
11	MOS-SF36	Medical outcome study – Short form with 36 items
12	KAP	Knowledge, attitude and practice
13	OSA	Obstructive sleep apnoea
14	RMR	Root-mean-square residual
15	RV-DQOL13	Revised version of Diabetes Quality of Life with 13 items
16	SDSCA	Summary Diabetes Self-Care Activities
17	WHO	World Health Organization

Table of Content

List of Tables

List of Figures

List of Appendices

Chapter 1

Background

A questionnaire is often the preferred platform for eliciting relevant information from a target population or its sample. Hence, it is defined as “a set of printed or written questions with a choice of answers, devised for the purposes of a survey or statistical study” (Oxford Dictionary, 2018). In other words, a questionnaire can also be referred as a set of relevant questions that are assembled together in a digital (soft copy) or nondigital (hard copy) format aiming to collect information from the target study subjects (who are also the respondents) for record-keeping and/or research purposes.

The quality of a questionnaire is vital for enabling the researcher to collect important and relevant information in an efficient and effective manner, because the use of a questionnaire for collecting information can result in a waste of time and effort if it is not well-designed. Therefore, proper steps are necessary to be taken in order to develop a high-quality questionnaire especially for research purposes. Therefore, this book aims to serve as a guide for undertaking questionnaire development studies in medical research. Despite the wide availability of research publications in the field of questionnaire development, a thorough discussion of the broader overall scope of the questionnaire development process particularly within the medical discipline is still lacking and can therefore be further enhanced (McColl and Thomas, 2000; Wilkinson and Birmingham, 2003; Boynton and Greenhalgh, 2004; Rattray and Jones, 2007). This underpins the purpose of this book, which aims to fill in the gap by compiling a very useful list of practical considerations when developing a scientifically valid questionnaire for collecting data in research.

Since research in the medical discipline mostly deals with data regarding patients and their medical condition (including their diseases, treatment modalities offered and treatment outcomes), thus the development of a questionnaire will need to follow a careful list of pre-defined steps which are planned out thoroughly as each of these steps will have a direct impact on the quality of data collected. Furthermore, the existence of a wide variety of different research designs for medical research adds complexity to this process. Generally, the process of questionnaire development starts with understanding the problem statement of a research proposal (through a review of relevant literature, and/or a discussion with subject matter experts), which shall next lead to the setting of research or survey objectives, and then to the formulation of an overall design of a questionnaire, which includes the planning and execution of an appropriate research design, and then finally to the assessment of its overall reliability and validity.

Besides, this book also places an emphasis on the questionnaire development process for medical research by providing a list of instructions for a researcher to follow when tailoring to the specific requirements for designing each of the various types of questionnaires. Although the majority of the examples cited in this book have been obtained from the published medical literature, the generic questionnaire development process can often be applied in many other fields. Hopefully, this book can provide novice researchers with valuable guidance for enabling them to design and validate a new questionnaire. A checklist which consists of a list of important pre-defined steps will be provided in the final chapter of this book to guide researchers in a step-by-step manner for conducting research in questionnaire development studies. However, it must be noted that this book caters for the development of a questionnaire for quantitative studies only; and hence it may not be applicable for developing a questionnaire in qualitative research, since qualitative research will involve the use of a set of different approaches for designing, developing and validating a questionnaire.

1.1 To obtain a clear understanding of the research topic

The first step for conducting a research study is to obtain a clear understanding of the background subject matter for the topic of interest. A comprehensive literature review shall initially be undertaken by the researcher in order to obtain an overview of all existing research work being conducted in this area. Likewise, a preliminary literature review will also be necessary for conducting a questionnaire development study. Below are some questions that a researcher should focus on answering when performing a literature review.

- What is the scope of my research?
- What is/are the study objective(s)?
- What is the current state of knowledge and/or latest updates on the subject matter of interest?
- What are the key findings and/or lessons learnt from previous research on a similar topic or study design?

- How will the new study findings add to current repertoire of knowledge in this area of interest?
- Based on the researcher’s capacity, will this study be feasible?
- If preliminary research reports for a benchmark study (i.e. any published research papers on the topic of interest with a similar scope) are available; then based on these reports, what approach should the researchers be adopting when conducting this new study?

- What are the potential and/or actual impact of this new study?

Apart from performing a literature search, the researcher will also need to hold a brainstorming session with other investigators, and to discuss with subject matter experts from the relevant field, or to conduct interview with potential respondents and to make certain observations about the area of interest to identify any knowledge gaps. This means that a researcher will have to gain relevant knowledge in the subject matter, and to review all relevant research work done on the subject, and then to discuss with the other investigators until he/she has acquired a clear understanding of the right direction in which a new study should be heading towards. It is therefore advisable for him/her to formulate a clear and answerable research objective(s) before embarking on a new study. It is not surprising that a researcher may occasionally have to suspend a study by taking one major step backward for making a major revision of the study protocol in either its study design or methodology (or both) midway through conducting a new study, probably due to a major flaw arising from inadequate literature review and/or research planning. This will lead to a waste of time spent and effort made by both investigators and study subjects.

In a typical questionnaire development study, a researcher should know what questions should be considered as essential and what is the most appropriate method to frame the question in the questionnaire for eliciting the answers to these questions. For example, a research team who wishes to develop a quality of life (QoL) questionnaire should carefully study the definition, underlying theory and the overall concept of quality-of-life (QoL). As the theory of QoL may evolving, it is thus necessary to perform a careful review of both historical and contemporary scientific literature on QoL.

A QoL questionnaire has been studied for many decades and there are many well-developed instruments such as MOS-SF36, EQ5D and WHOQOL-BREF, (Ware & Sherbourne, 1992; EuroQol, 1996; The WHOQOL Group, 1998). Therefore, one of the questions which a researcher can potentially pose is that whether it is necessary to develop another QoL measure. In addition, a series of other related questions will often be posed by a researcher as well. These include: ‘What is the basis for developing a new QoL measure?’, ‘What are the benefits which can accrue from this new QoL measure?’, ‘Is it really necessary to develop a new QoL measure?’, ‘Will the domains in the existing questionnaire have to be further enhanced?’, ’What are the weaknesses and/or limitations of the previous QoL measures?’, ’What possible advantages will this newly developed QoL instrument offer?’, etc..

All these questions should ideally be carefully considered by the research team before they make a decision for proceeding with the new study, in order to be able to provide a valid justification for it.

Key messages:

- *Understanding the subject matter is an important pre-requisite before embarking on a new scientific research study.*

- *This understanding can be acquired through a careful review of the literature, or via holding a brainstorming session among the investigators, or engaging in fruitful discussion with other subject matter experts, or via conducting open-ended interviews with potential* *respondents and making certain observations about the subject matter.*

- *Once the researcher has acquired a clear understanding of the subject matter, he/she should be able to define their proposed scope of study, the tentative study objective(s) and will thereby be able to select an appropriate research design.*

Chapter 2

Types of Questionnaires

This book introduces four types of questionnaires, namely: a questionnaire is designed to (1) retrieve facts or information, (2) measure knowledge, (3) measure latent variables and (4) screen for or diagnose a medical condition. To the best of our knowledge, the categorization of four different types of questionnaires has not been described or emphasized in the existing literature. This categorization of questionnaires will be introduced in this book, which is based on an explicit understanding of the approach taken by scholars in developing and validating various types of questionnaires, especially within the existing literature on medical research.

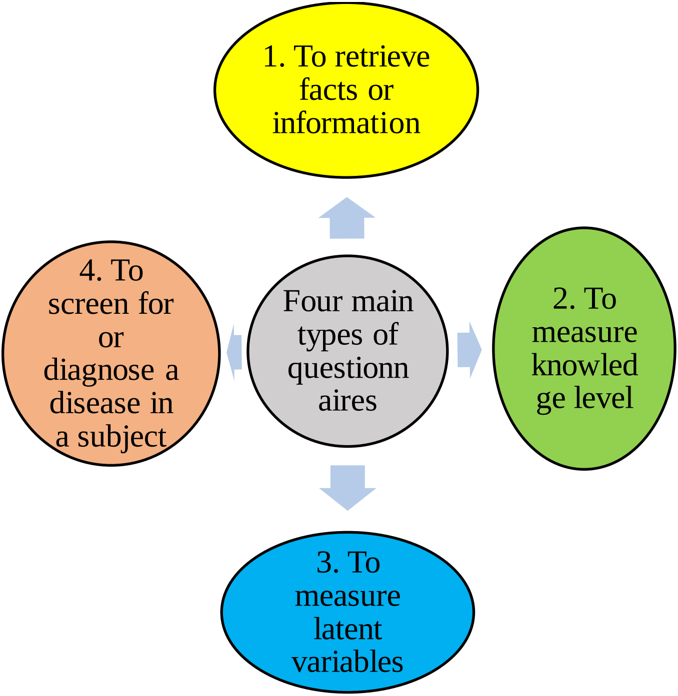


Figure 2.1: Four main types of questionnaires

2.1 Questionnaire that is designed for collecting facts or information

Fact or information about an individual belongs to actual data which describes the true characteristics of an individual. Examples include any information that are related to the demographic profile of the study respondents such as gender, age and ethnicity. Besides, such information can also be related to their socio-economic class such as their monthly income or medical history, such as co-morbidities, treatment outcomes and/or complications for a medical condition (NHAM, 2015). As long as such information is depicting the true characteristics of a study subject, then the questionnaire that is designed this type of information shall fall under this category.

This category of questionnaire is widely utilised for collecting data for record-keeping purposes in various fields of research, such as in the medical field where specific details of individual patients are being collected; or for administrative purposes where details of staff and customers and/or particular information related to products or services such as brands, or product specifications will be collected; along with many others. Although such data will initially be collected for future reference only; however, after having accumulated a large pool of data, they can subsequently be extracted and then analysed further, in order to derive useful information by observing any trends and/or patterns which possibly emerge from such data. Indeed, such information can often be useful for guiding the managers or researchers in their future planning and/or making important decisions.

In an established registry such as a patient registry or a disease registry, patient information can often be used for research purposes, although they are initially collected for the purpose of monitoring their disease progression (NHAM, 2015). The usual approach for asking the questions for this type of questionnaire are simply stated without having to resort to using an interrogative approach such as asking “what”, “when”, “who”, “why”, “where” and “how” (5W1H) questions. These questions are often straightforward and supposed to be easily understood by the study subjects or study administrators. For those complicated surveys like a patient registry, the questionnaire will usually be accompanied by a variable dictionary which provides each and every variable with a specific definition (NHAM, 2015).

2.2 Questionnaire that is designed to measure the level of knowledge

The second type of questionnaire is one which aims to measure the level of knowledge attained by students or study respondents in a specific module of field. For example, the standard examination questionnaires are designed for formal academic purposes to evaluate the academic performance of students from various education levels. On the other hand, there is also a questionnaire that aims to measure the level of knowledge attained by a study respondent in a specific subject, such as the level of knowledge about antibiotics among medical officers or the patients' level of knowledge regarding diabetes mellitus (Thriemer et al., 2013; Gulabani et al., 2008). Hence, this type of questionnaire aims to measure a student's or a study respondent's performance in their level of knowledge attained, rather than for the purpose of keeping such records for future reference.

The guidelines presented in this book are not meant to measure the level of knowledge attained by a study respondent, as shown by an examination questionnaire for the purpose of assessing academic performance. Different academic institutions or organizations may have their own processes in developing and validating the assessment criteria for such examination papers. Conventionally, the set of questions specifically designed for examination papers will vary in each examination to prevent these students from focusing on how to spot the potential examination questions instead of truly mastering the subject by properly learning and assimilating the teaching materials. Therefore, the scope of this book is merely to measure the magnitude of knowledge in a specific subject matter which has been mastered or acquired by a specific population; since its ultimate aim is usually to improve an existing intervention or to introduce a new intervention, apart from measuring the knowledge level of a subject which has been acquired by a specific population.

Some of these ‘knowledge’ questionnaires can be fixed or just slightly modified from the existing sets of questionnaires. This type of ‘knowledge’ questionnaire is usually designed for testing the level of knowledge attained by the customers, patients and other prospective study respondents with the ultimate intention of improving an intervention, product or service rather than ranking the study subjects based on their knowledge (Berikai et al., 2007; Wirtz et al., 2009). For example, knowledge about antibiotic usage can often be necessary to be assessed among patients so that proper counselling and/or consultation on judicious antibiotic use can then be provided to the patients.

This type of questionnaire is usually self-administered and the knowledge level will be assessed by applying the 5W1H questions. The response can be elicited in a binary manner (i.e. Yes or No), or via answering a multiple-choice question, or by matching, or by assigning a score and by answering with an open-ended text. When answering the questions, the study subjects will need to provide definite answers based on their current level of knowledge. The most important characteristic of a 'knowledge' questionnaire is that the questions should only relate to facts or common understanding of an actual state or condition. Thus, the answers for all the questions should be made available in peer-reviewed published materials such as textbooks or from published guidelines by reputable institutions including the World Health Organization (WHO) (WHO, 2018).

2.3 Questionnaire that is designed to measure latent variables

The third category of a questionnaire is one which is designed to measure one or more latent variables. A latent variable is a variable that is almost impossible to be measured by using any machine or device. In other words, the latent variables are variables inferred from other variables that can be observed directly (Bollen, 2002). For example, there is no device that measures health-related quality of life of a study subject and therefore, the feedback provided by the study subjects will be derived from their answers to a set of questions which are currently used to measure their health-related quality of life. Based on Figure 2.1, there are eight latent variables that represent the measure for health-related quality of life which consists of physical activities, social activities, role activities due to physical problems, bodily pain, mental health, role activities due to emotional problems, vitality, and the general health condition (Ware and Sherbourne, 1992).

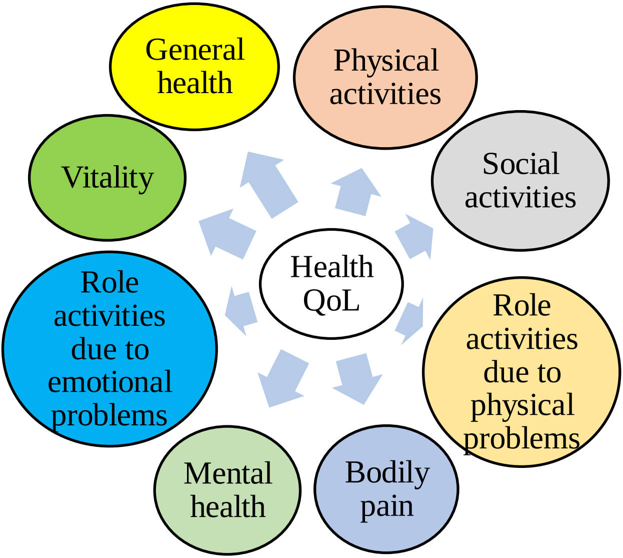


Figure 2.2: An example of latent variables to represent a measure of health-related quality of life

Questionnaires that are designed to measure latent variables are the most difficult to develop and validate because the measurement of latent variables can often be subjective and how the study subjects can perceive the parameters is likely to differ between individuals. 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 and Sherbourne, 1992; Lovibond and Lovibond, 1995;

Toobert and Glasgow, 1994); which are widely used for measuring latent variables.

The response to be elicited by this type of questionnaire is usually measured by a Likert scale. Such questionnaires are validated by using exploratory factor analysis (EFA) which will sometimes be followed by structural equation modelling (SEM). Both are statistical techniques to help researchers in dividing many items or questions into several possible and logical domains or factors. Such domains can later be defined to represent a specific characteristic of the population, and so the next chapter provides the recommended steps for designing a questionnaire that aims to measure the magnitude of a specific parameter.

2.4 Questionnaire that is designed to screen or diagnose a subject

Diagnosing a medical condition in a study subject is very common in the medical field. The medical condition in these study subjects can often be diagnosed by using methods like a structured interview such as one that is usually conducted by psychiatrists to diagnose depression. On the other hand, in a variety of many other instances, the diagnosis can often be confirmed by using a specialized machine. For example, the gold standard for diagnosing obstructive sleep apnea (OSA) is the attended overnight level I polysomnogram (PSG). It requires the patient to be admitted overnight at a medical facility which has a sleep study centre. Polysomnography is used to measure the apnoea-hypopnoea index (AHI) and various other parameters; and so the severity of OSA will be measured with the AHI, which is the number of apnea and hypopnea events per hour of sleep. AHI reading exceeding 5 indicates that a patient has OSA (Young et al., 2002).

However, these procedures (i.e administering a structured interview or using a complicated device) are often time-consuming and costly especially when many subjects are suspected to have depression or OSA, and it is necessary to perform a diagnosis for each subject. Therefore, questionnaires are specifically developed by scholars to help screen subjects for conditions such as depression and OSA (Netzer et al., 1999; Chung et al., 2008). This will save time, manpower and cost since only those study subjects that potentially have a higher risk which will be detected by the screening questionnaires will then be required to undergo further diagnostic tests for confirmation. This type of screening questionnaire is important for complementing the existing diagnostic procedures.

This type of questionnaire is also related to facts, science or knowledge. The outcome has to be scientific with a list of distinct criteria and which must be answered in a binary manner such as ‘diseased’ or ‘not diseased’ (i.e ‘depressed’ versus ‘not depressed’). All the questions posed by this type of questionnaire will need to link and directly correspond to the outcome. Hence, the responses to all the questions will have to be supported by research or scientific evidence. These responses will usually be transformed into a score and then be divided into two categories such as ‘with disease’ or ‘without disease’. The accuracy of this type of questionnaire will be tested and considered valid if the performance of the questionnaire in detecting a ‘positive’ or ‘negative’ event of interest is substantially high although its accuracy of such a detection is often not as high as the gold standard for such a measurement.

Ideally, a screening questionnaire is usually not designed to substitute the gold standard for this measurement (i.e. structured interview to diagnose depression or PSG machine to diagnose OSA) since it is often very unlikely for the performance of a ‘screening’ questionnaire to be as good as the gold standard in the detection of the event of interest (i.e. depression or OSA). However, this type of questionnaire will be developed with an intention to screen for a medical condition in the study subjects which can also be used as preliminary evaluation to rule out the ‘non-disease’ cases. This is a useful approach for minimizing the overall cost of performing a diagnosis and screening for a medical condition in a study subject especially when the procedure for diagnosing a case can often be too costly and time-consuming.

Chapter 3

Framing of Questions for a Questionnaire

In a survey research, asking appropriate questions and eliciting the ‘correct’ types of responses are two important elements for developing a versatile questionnaire. Both should complement each other and the ideal combination of the two will lead to a successful survey research by collecting the right information to answer the study objective(s).

3.1 The necessity to adopt the right approach for asking questions or obtaining information

This chapter discusses a variety of differing approaches in asking questions, eliciting the responses from the respondents and developing valid measurement scales. Generally, questions can often be developed in many various ways but overall it is always the prerogative for the researcher to choose the most appropriate way to frame the questions that aim to address the study objective(s). The fundamental tenet for developing a good questionnaire is to create straightforward and unambiguous questions that will both answer the research questions and address the study objectives. In addition, it is also necessary to decide beforehand the appropriate responses for each question. For instance, the researcher will have to decide whether it is more appropriate to elicit an open-ended response, or a binary response ("Yes" or "No") or a response based on a Likert-scale. This is because the researcher must ensure that the responses elicited from the study respondents will be able to provide enough information, which are packaged in a sufficient level of details which is of sufficient depth, and can be amenable to meaningful analysis.

In a qualitative survey research, questions are usually being framed by using interrogative words such as ‘What’, ‘Where’, ‘Who’, ‘Why’, ‘When’ and ‘How’, which belong to the type of questions that are designed to elicit subjective responses. Such an approach will enable the researcher to obtain in-depth information from the respondents. Apart from this, the researcher will need to understand how to analyse the respondents’ feedback by using methods such as thematic analysis. Quantitative analysis is rarely needed in qualitative research. Contrarily, in quantitative research, the types of responses elicited from the study respondents are very important because these quantitative data must be presented in way that allow them to be analysed by using appropriate statistics.

It is too lengthy to describe all the possible ways to choose an appropriate type of question for a questionnaire. Therefore, to keep things simple for a researcher, the following description provides two approaches which can often be used to frame a question in such a way that it will address the study objective for a survey research:

a. The ‘5W1H’ type of questions.

The 5W1H are the interrogative words that are used to ask questions such as “What”, “Where”, “Who”, “Why”, “When” and “How”. In addition, “whom” or “do/does” can also be used as interrogative words. Such questions will end with a question mark “?”. The choice of interrogative word shall depend on the type of information to be collected. Therefore, it will also be necessary to select an appropriate response which corresponds directly to the question. Generally, there are three main types of suitable responses to most questionnaires, namely: open-ended answers, dichotomous responses such as “Yes” or “No”, and multiple-choice type of answers.

b. The ‘non-5W1H’ type of questions.

It is possible for some data or information to be retrieved without using interrogative words. In such cases, the type of information to be elicited from study respondents is often simple or commonplace and can be easily understood by both researchers and respondents. An example of such information for a survey research is demographic profile of study respondents that can be asked without posing the ‘5W1H’ type of questions and the possible feedback options for them are usually designed to be a multiple-choice. For example, in a questionnaire, it is often sufficient to state “Gender” and then to provide a binary option which can either be (1) Male or (2) Female (see Figure 3.1). It is commonly understood that the respondent should answer by choosing one of the two options. In a patient registry, many questions can also be asked by adopting the same approach to collect clinical information such as type(s) of treatment offered, treatment complication(s) and outcome variable(s) (NHAM, 2015).

3.2 Tips for framing the questions in a questionnaire

Below are some of the useful tips for framing the questions in a questionnaire:

- *To ensure that all questions are designed to focus on addressing the study objectives*

  It is necessary to develop questions which are in line with the scope of a study and which can address the study objectives. So, an astute researcher will need to acquire a strong knowledge base of the subject matter in order to determine what question(s) should be posed.

- *To use simple language*

  The language to be used in a questionnaire should be easily understood by the study respondents because they will have to respond to each question. Thus, the researcher will need to gauge the level of knowledge and understanding of the respondent population.

- *To avoid double-barreled questions*.

  It is recommended to pose the question in such a way which will be asking for a single message at a time. For a more profound idea or a complex concept, it will be necessary to split it into multiple questions. For example, “Are your supervisor morally good and kind-hearted?” The problem with this question is being morally good is different from being kind-hearted.

- *To avoid leading/loaded questions*

  It is also advisable to phrase the question in such a way that it will not direct the respondent to choose any particular response(s). Such leading questions will often introduce bias to the results obtained from a questionnaire. For example, “Why did you have difficulty with online learning?” This question shall directly imply the answer which inadvertently leads to the assumption that online learning is the root of the problem. It also casts the blame on the user, which is a source of bias. Moreover, this question will become a loaded question if the respondent has no prior exposure in online learning; because a loaded question usually makes an assumption about the respondent and then imposes on them to select an answer which he/she will often not be familiar with.

- *To avoid double-negative questions*

Use of two negative words in a question can often confuse the respondents, who may misinterpret the question. For example, “Why you don’t do nothing?” “Don’t” and “nothing” are both negative words which can confuse respondents and then elicit a wrong response from them.

- *To avoid long questions*

  Long questions are often not likely to be focused and specific and therefore they tend to explore more than one thing at a time. The respondents may lose focus when they try to comprehend such questions and so they will end up having difficulty in understanding them and responding to them appropriately.

- *To refrain from asking questions on sensitive issues and to pose such questions only if really necessary*

  Avoid posing questions that are considered a sensitive issue for the community or the public. Questions exploring topics such as racism, sexuality and cultural differences may be deemed as sensitive for a study respondent. Such questions may potentially hinder the respondent from providing an honest, candid answer for their overt fear of judgment by the researchers or even worse, they may make them decide not to answer. Hence, it is imperative for the researcher to rephrase such questions very carefully and to consider an alternative way to elicit responses that pertain to sensitive issues. If deemed necessary, then such sensitive questions should be carefully and appropriately asking so that the respondent will not feel offended.

- *To understand who are the prospective respondents and their level of prior knowledge and understanding of the subject matter*

  First of all, researchers should not overestimate the respondents' capability and capacity to provide appropriate responses to the questions in a questionnaire. Do not pose any questions that are beyond the respondent's ability to comprehend and also do not over-estimate the knowledge level of a study respondent. Secondly, it is also necessary to refrain from making unfounded assumption(s). For instance, an elderly patient with type II diabetes may not necessarily have suffered from diabetic complications.

- *To give higher priority to those questions that are amenable to statistical analysis*

  For a quantitative research, the researcher will need to ensure responses collected are amenable to statistical analysis. Thus, the researcher will need to develop an appropriate measurement scale for determining the magnitude of these responses (which will be discussed later in this chapter) and then to apply such a scale when assessing these quantitative responses. It is therefore worth mentioning here that we must always bear in mind that a qualitative research may probably require a different approach when asking questions of a qualitative nature.

- *To provide a justifiable, relevant and valid response for each question*

  A good question should be paired with a good response option. Look at this question.

  What colour do you like most?

    - Blue ( )
    - Red ( )

    - Black ( )
    - Others ( )

  Ideally, the researcher should decide if the above three options are sufficient to be representative of colour preferences for most respondents.

- *To standardize the format for the expression of responses of questions*

  Some questions may require responses which are expressed in a standardized format. For example, the respondent can report a date in several different formats and therefore it will be easier for the researcher to standardize the format for the expression of this variable (such as “dd/mm/yyyy” or “dd.mm.yyyy”).

- *To exercise caution when dealing with questions that may require multiple responses*

  Some questions may have multiple answers to it. For example, common answers to the question “State your co-morbidities” are hypertension, diabetes mellitus, dyslipidemia and many more. Thus, the best way to elicit responses for this type of question is by providing an option for each possible co-morbidity, as shown in Figure 3.1.

Question:

Please mark "x" to indicate your status of having (or not having) co-morbidities in the bracket provided.

- Hypertension Yes (   )   No (   )
- Diabetes mellitus Yes (   )   No (   )
- Dyslipidemia Yes (   )   No (   )
- Others (Please state: ____)

Figure 3.1: Examples of questions with multiple responses

3.3 Types of responses and measurement scales

Regardless of whether the questions are using interrogative words or not, the responses provided by the respondents should always be amenable to analysis. This is the ultimate aim of a questionnaire, i.e. to collect relevant data or information from study respondents. The types of responses shall vary between each question, which will depend on the information intended to be collected by researchers. There are four types of measurement scales such as nominal, ordinal, interval and ratio. The brief explanation for each scale is as follows:

- **Nominal scale**

The nominal scale uses numbers to represent categorical data. Examples of nominal scales are shown in Figure 3.2. These numbers should not be added, subtracted or multiplied as it is merely a coding for the categorical data it represents. There is also no ascending or descending order for the nominal scale.

Instruction: Mark "x"

on each question below.

Gender

Male (x)

Female (   )

Ethnicity

Malay (   )

Chinese (x)

Indian (   )

Code's definition:

Gender:

Male = 1

Female = 2

Ethnicity:

Malay = 1

Chinese = 2

Indian = 3

Figure 3.2: Examples of questions where the responses are expressed using the nominal scale

- **Ordinal scale**

In ordinal scale, numbers represent the ranking and order of the data without a clear degree of variation between the ranks. Ordinal data are quantitative data of naturally occurring ranks to provide information about ‘more than’ or ‘less than’. Therefore, ordinal data can provide a measure of magnitude of non-numeric characteristics such as satisfaction, preference and happiness. One commonly used example of ordinal scale is the Likert scale that aims to describe for instance respondent’s satisfaction level, ranging from e.g. ‘Very Satisfied’ to ‘Very Unsatisfied’.

Likert scale can be a five-point, seven-point or even ten-point scale which is used by the study respondents to express their level of agreement to a particular statement. Instead of a simple ‘yes/no’ answer, the Likert scale allows the respondent to have varying magnitude of opinions. One possible downside of this scale is that respondents may not answer truthfully due to social desirability (or expectations for social norms) and hence, compromising the validity of data collected. For example, questions derived from the Revised Version of DQOL (RV-DQOL13) are formulated with responses which are expressed in a 5-point Likert scale to explore the study respondents’ satisfaction (Bujang et al., 2018) (see Figure 3.3).

The Revised Version of DQOL (RV-DQOL13) Questionnaire

Please read every statement carefully and circle the number that best describe your feeling or situation.

Sila baca setiap penyataan dengan teliti dan bulatkan nombor yang paling sesuai untuk anda.

1= Very satisfied/*Sangat puas hati*
2=Moderately satisfied
3=Neither satisfied nor dissatisfied
4=Moderately dissatisfied
5= Very dissatisfied/*Sangat tidak puas hati*

1. How satisfied are you with the amount of time it takes to manage your diabetes? *Sejauh manakah anda berpuas hati dengan jumlah masa yang digunakan untuk menguruskan diabetes anda?*	1	2	3	4	5
2. How satisfied are you with the amount of time you spend getting check ups? *Sejauh manakah anda berpuas hati dengan jumlah masa yang anda gunakan untuk mendapatkan pemeriksaan doktor?*	1	2	3	4	5

Figure 3.3: Examples of questions where the responses are expressed using the ordinal scale

- **Interval data**

In interval scales, the numbers are aligned with each other on a continuum and the division between each number has an equal difference. This is unlike in ordinal data, where the gaps between the numbers are usually not identical. For example, when measuring temperature, the 10-degree Celsius difference between 20 and 30 is the same as the 10-degree Celsius difference between 80 and 90. The zero in interval scales represents a value, rather than a true zero. Examples of interval scales are the Celsius and Fahrenheit temperature scales, in which a zero degree does not indicate an absence of temperature. Data that are presented on the interval scales can be expressed as a mean when analysing the data. Based on Figure 3.4, the responses to the questions are considered to be expressed as an interval scale since the intervals between each number are identical where the difference between each option in the above example is equal to 1 hour.

Instruction: Circle the best answer according to your own eating habit

1. Normally at what time will you be having your breakfast?

b. 6.00am

c. 7.00am

d. 8.00am

e. 9.00am

f. 10.00am

Code's definition:
6.00am = 1
7.00am = 2
8.00am = 3
9.00am = 4
10.00am = 5

Figure 3.4: Example of question where the responses are expressed and measured using the interval scale

- **Ratio**

Ratio scale is similar to interval scale, except that it has a true zero which indicates a complete absence of the item on the measurement scale. Examples of ratio scales are height, weight and money. A reading of zero in height scale indicates non-existence. Ratio scale also provides information about the relative magnitude of scores. For instance, it can be interpreted that 20kg is twice as heavy as 10kg. Based on Figure 3.5, all the six questions require responses to be expressed in the ratio form. However, the ratio scale does not need any coding unlike nominal, ordinal and interval scale because the values refer to the exact measurements obtained from a study subject. The measurement units for such questions must be clearly specified in order to prevent inconsistency in eliciting responses from the study respondents and also to obviate the need for the researchers in making assumptions when reviewing the responses.

Instruction: Please answer all questions

1. Age: _______ years
2. Height: _______ cm
3. Weight: _______ kg
4. BMI: _______ kg/m^2
5. BP systolic: _______ mm Hg
6. BP diastolic: _______ mm Hg

Figure 3.5: Example of questions where the responses are expressed and measured using the ratio scale

For the sake of brevity, it can be summed up by saying that as long as the researchers have been equipped with the knowledge on the four different measurement scales, they will be able to design questions that can be analysed, particularly for a quantitative research study. This will also help them to identify appropriate statistical tests for performing subsequent data analysis. For example, frequency along with percentage data (%) is suitable for nominal and ordinal data; while mean with standard deviation is suitable for interval and ratio scales (provided that such data are normally distributed).

Despite the above contention, we must emphasize that it is not our proposition that open-ended answers are not allowed for a questionnaire and therefore should not be condoned. In fact, open-ended answer plays a very important role in capturing detailed information and in exploring new concepts or knowledge areas. For a medical survey, the detailed information captured from the open-ended question may not be analysed directly but the information can be kept as a reference for future use if necessary. If a researcher plans to analyse feedback obtained from open-ended questions, he/she will need to convert these feedbacks to categories or multiple-choice responses. For example, instead asking “State your co-morbidities” and provide an open-ended answer, researchers can convert the response to multiple questions with a close-ended ‘Yes/No’ response. To visualize this, please refer to Figure 3.1.

Chapter 4

Step-by-step Procedures for Questionnaire Development

Ideally, the recommended process for questionnaire development shall include (1) understanding the content and scope of the subject matter, (2) designing the structure and format of a questionnaire, (3) evaluating for both its content and face validity, (4) conducting a pilot study and (5) doing a fieldwork study to assess the reliability and validity of the questionnaire. Before starting to develop a new questionnaire, researchers will need to pre-specify the ultimate aim of a new questionnaire by answering the following questions, such as:

(i) Is this questionnaire going to collect only facts and information, or

(ii) Is it a knowledge-based type of questionnaire which is designed to assess the level of knowledge of respondents, or

(iii) Is it a questionnaire designed to measure latent variables or

(iv) Is it a questionnaire designed for screening/diagnostic purposes?

This is an important step because it is must be made absolutely clear on the specific nature and type of questionnaire to be developed at first instance since this will determine the optimal approach to be applied for developing a questionnaire.

4.1 Step 1: To explore knowledge and understand in depth the chosen research topic

Irrespective of the type of questionnaires, the preliminary step of knowledge exploration is always necessary to acquire a thorough understanding of the overall scope of background subject matter. This can often be achieved via literature review of published journal articles or academic books, or through direct observation, or engaging in purposeful communication with experts, prospective respondents or holding a group discussion with key stakeholders and/or subject matter experts. It is a good practice to carefully review the existing study instruments or questionnaires because it can be very helpful for a researcher to first explore how the past researchers had perceived the scope and background content of the subject matter and the approach adopted by past researchers for designing the questionnaire. The overall goal is to better understand the scope of the subject matter so that the questionnaire that is awaiting development will be at least of equal or even higher standard (i.e. add value) than the existing questionnaires.

After undertaking an initial exploratory work, the researcher should be able to answer the following questions:

- Why is development of this new questionnaire necessary?
- What is the overall scope of the questionnaire? The determination of its scope is important because it will ensure that only important and relevant questions will be asked.
- From among all the questions which are considered necessary to be asked, which are regarded as the truly important ones?

- How to phrase the questions during the development of the questionnaire?
- What information will be expected for the researcher to derive from the questionnaire?
- Is the new questionnaire going to be of the same standard or better than existing questionnaires that are measuring the similar subject matter?

After a preliminary exploratory work has been conducted properly, the researcher will ideally be equipped with sufficient knowledge to develop appropriate questions for their questionnaire. The success of questionnaire development is highly dependent on the initial acquisition of thorough understanding of the subject matter. It is not a one-off process but rather, a continuous process where the researcher will have to acquire an in-depth understanding of the background subject matter until the questionnaire has been successfully developed and finalised.

4.2 Step 2: To develop the questionnaire

Item development is the most crucial step in the development of a questionnaire. The researcher will need to have sufficient knowledge on the background subject matter before they can design appropriate questions. It is recommended to review the reporting standards of the current available questionnaires, in particular, those of the benchmark questionnaire(s). If there is no benchmark questionnaire available on a particular subject matter, it will be helpful to review other questionnaires from the same research topic. After having identified the minimum standard expected of the questionnaire that measures a similar subject matter, the researcher will need to add value in terms of the overall content or measurement scale of the questionnaire.

The approach for developing and designing a questionnaire varies according to the type of questionnaire. Hence, the researcher will first need to understand all the different types of questionnaire and each of their utility and versatility for designing a questionnaire with a specific purpose. To the end, this book will be recommending a few important steps for developing a questionnaire and also providing a brief introductory guide for each type of questionnaire, as presented in Table 4.1.

Table 4.1: Recommended steps for questionnaire development

Steps	Details
Determine the overall scope of a questionnaire and its target population "The target population can be any sub-population such as patients (for which the diagnosis of a disease will be established), employees, or healthy population. The scope of the questionnaire shall depend on the	***Questionnaire to collect facts/information*** The scope can be set based on suitability of subjects' characteristics, logistical or other scientific considerations of the underlying subject matter. Say, for example, in a questionnaire that aims to collect data from a patient's medical record, the information needed to fill in this questionnaire may include demographic profiles of subjects, and their medical history,

research question(s) and the study objective(s)."	treatment and other clinical outcomes. ***Questionnaires to measure knowledge*** Knowledge regarding specific diseases can be broad. Researchers need to identify the scope of knowledge to be explored: whether to assess general knowledge on e.g. diabetes mellitus, or a specific area e.g. insulin therapy. Researchers will also need to understand the level of knowledge expected of the target population. Researchers

	may set questions of a higher level of difficulty for health professionals as compared to questions that aim to assess the patients' knowledge levels. Hence, it shall also depend partly on the access to information by the target population. ***Questionnaires to measure latent variables*** A simple definition for latent variable(s) is recommended in order for a common understanding of the concept of 'latent variable' to be standardized. The

	definition of the latent variable within the context of the questionnaire will then reflect the scope of the questions to be asked. Latent variables usually refer to an existing theoretical framework for the measurement of a construct such as those underpinning theories on the measurement of a ‘job satisfaction’ scale. These latent variables can also be based on an entirely new conceptual framework which has been designed by the researchers.

	Most importantly, the design of the theoretical framework should be robust and has been established by a strong basis which is in line with the current literature (See Appendix 1: The example of a conceptual framework for developing the job satisfaction questionnaire). ***Questionnaires to screen or diagnose a patient*** Set the definition of the outcome parameter such as with or without the event (which means that the outcome of interest

	can either be with the disease (poor outcome) or without the disease (good outcome). Then, choose the target population who are to be screened or diagnosed with the specific event of interest. The target population can be a healthy population or among population who are at risk such as patients.
~	
Determine the appropriate number of sections and the required questions or	***Questionnaire to collect facts/information*** The questions are usually straightforward. Asking subjects' "age" may give

variables within each section. "The number of variables in a questionnaire depends on the scope of study and practicality of the time requirements to fill in the questionnaire. The variables should be sufficient to collect all the necessary information but not too many that will deter the subjects from completing the questionnaire and thus leading to the presence of	rise to a different understanding such as whether it is referring to age at notification or age at diagnosis or age at current year. Therefore, questions need to be asked in a specific manner. All the questions should be accompanied by an appropriate definition and its unit of measurement. For example, age of an infant will be in weeks or months as compared to years for adults. Avoid open-ended questions unless they are absolutely necessary.

missing data. The variables should also be clearly defined and unambiguous. It is recommended to start the entire process by determining the list of appropriate sections or domains. Next, it is necessary to list down all the appropriate questions within each section or domain. Finally, it is then necessary to	The definition of each variable will need to be established. The appropriate unit of measurement for a variable has to be defined and its format must be standardized. All questions in the survey questionnaire are usually stated in a simple way such as respondent's date upon diagnosis of a disease, gender and age at diagnosis. Hence proper definitions for the three variables are necessary, such as "the date when the patient had been notified to the

carefully review all the sections and questions again until the draft set of a questionnaire is completed and finalized. This review should be based on an understanding of the subject matter and findings obtained from Step 1 (i.e. To explore knowledge and understand in depth the chosen research topic”	registry based on the format dd/mm/yyyy (eg: 01/05/2015)”, patient’s gender where male is coded with 1 and female is coded with 2” and “Patient’s age at disease onset in years” respectively. The variable definition is important because it serves as a useful reference, especially for the investigators and statisticians. ***Questionnaires to measure knowledge*** Knowledge can stand alone as a single domain or can be divided into a

	few specific sub-domains. For example, knowledge of antibiotics may include other sub-domains such as theory on antibiotics and prescribing practices of antibiotics. In addition, both theories and practices can also be subdivided into a few other minor categories if necessary. Therefore, researchers have to set the scope of the knowledge questionnaire which is well-suited to the study objective and also to the knowledge level

	expected of the target population. Knowledge question is based on facts which one can readily find their answers from reputable sources. Thus, questions that do not have answers which are presented as clear facts should be avoided. For example, risk factors for a clinical outcome which are published in scientific journals can often vary according to their source. Thus, they are not usually regarded as knowledge questions yet.

	The answer for each question is clear and unanimous, and is well-supported by facts, theories or practices. The answer is usually derived from a consensus among various scholars and is usually found in reputable resources such as published text books and standard guidelines. Responses which are expressed in a binary form such as “Yes” or “No” answer and in a multiple-choice answer are usually applied. Feedback which are rated by study respondents that base on a measurement

	scale such as Likert scale is discouraged. (See Appendix 2: Example of presentation to support evidence for knowledge questions). ***Questionnaires to measure latent variables*** The development of the items/questions usually begins with making an initial reference to the existing model of theory or framework. Of note, it must be borne in mind that it can often be a challenging task to develop the questions for a new domain which has

	been introduced by the researchers. This process requires researchers to explore the concept of a particular variable parameter that is based on previous scholarly work in order to understand how the variable parameter is being perceived by subject matter experts from the relevant field. However, it must be emphasized that those domains which are identified by the researchers based on their preliminary exploratory literature

	review may not be the same as the results derived from the subsequent statistical analysis. This is one of the major challenges in developing a questionnaire for measuring latent variables. The ultimate aim is to match the domains identified by the theoretical framework with those identified by key statistical findings. In other words, a strong basis for the construct of the domains will be established, if and only if the domains are

	consistent with the key findings obtained from statistical analysis. The questions within each domain should clearly reflect the overall scope of the intended domains. Therefore, the degree of success achieved by the construct is closely related to the extent for which the researchers are able to include the most ideal domains of a variable parameter in the questionnaire and then to develop relevant items/questions for each included domain.

	At this preliminary stage, the researchers will usually introduce many questions in each domain as it is likely for some of these questions to be omitted based on results of subsequent statistical analysis. In most cases, only responses based on a Likert scale format will be used. ***Questionnaires to screen or diagnose a patient*** The questions on risk factor or predictor of a clinical outcome should be specific to the outcome or disease of interest.

	Only those risk factors with a strong association with the outcome or disease of interest should be included so that each question will hold a sizeable weight for discriminating between disease and non-disease. This process requires the researchers to understand the variable parameter based on previous scholarly works. Extensive research is required to obtain a sound understanding of the risk factors or predictors which are strongly associated with

	a positive event (i.e. outcome of interest such as with the disease). Hence, the researchers are usually subject matter experts in this field. The relationship between the risk factors or predictors and the occurrence of a positive event (i.e. outcome of interest such as with the disease) should be at least moderately strong, and also be backed up by reputable sources. The validity of the questions is highly dependent on how well they are able to discriminate between diseased and non-

	diseased groups (or event group versus non-event group). Responses which are based on a measurement scale such as the Likert scale, multiple-choice answers and binary responses (such as “Yes” or “No”) are generally used.
~	
Define the scoring scheme for overall or/and each domain if necessary “Not for Questionnaire to	Researchers have to define the scoring scheme to determine the overall score and individual scores for each domain level. It is preferred to fix the same weightage for each question, but researchers

collect facts/information”	may give higher weightages for selected questions. The conventional way of calculating the final score is usually by summation which is then converted into a percentage. A scoring manual should be prepared in advance for future reference. Some questionnaires use raw score although the majority converts the raw score into a percentage. This scoring will need to be validated subsequently to ensure that it can accurately

	reflect the magnitude of knowledge level (i.e. measure knowledge level), the magnitude of domains (i.e. measure latent variables) or the ability to discriminate between disease and non-disease groups (i.e. screen for or diagnose a disease in a patient).
~	
Determine the overall layout, structure and format of a Questionnaire	At this stage, researchers will be ready with the questions and each of their respective measurement scales. The questionnaire should be presented in a simple and attractive manner to ensure that respondents are well-motivated to fill

	in the questionnaire completely. The following provides a list of recommendations on how to design the overall structure and format of a questionnaire: i. Provide a clear title of survey at the front page of the questionnaire. ii. Provide a space for researchers to indicate the subject code for each completed questionnaire. iii. Insert page numbers in the whole questionnaire so that the subjects

	can identify any missing page(s) of the questionnaire, if they occur. iv. Briefly introduce background, objective and expectation of the study. v. It is also a good practice to request the subjects to fill in the questionnaire by using a pen instead of a pencil. vi. Arrange the questions into different sections in order to

	organize them under relevant parts of the questionnaire and then provide clear instructions within each section. vii. Use a reasonably large font size. Insert a statement of appreciation to express thanks and gratitude to the respondents at the end of the questionnaire.

4.3 Step 3: To assess both content validity and face validity

Topics on content validity and face validity have already been discussed thoroughly in existing literature including standard textbooks and research journals. This section aims to describe both content and face validity specifically for conducting questionnaire development studies. After the questionnaire has been finalized, its content validity and face validity must be assessed before proceeding to a pilot study. This is considered to be an early screening process to prevent problems from arising during the pilot study.

Content validity refers to the degree to which the study instrument is covering the content which it is supposed to measure. There are basically two important measures that must be addressed, the first is the measurable extent of each item for defining the traits and the second is the defined set of items that represents all the aspects of traits (Yaghmaie, 2003). Thus, content validity is a process in which the panel of subject matter experts (SME) for a particular field are elected to review the validity of each item, question or domain from the questionnaire and then provide their constructive comments on how to improve them further. This process is considered to be one of the most crucial step when the SME or its elected panel will carefully review and evaluate the content: whether the questions carry scientific or theoretical value, are aligned with the scope of study, are reflecting the intended section or domain and also whether the questions are able to answer the research question and to address the study objective(s).

Content validity is commonly assessed by a qualitative approach when all the comments are being addressed and subsequently the researcher will improve the wording of these questions to maintain the validity and quality of the content of these questions or items. In other cases, the process can be completed through a quantitative approach such as by using content validity index (CVI) where the SMEs or its elected panel will rate each item in terms of its content validity and a specific formula will be used to assess each question or item (Rubio et al., 2003).

The conventional way to check for its content validity is by conducting a face-to-face approach whereby the panel of subject matter expert will assess and provide their recommendations for improvement straightaway. The content validity (i.e. the initial development and subsequent amendment of the questions) can be conducted several times until the researcher is satisfied with the wording of the content of these questions.

Another way to assess content validity is shown in the following example. In this example, the researcher aims to develop a job satisfaction questionnaire that consists of a few domains. Say, for example, the domain that requires to be assessed is “leadership”. Here is a description of all the necessary steps the researcher should take in order to assess the content validity of the three questions that are developed to represent the domain for leadership in the questionnaire.

Step 1: To elect at least three expert panels from the relevant field.

Step 2: To prepare the questions and arrange them according to a structured format such as in Figure 4.1.

Step 3: To send an initial draft of the questionnaire to the elected expert panels for a careful perusal and review. They will then critically evaluate each question and provide suggestions for improvement.

Step 4: To make all necessary amendments in the draft questionnaire after receiving feedback from the panels of subject matter experts.

(Step 2 until Step 4 can be repeated a few times if necessary)

Based on a quantitative approach, the responses provided by the expert panels can be assessed as shown in Figure 4.2. The calculation of the score will be shown in Step 4. In this example, the content validity is assessed by a quantitative method based on Content Validity Index (CVI), Item-level Content Validity Index (I-CVI) and Scale-level Content Validity Index, universal agreement calculation method (S-CVI/UA) (Polit & Beck, 2006). In this example, each question is assessed in a four Likert-scale such as “Not relevant = code 1”, “Somewhat relevant = code 2”, “Quite relevant = code 3” and “Highly relevant = code 4”. The experts’ responses which are given as “Quite relevant” and “Highly relevant” will be coded ‘1’ and those given as “Not relevant” and “Somewhat relevant” will be coded ‘0’. Based on the example above, the mean I-CVI

= 0.93, S-CVI/UA = 0.80 and mean expert proportion = 0.93. The value of I-CVI of at least equal or more than 0.90 shall be regarded as the items having sufficient evidence of content validity (Polit & Beck, 2006).

Instruction: Kindly provide your evaluation for each question in terms of the suitability of its content pertaining to the knowledge domain

≠	Domain: Knowledge	**1** **Not Relevant**	**2** **Somewhat relevant**	**3** **Quite relevant**	**4** **Highly relevant**
1					
2					
3					
.					
.					
20					

Instruction: *Kindly provide your suggestions for those questions that will require further revision.*

Domain	**Item**	**Remark**
Knowledge		

Figure 4.1: Example of feedback form to assess the content validity of a questionnaire

Table 1: Ratings on a 20-Item scale by four experts: items that received a rating of 3 or 4 on a 4-point relevance scale

Item	Expert 1	Expert 2	Expert 3	No. in agreement	Item CVI
1	1	1	1	3	1.00
2	1	1	1	3	1.00
3	1	1	1	3	1.00
4	0	1	1	2	0.67
5	1	1	1	3	1.00
6	1	1	1	3	1.00
7	1	1	1	3	1.00
8	1	1	1	3	1.00
9	1	0	1	2	0.67
10	1	1	1	3	1.00
11	1	1	1	3	1.00
12	1	1	1	3	1.00
13	1	1	0	2	0.67
14	1	1	1	3	1.00
15	1	1	1	3	1.00
16	1	1	1	3	1.00
17	1	1	1	3	1.00
18	1	1	1	3	1.00
19	1	1	0	2	0.67
20	1	1	1	3	1.00
				Mean I-CVI	0.93
				S-CVI/UA	0.80
Proportion relevant	0.95	0.95	0.9	Mean expert proportion	0.93

Note:

CVI: *Content validity index*

I-CVI: *item-level content validity index*

S-CVI/UA: *scale-level content validity index, universal agreement calculation method*

Figure 4.2: An example to illustrate the calculation of CVI, I-CVI and S-CVI/UA to assess content validity

Once the results obtained from an assessment of content validity is satisfactory, then the researcher can proceed to the next step, which is an assessment of its face validity. Face validity refers to the suitability of the content of an item(s) or a test which is specifically designed for an intended purpose as perceived by test takers, users or respondents (Secolsky, 1987). When a study instrument has high face validity, this will help to increase the level of cooperation and motivation among the examinees or respondents and then attract them to fill in the instrument (Nevo, 1985). This process will therefore aim to ensure the relevance of a test when it is being administered to the test participants (Holden & Ronald, 2010; Gravetter, et al., 2012).

Face validity can be explored via an interview or a focus group discussion involving a qualitative approach, during a two-way communication between researchers and a few (preferably five to ten) respondents. The emphasis is on whether the respondents' level of understanding of the questions in terms of their use of language and terminology can be regarded as satisfactory or not. The respondents will then provide honest feedback to these items or questions. This process can also include statistical analysis (by using a quantitative approach) e.g. based on feedback which is measured on a Likert-scale to lend support to the evidence of face validity (Bujang et al., 2016). Such feedback will then be evaluated by researchers who shall decide if the questionnaire requires further amendment.

Since the findings obtained from an assessment of its face validity are usually derived from the respondents, therefore feedback elicited from them will not be necessarily valid. This is because not all respondents have sufficient level of knowledge or prior experience with the meaning of the items or questions in the questionnaire. In addition, they might not understand the objective of the survey, or the overall purpose of its structure and format. Nevertheless, the researcher will still need to carefully examine the respondents' responses, comments, feedback, and/or justifications for such responses. Such comments should still be accepted if they are valid. Based on the assessment of its face validity, the researcher should be able to determine whether or not the item or question will need to be rephrased, amended, expanded or omitted.

Unlike other types of questionnaires, a knowledge-based questionnaire especially for examination purposes may not necessarily require face validity testing because the researcher (who is also the subject matter expert) will already know the probable level of knowledge among the study subjects. Hence, the researcher will develop the questions which are tailored to an estimated level of knowledge of the target population. Nevertheless, feedback obtained from the study subjects is still valuable for estimating the difficulty level of these questions in order to decide what further amendments are necessary.

In summary, both content validity and face validity are important measures to ensure that the questionnaire is well-developed and all questions have been appropriately set. These processes can be repeated several times until the researchers are satisfied with both the content and face validity measures. Such processes should be completed ethically and in a trustworthy and transparent manner. Ignoring both these processes will eventually create many problems such as the elicitation of invalid responses from the respondents and thus introducing bias into interpretation of data and conclusion-generation.

This explains why the researcher must always bear in mind that it is always necessary to ensure that both content and face validity are properly established, which makes it highly probable for the study instrument to achieve two major successes in attaining high levels of both reliability and validity. The first success in achieving high level of reliability can usually be seen from a pilot study.

4.4 Step 4: Pilot study

A pilot study is a preliminary investigation that is conducted before proceeding to the actual survey or experiment which aims to address the objectives of the study. Conducting a large, full-scale study is usually highly cost-consuming and resource-consuming. Therefore, any major flaws in the actual studies will very likely to result in a wastage of resources such as time, manpower and money. Thus, a pilot study shall serve as an important prerequisite for the main study (with a larger-scale) by ensuring that it is feasible to run the main study with a realistic ability to address its research objectives.

For a questionnaire development study, the pilot study is considered to be a preliminary evaluation of the questionnaire before the researchers are finalizing the questionnaire and confirming its structure and layout. In a pilot study, a small number of respondents will be asked to provide their responses to the newly-designed questionnaire based on a self-administered survey. Their responses will then be analysed and assessed based on a quantitative approach. Usually, pilot study requires a small sample size of less than 50 subjects. At this stage, emphasis is only on reliability assessment since evidence for validity basis usually requires larger sample size which can only be achieved in fieldwork phase.

The aim of a pilot study is to design a questionnaire in such a way that minimises potential errors that can occur in the questionnaire, especially in terms of its content. Common statistical techniques for use in a pilot study are test-retest reliability and internal consistency whereby a sample size of 30 to 50 subjects is usually regarded as adequate (Bujang et al., 2018; Bujang and Baharum, 2017a; Bujang and Baharum, 2017b). Reliability and validity testing are discussed in subsequent sections. Problematic item(s) can be detected through statistical analysis so that necessary action can be taken whether to amend, rephrase or omit the item(s).

4.5 Step 5: Fieldwork study

The final step is to conduct a fieldwork study. Fieldwork study involves conducting a survey with a sufficiently large sample size to assess the reliability and validity of a newly developed questionnaire. In this stage, researchers will need to design a study in such a way that the newly-developed questionnaire will have sufficient evidence of both reliability and validity. To conduct a fieldwork study for assessing a questionnaire’s reliability and validity, there are a list of important considerations that should be taken into account, including the types of study design, subjects’ eligibility criteria, sampling technique, sample size planning and statistical analyses.

It is commonly understood that any mistakes in statistical analysis should be rectified since it is always possible for such analysis to be redone. In contrast, any errors in the selection of a suitable research design can lead to disastrous results because it will be too laborious and time-consuming to repeat a survey especially one with a large sample size. Therefore, careful planning is necessary for establishing both the research design and data analysis, which should be well thought-out and properly accounted for. One of the important elements for a questionnaire development study is to decide on the appropriate measures for performing reliability and validity testing for the new questionnaire. In subsequent section (Chapter 5 and Chapter 6), this book discusses regarding the common approaches and statistical techniques which are used to test reliability and validity of a questionnaire.

In summary, an overall process for questionnaire development is presented schematically in Figure 4.3. This process involves five important and related phases, which is illustrated in a step-by-step fashion in Figure 4.3. As each of these steps has the same weightage in terms of its significance, right from the very beginning until the end; therefore, the researcher should not ignore any of the steps included in this process.

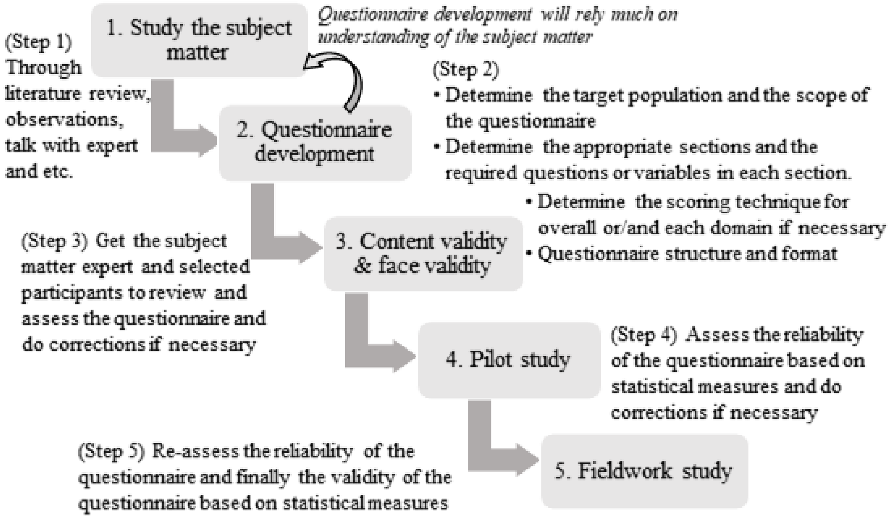

1. Study the subject matter
(Step 1) Through literature review, observations, talk with expert and etc.
Questionnaire development will rely much on understanding of the subject matter
2. Questionnaire development
(Step 2)
• Determine the target population and the scope of the questionnaire
• Determine the appropriate sections and the required questions or variables in each section.
• Determine the scoring technique for overall or/and each domain if necessary
• Questionnaire structure and format
3. Content validity & face validity
(Step 3) Get the subject matter expert and selected participants to review and assess the questionnaire and do corrections if necessary
4. Pilot study
(Step 4) Assess the reliability of the questionnaire based on statistical measures and do corrections if necessary
5. Fieldwork study
(Step 5) Re-assess the reliability of the questionnaire and finally the validity of the questionnaire based on statistical measures

Figure 4.3: Overall process of questionnaire development for quantitative medical research

Chapter 5

Testing the Reliability of a Questionnaire

Previous chapters have discussed regarding the process of questionnaire development from the beginning up until conducting a fieldwork study. After the recruitment of respondents has been completed, then the next step is to use the data to evaluate the newly developed questionnaire in terms of its reliability and validity. The determination of the appropriate statistical analysis to assess a questionnaire’s reliability and validity should be done upfront during the initial stage of research design. By knowing what types of questionnaire that should be developed and what types of evidence to be gathered in order to address the study objectives, the researcher will be able to determine what types of reliability and validity testing are required and subsequently they will decide what types of statistical analysis should be performed.

Earlier on, the four types of questionnaire are introduced in Chapter 2. The reason for introducing the four different types of questionnaires right at the beginning is that different types of questionnaire may necessitate different types of statistical analysis for the purpose of demonstrating evidence of both its reliability and validity. This has been proven time and again in the existing literatures. For example, a newly-developed questionnaire which aims to measure a latent variable such as job satisfaction might require Cronbach's alpha test to measure the reliability of the domain. On the other hand, a newly-developed questionnaire that measures knowledge may not require the same test to assess the reliability of the questionnaire.

Basically, all types of study instrument that are used for research must be validated. For example, it is compulsory to validate a weighing machine that is designed for use in research to ensure that all the observations made for these measurements of weight are precise enough. This applies similarly to questionnaires. However, the basis for determination of reliability and validity of a questionnaire can be either similar or different depending on the specific type of questionnaire. In this chapter, the book will discuss various statistical techniques for assessing the reliability of a questionnaire in general. Subsequently in Chapter 6, the book will summarize the appropriate statistical technique or statistical test that can be performed to test the reliability and validity for different types of questionnaires.

A valid questionnaire should ideally fulfil the specific requirements for both reliability and validity testing. Hence, researchers are responsible for providing as much evidence as possible on the reliability and validity of a questionnaire to convince the audience that the newly-developed questionnaire is valid. Reliability is a measure of consistency, stability and reproducibility (AERA, APA, NCME, 2014). For example, a weighing machine is considered reliable when repeated measurements of a particular weight shall produce almost the same value each time. In questionnaire development studies, reliability can be measured in two ways based on: (i) test-retest reliability or an agreement test, and (ii) perform the inconsistency test by calculation of the measure for an inconsistency.

5.1 Test-retest reliability

In test-retest reliability, when the subjects' responses for a questionnaire are found to be consistent after being self-administered for several times (usually at least twice), this indicates evidence to show that the questionnaire has some level of reliability. Any disagreement could be due to the varying contexts in which the questions are set which may possibly confuse the subjects. Another reason is that any recent changes in the subject's medical condition or environment that may possibly affect the subject's responses. For example, a subject may have symptoms of depression during the first survey but his/her symptoms may not be the same after 1-2 weeks' time. However, in test-retest reliability, the researchers shall assume that the subjects' medical or external condition remains constant between the two time points.

Considering that subjects may be influenced by their medical condition or their circumstances, researchers will usually provide a margin of allowance for the values of agreement. The usual procedures for performing the statistical analysis to test for the level of agreement is either via the intra-class correlation test for numerical values or kappa agreement test for categorical values. The acceptable range of coefficient values for demonstrating sufficient level of agreement in both statistical tests is more than 0.4 (Cicchetti, 1994; Bujang & Baharum, 2017a). For questionnaire development, the authors recommend researchers to aim for higher range of coefficient values, i.e. more than 0.5.

Table 5.1 shows a set of sample data which is collected by a questionnaire development study where a researcher aims to test the test-retest reliability of two questions (question 1=Q1 and question 2=Q2) and total score (TOT) at week 1 and week 2. The result is presented in Table 5.2. Result shows that question 1 is demonstrating an acceptable and moderate agreement with kappa coefficient= 0.565 (p=0.010) when compared with question 2 with a value of kappa coefficient= -0.316 (p=0.143). Looking at the red box, Q2 has more instances of disagreement when compared with Q1. Assuming all the other factors remain constant, it is likely that respondents might erroneously conclude that Q2 has resulted in higher number of inconsistencies in their responses. Thus, researcher may need to study the real cause of disagreement in the responses to Q2 and then to make necessary amendment and/or substitution for Q2.

Table 5.1: Sample data of a new Questionnaire A

ID	Q1W 1	Q1W 2	Q2W 1	Q2W 2	…	TOTW 1	TOTW 2
1	1	1	1	0		7.7	8.9
2	0	0	1	1		7.6	7.6
3	1	1	0	1		7.8	7.8
4	1	1	1	1		8.0	8.0
5	0	1	1	1		7.6	7.6
6	1	1	1	0		7.8	7.8
7	0	1	0	1		8.0	8.0
8	1	1	1	1		8.9	8.9
9	0	0	1	1		7.5	7.5
10	1	1	0	1		7.6	7.6
11	1	1	1	0		7.7	8.9
12	0	0	1	1		7.6	7.6
13	1	1	0	1		7.8	7.8
14	1	0	1	1		8.0	8.0
15	0	1	1	1		8.3	8.3
16	1	1	1	0		8.2	8.2
17	0	0	0	1		8.2	8.2
18	1	1	1	1		8.9	8.9
19	0	0	1	1		7.5	7.6
20	1	1	0	1	…	7.6	7.8

Note:
Q1W1: Question 1 observed at week 1
Q1W2: Question 1 observed at week 2
Q2W1: Question 2 observed at week 1
Q1W2: Question 2 observed at week 1
TOTW1: total score observed at week 1
TOTW2: total score observed at week 2

Table 5.2: Results of a test-retest reliability for Questionnaire A

Question	Kappa Coefficient	ITC	p-value
Q1	0.565		0.001
Q2	-0.316		0.143
Total score		0.649	0.001

Note:
ITC: Intra-class correlation coefficient

The test-retest reliability for the total score (TOT) will be evaluated by using the intra-class correlation test. Based on the result, an assessment of the level of agreement via the TOT has reported a value of intra-class correlation coefficient to be

0.649 (p=0.001). The agreement is considered acceptable since the coefficient is calculated to be more than 0.40. By amending, rephrasing or substituting question Q2 (to another better and more valid question), then both the level of agreement of Q2 and that of TOT will improve.

5.2 Internal consistency

Another reliability measure is internal consistency. The internal consistency can be measured for questionnaires that use rating scales like the Likert scale. A common type of Likert scale is a 5-point Likert scale (1 = strongly disagree, 2 = disagree, 3 = not sure, 4 = agree and 5 = strongly agree). A higher level of internal consistency can be achieved if the majority of the subjects gives the same rating as their responses for the same question. The usual statistical test which is used to measure internal consistency is Cronbach's alpha test with an acceptable coefficient value of more than 0.5 (Nunally, 1967; Bujang et al., 2018). For questionnaire development, this book is advocating a recommendation to aim for higher values of the coefficient for Cronbach's alpha such as more than 0.70.

Another statistical test that can be performed to assess reliability is RASCH analysis. A RASCH analysis is a technique for measuring reliability for both the person's and the item's reliability. A good item will yield a sufficiently high level of person's and item's reliability, with coefficient values of more than 0.70 (Tennant et al., 2004). The person's reliability estimates how well we can discriminate people based on their estimated visual ability (Bond & Fox, 2001). On the other hand, the item's reliability indicates how well the items can be distinguished from each other on the basis of their difficulty level (Bond & Fox, 2001). Of note, RASCH analysis also offers many other statistical measurements beyond reliability testing, including validity testing.

Reliability is usually a preliminary assessment of questionnaire development that can be performed during pilot study. One of the key advantages of using statistical techniques such as intra-class correlation, kappa agreement test and Cronbach’s alpha for assessing reliability is that the sample size requirement to determine high effect sizes is usually small (Bujang and Baharum, 2017a; Bujang and Baharum, 2017b). Therefore, any coefficient values derived from the analysis that are found to be below the acceptable values shall indicate that the questions must be rephrased, amended or omitted before the questionnaire can undergo a validation process during the subsequent fieldwork phase.

Table 5.3: Sample data of a new Questionnaire B

ID	Q1	Q2	Q3	Q4	Q5	Q6
1	4	4	4	4	1	1
2	3	3	3	3	2	4
3	4	4	4	4	2	3
4	4	4	3	4	3	4
5	4	4	4	3	1	1
6	4	4	4	4	3	4
7	3	3	3	3	3	3
8	4	4	4	4	4	4
9	4	4	3	4	3	1
10	4	4	4	4	2	3
11	4	4	4	4	1	1
12	4	4	4	3	2	4
13	4	4	3	4	2	3
14	4	4	4	4	3	4
15	4	4	4	3	1	1
16	4	4	4	4	3	4
17	4	4	4	3	3	3
18	4	4	3	4	4	4
19	4	4	4	4	3	1
20	4	4	4	4	2	3

Domain 1

Domain 2

Table 5.3 shows a sample set of data for a questionnaire development study for Questionnaire B. In this example, Q1, Q2 and Q3 are hypothesized to be grouped together in one unique domain (Domain 1) and Q4, Q5 and Q6 are hypothesized to be grouped in another unique domain (Domain 2). The researcher aims to measure the internal consistency of Domain 1 and Domain 2. The result is presented in Table 5.4.

Table 5.4: Results of internal consistency for Questionnaire B

Domain/item	Cronbach's alpha	Corrected item-Total correlation	Cronbach's alpha if item deleted
Domain 1	0.812		
Q1		0.804	0.636
Q2		0.804	0.636
Q3		0.509	1.000
Domain 2	0.587		
Q4		0.180	0.723
Q5		0.642	0.086
Q6		0.508	0.372

Domain 1 reports a value of Cronbach's alpha of 0.812 and a value of minimum Corrected item-Total correlation (CITC) of 0.509. The value of Cronbach's alpha is reasonably high and this refers to the strength of consistency of the items for a particular domain. The values of CITC are also calculated to be relatively high and well above the minimum cut-off of 0.30. It is recommended to delete any items with a corrected item-total correlation of <0.30 (Kline, 1993). On another note, Q1 and Q2 is highly congruent with each other, which results in the same values obtained for both Corrected Item-Total Correlation (CITC) and Cronbach’s Alpha if Item Deleted (CAID). Besides that, the result shows that the coefficients of Cronbach’s alpha for Q1 and Q2 are the same. Although it seems that Q1 and Q2 are perhaps indicating for redundant questions, however it is not necessary for deleting any of the items. As suggested in the existing literature, questions of a particular level of importance (i.e. with an acceptable level of content validity) should ideally be retained regardless of their results via assessment by the correlation analysis (Rattray and Jones, 2007).

Domain 2 reports a value of Cronbach's alpha of 0.587 and a value of minimum Corrected item-Total correlation (CITC) of 0.180. The result shows that Domain 2 is weak in terms of its consistency. For questionnaire development studies, all the inconsistent items found in this domain should be rephrased or omitted. If no action is taken, then the problematic item or domain will jeopardize the validity of the whole questionnaire. This is the reason why it is always strongly emphasized that the reliability of a new questionnaire must be carefully evaluated during pilot study, so that any amendments necessary to be made in the questionnaire have already been made before embarking in the fieldwork phase of validity testing. For example, Q4 might be a problematic item and should be deleted from the whole questionnaire. Hence, the deletion of Q4 will cause the value of Cronbach's alpha to increase to a high value of 0.723.

In an effort to improve the consistency and reliability of items and domains, the process of validation such as assessing the content validity, or face validity and conducting a pilot study can often be repeated several times. Although such efforts are time-consuming, this remains as a very important step for the researcher to take for strengthening the reliability and also the validity of the new questionnaire. Taking into consideration that it is usually rather costly to conduct the fieldwork phase of validity testing of a questionnaire, it is always worthwhile to ensure that the questionnaire has an acceptable level of reliability before subjecting it to validation process in the fieldwork phase. This is because during the fieldwork phase, the main emphasis is to test the validity of the new questionnaire since it usually requires a much larger sample size for testing its validity.

Chapter 6

Testing the Validity of a Questionnaire

Validity of a questionnaire shall be based on statistical measures which are often considered as the strongest level of evidence to confirm the validity of a particular questionnaire. In the previous chapter, this book describes about the importance of content validity and face validity. Both these processes happen before the fieldwork testing of a questionnaire and most of the times they will not be recouping any statistical evidence.

The type of validity discussed in this chapter is referring to that measure of validity which should be assessed by statistical methods. It is again emphasized that different types of questionnaire will require different types of statistical analysis to garner evidence for supporting its validity. In fact, a newly-developed questionnaire which aims to collect facts and information may not require any validity testing, apart from a basic assessment of its content validity and face validity.

For illustration purposes, let's say a weighing machine is said to be reliable if it always produces reproducible or consistent measures.

However, the weight of the study subject is not necessarily truly accurate although the same results will be obtained after several rounds of measurements. The weighing machine is said to be valid if it is able to produce an accurate measure of what it is intended to measure. Thus, validity also refers to an assessment of the accuracy of the measurement. This explains why the validity of a questionnaire is referring to “the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of tests” (AERA, APA, NCME, 2014).

In questionnaire development studies, the questionnaire is said to be ‘valid’ when it is able to accurately measure what it is supposed to measure. For example, if a questionnaire is designed to measure the magnitude of depression of a patient, a higher score will reflect that the patient has more severe symptoms of depression compared to those who have lower scores. We have discussed two types of validity (i.e. content validity and face validity) to be conducted before or during the pilot study. Subsequently, this section will discuss the need for validity testing to be conducted during fieldwork phase. The validity of the questionnaire will be assessed by appropriate statistical methods which shall base their analyses on an adequate sample size. Figure 6.1 shows different types of validity testing. By utilising a quantitative approach, there are two differing measures for validity tests such as criterion validity and construct validity.

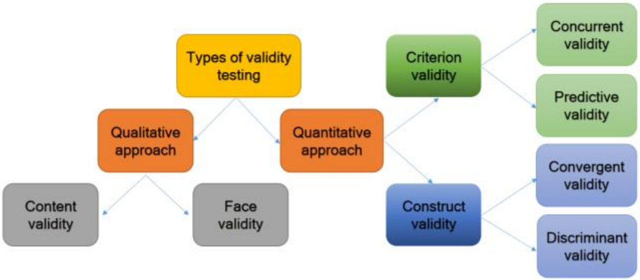


Figure 6.1: Different types of validity testing

6.1 Criterion validity

Criterion validity is the extent to which an operationalization of a construct, such as a test, which can be related to, or which can predict, a theoretical representation of the construct (Cronbach et. al., 1955). For example, a manager gives an interview score (test score) to a candidate and he/she wants to assess whether or not there is a positive (linear) correlation between a candidate's test score and his/her job performance (sales figure). Meaning to say, the manager is predicting that those employees who score highly on their pre-employment test shall also produce high sales figures? If there is a positive correlation between the test score and job performance, then the interview or test score can be considered as having a good criterion validity.

In questionnaire development study, a particular item or test score is expected to have a degree of association or correlation with another variable which can often be an outcome or a test score from another validated questionnaire. Two fundamental measures for the determination of criterion validity are concurrent validity and predictive validity. Criterion validity can sometimes be regarded as a measure of effectiveness.

6.1.1 Concurrent validity

For questionnaire development studies, concurrent validity refers to a measure of the degree of correlation between the test scores derived from a newly validated questionnaire and those from another set of criterion scores. One common practice is for the researcher to use another previously validated questionnaire to support the claim for evidence of validity of the newly developed questionnaire. A validated questionnaire simply means that the work involved in the development of the questionnaire has already been published in peer-reviewed journals or sometimes in books or reports. To check for the concurrent validity of a questionnaire, the two different questionnaires must have some degree of similarity whereby the overall score or domain scores are found to be clearly associated.

This can often be tested by using correlation test such as Pearson’s chi-square test, Pearson’s test or Spearman’s rank test whereby the researcher can associate or correlate the scores between various categories or domains from two different questionnaires. For example, SF-36 questionnaire had been used to validate Diabetes Quality of Life (DQoL) questionnaire (See Figure

6.2). The results showed that ‘satisfaction’ and ‘impact’, which were two of the domains from DQoL, had been correlated with the domains of ‘physical functioning’, ‘social functioning’, ‘role physical functioning’, ‘pain score’ and ‘general health’ domains from the SF36 (Jacobson et al., 1994).

Concurrent validity can often be regarded as a fairly weak type of a measure of validity. The underlying problem with concurrent validity is that the benchmark test (i.e. previous validated questionnaire) may intrinsically have some inaccuracies and, if the new test (i.e. newly-developed questionnaire) does show some degree of correlation, it merely shows that the new test also has the same problem. Therefore, the researcher will need to make sure that the benchmark test has been accorded an appropriate measure of validity, which is deemed satisfactory for the purpose of the research.

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 6.2: Example of testing for validity based on an assessment of concurrent validity (Results obtained from Jacobson et. al., (1994))

Another assessment of concurrent validity can also occur in a following scenario. For example, let’s say a group of medical students take two final exams to assess their knowledge. One exam is a practical test and the second exam is a paper test. If the medical students who score well on the paper test also score well on the practical test, then concurrent validity has occurred. If, on the other hand, students who score well on the paper test score poorly on the practical test (and vice versa), then there is a problem with concurrent validity. In this particular example, the ability of either test to assess knowledge can be questioned.

6.1.2 Predictive validity

Predictive validity refers to the extent to which the questionnaire is sufficiently valid for using the score on a scale or a test to predict the value of some other variable in the future (See Figure 6.3). In medical research, a newly-developed questionnaire will be validated based on its ability to be associated with some other real-life outcomes. Therefore, predictive validity of a questionnaire usually requires a longer time to evaluate as it is necessary to wait for the outcome to occur. Predictive validity can be most suitably applied in circumstances when the measure of the questionnaire is likely to be directly attributed to either a good or a bad outcome in the future.

For example, in a study by Levi and colleagues (1999), the domains were validated only after two years to see whether the negative indicators such as psychiatric hospitalization and drug use could potentially be associated with personality disorders (Levy et al, 1999) (See Figure 6.4). Since predictive validity requires more time to assess the changes in the endpoint or outcome, it is often less popular among researchers. However, predictive validity can still be useful to indicate the extent to which a questionnaire is able to predict future outcomes, which can often be highly relevant in various clinical applications.

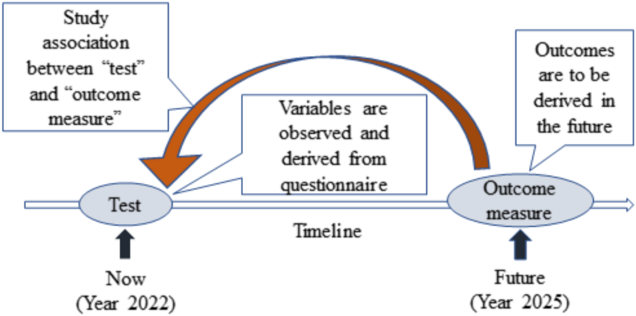


Figure 6.3: Schematic diagram that provides a visualization of the concept of predictive validity

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 6.4: Example of testing for validity based on an assessment of predictive validity (Results obtained from Levy et. al., (1999))

6.2 Construct validity

The conventional concept of validity was traditionally defined as "the degree to which a test measures what it claims, or purports, to be measuring" (Brown, 1996). A construct, or psychological construct as it is also called, is an attribute, proficiency, ability, or skill that is often found in the human brain which is defined by established theories (Brown, 2000). For example, ‘job satisfaction’, ‘diabetes self-care activities’, ‘quality of life’, ‘depression’ are all regarded as constructs. A construct often originates from theory which has also been observed to exist in practice. For example, Toobert and colleague developed a construct for diabetes self-care activities and the four main domains for this construct consisted of diet, exercise, blood glucose monitoring and foot care; all of which were empirically developed from the construct (Toobert & Glasgow, 1994). The two fundamental measures for determination of the construct validity are convergent validity and discriminant validity.

Construct validity can often be analysed statistically via exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). As shown in Figure 6.5, items are hypothesized to represent three domains (as illustrated by three different colours: green, purple and red). It is expected that three domains will be constructed through EFA. A good construct is derived when the construct based on EFA is similar to the construct which was based on the originally hypothesized framework. However, a problematic construct can be obtained if it has at least one item which falls under several different hypothesized domains.

In questionnaire development studies, researchers will have hypothesized domains with several items in each hypothesized domain. The success in the construct is highly dependent on the content of the items which are being introduced in the questionnaire. Ideally, each of the items should represent its own domain only, and therefore should not overlap with the other domains. When a problematic construct is found, the researcher can either substitute or delete those items which overlap with the other domains until an ideal construct has been identified.

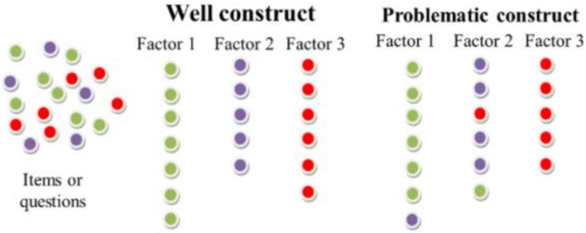


Figure 6.5: A visualization of how an exploratory factor analysis works

As shown above, a good construct consists of items which represent their own domain only, which will enable it to identify common factors that can explain the order or structure among measured variables. On the other hand, a problematic construct consists of items which overlap with the other domains, which will preclude it from identifying those factors that can explain the order or structure among measured variables. A real-life example is shown in Figure 6.6. It was a study by Ahmad et. al., (2020) where the researchers had successfully constructed 8 domains of a job satisfaction model which consisted of *Teamwork (TW), Leadership (LD), Rewards and recognitions (RR), Empowerment (EP), Training & development (TD), Flexibility of working hours (WH), Communication (C), and Working condition (WC)*. This is an illustration based on findings obtained from exploratory factor analysis. To summarize the result, all items were successfully grouped together within their own respective domains with reasonably high factor loadings (>0.500). For example, the five items TW1, TW2, TW3, TW4 and TW5 are measuring TW and none of TW's items are measuring the other domains. The same applies to other domains.

Result of EFA and internal consistency for JS-Q which consists of 34 items and 8 domains

Items	Domains in JS-Q								Cronbach's alpha
	TW	LD	RR	EP	TD	WH	C	WC	
TW1	0.844								0.924 (0.781)
TW2	0.797								
TW3	0.787								
TW4	0.785								
TW5	0.699								
LD1		0.799							0.930 (0.755)
LD2		0.778							
LD3		0.758							
LD4		0.741							
LD5		0.628							
RR1			0.87						0.879 (0.720)
RR2			0.789						
RR3			0.771						
RR4			0.766						
RR5			0.607						
EP1				0.827					0.876 (0.608)
EP2				0.821					
EP3				0.808					
EP4				0.673					
EP5				0.529					
TD1					0.783				0.878 (0.716)
TD2					0.747				
TD3					0.693				
TD4					0.679				
TD5					0.600				
WH1						0.840			0.826 (0.590)
WH2						0.789			
WH3						0.592			
C1							0.735		0.871 (0.725)
C2							0.719		
C3							0.656		
WC1								0.795	0.751 (0.548)
WC2								0.740	
WC3								0.525	

Notes:

Cronbach's alpha (min CITC)

TW = Teamwork
LD = Leadership
RR = Rewards and recognitions
EP = Empowerment
TD = Training and development
WH = Flexibility of working hours
C = Communication
WC = Working condition

Figure 6.6: Example of testing for validity based on an assessment of construct validity (Results obtained from Ahmad et. al., (2020))

6.2.1 Convergent validity

Convergent validity refers to the degree for 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. Based on Figure 6.6, domains TW, LD, RR, EP, TD, WH, C, and WC are measuring the same construct or a common latent variable: namely, job satisfaction. Based on this design, the higher the score in each domain, the more likely for it to be contributing to the higher score on the 'job satisfaction' scale. Thus, it is expected that the high correlation between the scores on different measures that assess the same domains will yield a moderate to strong level of positive correlation between the different study instruments, thereby demonstrating convergent validity.

To assess construct validity of a newly developed questionnaire, the researcher should first establish its convergent validity, before testing for its discriminant validity. The basic difference between convergent and discriminant validity is that convergent validity aims to test whether those constructs which should be related are actually related, while discriminant validity aims to test whether those constructs which should not be related are, in fact, unrelated. Convergent validity can be evaluated using correlation coefficients either using Pearson's correlation test or Spearman's correlation test. However, since exploratory factor analysis can also verify any evidence of both convergent validity and discriminant validity at one go, so it is a well-known fact that reporting the results based on exploratory factor analysis alone is usually sufficient and preferable. However, some scholars prefer to report those results which are based on confirmatory factor analysis and/or RASCH analysis (Adnan et. al., 2018; Bujang et, al., 2018).

For a questionnaire development study, an assessment of convergent validity and concurrent validity seems to be similar in terms of their definitions and proposed statistical analyses. To distinguish between convergent validity and concurrent validity, it must be noted that concurrent validity usually applies to two different sets of questionnaires where one of them is a newly-developed questionnaire and another is derived from a benchmark test (and it is also a validated questionnaire which was published earlier). On the other hand, convergent validity involves those domains from only one construct and also a latent variable which is from the newly-developed questionnaire.

6.2.2 Discriminant validity

Discriminant validity tests whether concepts or measurements that are not supposed to be related are actually found to be totally unrelated. In other words, discriminant validity provides evidence that measures of constructs that theoretically should not be highly related to each other are, in fact, not found to be highly correlated to each other.

Correlation tests (i.e. Pearson's correlation or Spearman's rank correlation) can be applied to demonstrate that the correlation between domains are low which will be indicating that the two domains have an ability to discriminate between each other (See Figure 6.7). Based on this Figure 6.7, there are two latent variables X and Y. Domain A and domain B are the constructs for latent variable X and domain C and domain D are the constructs for latent variable Y. Results show that since domain A and domain B are associated with latent variable X, but not Y. Therefore, domain A and domain B should ideally not be correlated with domain C and domain D. The above observation had already been proven in previous research studies that are based on low correlation coefficients. Practically speaking, discriminant validity coefficients should usually be noticeably smaller in magnitude than convergent validity coefficients, in order for it to hold true.

Previously, it was explained that exploratory factor analysis, confirmatory factor analysis and RASCH analysis are also providing evidence for demonstrating both convergent and discriminant validity. However, it is always recommended to first assess both the convergent and discriminant validity by using a simple test such as correlation test, since it is very useful to avoid any errors or problems which may arise in the subsequent stage of analysis.

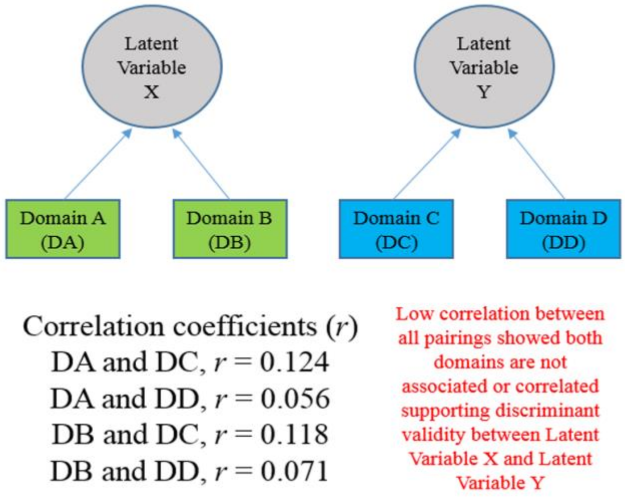


Figure 6.7: Schematic diagram that provides a visualization of the concept of discriminant validity

Low correlation between all the pairings showed that both domains are not associated with each other, which established a high level of discriminant validity between latent variable X and Y.

6.3 Diagnostic assessment

It can be debatable whether or not diagnostic assessment falls under any type of validity testing (concurrent, predictive, convergent and discriminant) as mentioned previously. Probably such a definition of diagnostic assessment bears a close resemblance to criterion validity because researchers are comparing the scale (which is regarded as an alternative) to the gold standard (which is regarded as the truth). In medical research, technique of diagnostic assessment has been widely used to develop a questionnaire, particularly for one that is used to screen or diagnose a patient. In this book, this type of questionnaire shall fall under the fourth category (i.e. questionnaire to screen for or diagnose a medical condition in a study subject).

For example, a newly-developed questionnaire is said to be valid if it is able to discriminate between the ‘diseased’ and ‘non-diseased’ groups. Discriminant validity can be analysed by using sensitivity and specificity test. 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). From the analysis, the researchers had identified a newly-developed questionnaire which is expected to produce moderate to high values for both sensitivity and specificity tests.

However, it is quite rare for a questionnaire or scale to be so highly sensitive in establishing or ruling out the diagnosis of OSA in a patient. In clinical practice, the gold standard for diagnosing OSA is the attended overnight level I polysomnogram (PSG). As mentioned earlier in

Chapter 2, the severity of OSA is measured with the AHI, which is the number of apnea and hypopnea events per hour of sleep. The AHI reading which exceeds 5 shall indicate that a patient has OSA (Young et al., 2002). The gold standard for establishing OSA diagnosis in a patient is too time-consuming and expensive. Therefore, both the Berlin OSA and STOP-Bang questionnaire are now developed to provisionally screen for any patients with a probable diagnosis of OSA, before confirming their diagnosis at a later stage.

6.4 Known group 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 group comparison can be applied 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. For instance, let's say a particular set of questions under one domain is representing the magnitude of distress whereby a higher score indicates a more severely distressed state of the subject. This domain shall be considered valid if there is a statistically-significant difference in the mean score or median score between the distressed group and non-distressed group (Tiong et al., 2018). Similarly, it is also a regarded as a suitable technique that can be applied for all types of questionnaires except a questionnaire which is designed to collect information from the study subjects. Statistical techniques such as independent sample-test or Mann Whitney-U test can be applied since the outcome is expressed in a numerical form (score) and the purpose for statistical analysis is to compare the outcomes between the two distinct groups (See Figure 6.8). On the other hand, Pearson’s Chi-square or Fisher’s exact tests can be applied for the outcomes which are expressed in a categorical form.

Table 4. The comparison of scores of K10 between healthy individuals and individuals who are receiving treatment in psychiatric clinic

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

Figure 6.8: Example of testing for validity based on known-groups comparison (Part of the results obtained from Tiong et. al., (2018))

6.5 Summary of reliability and validity testing

This book has discussed two types of reliability testing (test-retest reliability and internal consistency) and six types of validity testing besides the usual content and face validity (i.e. concurrent validity, predictive validity, convergent validity, discriminant validity, diagnostic assessment and known groups comparison). Content and face validity can be assessed for all types of questionnaires and both these measures of validity should be established before conducting a pilot study. However, it must be borne in mind that not all techniques discussed in this book should be applied for a newly-developed questionnaire. The researcher should wisely select the appropriate statistical techniques for reliability and validity testing of a newly-developed questionnaire.

The overarching idea is to provide as many sources of evidence for both reliability and validity testing as possible. For a minimum requirement, this book recommends that the researcher can choose to apply at least one type of reliability testing and then followed by at least one type of validity testing, besides the usual content validity and face validity. This is because it is already known that the more statistical tests on reliability and validity testing are being applied to a questionnaire, the more robust the cumulative evidence for its reliability and validity will be. Table 6.1 summarises the common statistical techniques that are used to test for reliability and validity in many different types of questionnaire.

Using the following example as an illustration. Let’s say, a newly-developed questionnaire that is designed to collect facts or information is currently awaiting validation. For this type of questionnaire, it is often considered optional to perform test-retest reliability or to test for the level of agreement. In most cases, it may not be necessary to assess the test-retest reliability for this type of questionnaire. However, it can still be done if the researchers choose to do so. Although it is not necessary to garner sufficient statistical evidence for the measurement of reliability and validity in this type of questionnaire; however, it should nevertheless still have both its content validity and face validity be subjected for assessment at a bare minimum.

Table 6.1: Recommended techniques used to test for reliability and validity for different types of questionnaires

Types of reliability and validity testing	Questionnaire designed to			
	collect facts or informati on	measure knowled ge 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
Convergent validity			X	
Discriminant validity			X	
Diagnostic assessment				X

Note: “X” indicates possible technique(s) for validity and reliability testing that can be applied

For a questionnaire that aims to measure latent variables, almost all the statistical techniques for reliability and validity testing can be applied (except for diagnostic assessment). However, the researcher will need to decide whether he/she chooses to apply all the techniques or to select only a few of these techniques. At a bare minimum, both content validity and face validity are required during or before conducting a pilot study. Measuring internal consistency can also be done during a pilot study. Citing an example obtained from a study reported by Ahmad et al., (2021) regarding a questionnaire development study for developing the job satisfaction questionnaire, this newly-developed questionnaire had been subjected to validity testing, which included both convergent and discriminant validity; and these two types of validity are the fundamental aspects of construct validity.

Chapter 7

Other Issues Relevant to Questionnaire Development

To develop a new questionnaire from scratch can be a very challenging task. Previously, we had discussed the processes involved in developing four different types of questionnaires. In this chapter, the authors highlight a few issues relevant to questionnaire development.

7.1 Is there any other types of questionnaire?

In the existing literature particularly for the medical discipline, there are four main types of questionnaire as discussed in earlier sections. Therefore, the next question which might arise is that: “whether there are any other types of questionnaire for survey research in medical field?”

7.1.1 Knowledge, Attitude and Practice (KAP) questionnaire

A common type of study for medical research involves the use of a ‘Knowledge, Attitude and Practices (KAP)’ type of questionnaire. Although a KAP questionnaire is usually regarded as one questionnaire; it actually consists of three different categories such as Type 2 (to measure knowledge – “K”), Type 3 (to measure latent variable – “A”), and Type 1 (to collect facts and information – “P”). Since each domain (i.e. ‘Knowledge’, ‘Attitude’ and ‘Practices’) aims to collect a different type of information, each domain may require a different approach for validation. Thus, this type of questionnaire still falls under the four types of questionnaire as discussed earlier.

Similarly, every survey will be collecting different types of variables such as demographic profile of respondents and questions pertaining to knowledge, attitude and practice of the respondents. In this case, although the research study will be regarding both these sections as one set of questionnaire; in reality this set of questionnaire actually consists of two types of questionnaire: Type 1 (to collect facts or information) and Type 3 (to measure latent variable). This means that all these four fundamental types of questionnaire are being considered universally applicable to all types of survey research for the medical discipline.

However, it is always recommended to exercise caution by not combining any two (or more) of the four different types of questionnaire into a single set of questionnaire or within a single section. For example, a set of questionnaire has been drafted to consist of questions related to demographic profile (Type 1) and also questions to assess knowledge (Type 2) within the same section. Such a questionnaire is very difficult to be validated and should therefore be avoided.

7.1.2 To develop a framework or theoretical model by using the path analysis

This type of questionnaire is an extension from questionnaire type 4 – to measure latent variable. In previous sections, it was emphasized that the construct of latent variables should follow an ideal framework which is supported by existing theory retrieved from the contemporary literature. For example, let’s refer to Appendix 1 for an example of a proposed framework for job satisfaction. There are seven latent variables that are contributing to the concept of ‘job satisfaction’, all of which are being validated by using exploratory factor analysis and the cumulative evidence can then be strengthened by measuring the goodness of model fit via confirmatory factor analysis (Ahmad et al., 2020). Hence, this involves an evaluation of “many latent variables which are reduced to a single latent variable” or in other words, the model specifies a list of the significant domains that contribute to a single outcome.

However, there is a type of questionnaire that also measures latent variables but the model is showing a path relationship (example shown in Figure 13). Here, “User Satisfaction” is determined by four latent variables (i.e. Perceived Usefulness, Perceived Ease of Use, Behavioural Intention and Actual System Use) but the model is showing a path relationship in which there is no single dependent variable. For this type of questionnaire, the researcher will develop and design questions for each domain (e.g. five questions for each domain). Later, instead of using exploratory factor analysis, the researcher will apply confirmatory factor analysis to evaluate the model based on a path relationship diagram

(Figure 13). This technique is called path analysis. The model is considered valid if several relevant indicators such as Goodness-of-fit index (GFI), Adjusted goodness-of-fit test (AGFI), Comparative fit index (CFI) and Root-mean-square residual (RMR) have been proven to support the goodness of model fit.

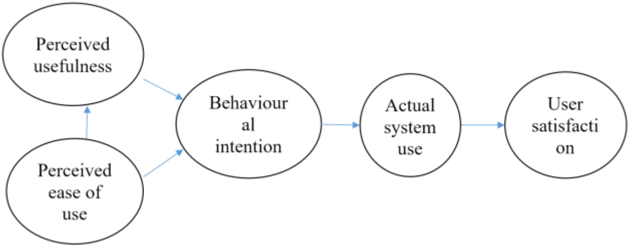


Figure 7.1: An example of a conceptual framework of user experience dimension that leads to the concept of user's satisfaction

7.1.3 Visual Analogue Score (VAS)

A Visual Analogue Scale (VAS) is a measurement instrument that aims to measure a characteristic of variables that is believed to range across a continuum of values and cannot be directly measured (Gould, 2001). It is often used in epidemiologic and clinical research to measure the intensity or frequency of various symptoms or outcomes (EuroQol, 1996; Dauphin et. al., 1999). For example, the amount of pain that a patient feels ranges across a continuum from none to an extreme amount of pain. The simplest VAS is a straight horizontal line of fixed length, for instance

100mm with a proportionate length of 10mm where every 10mm indicates different scale (e.g. 0 to 10) representing a continuum from worst (e.g. 0) to best (e.g. 10). The ends are defined as the extreme limits of the parameter to be measured (e.g. pain) which will be orientated from the left (worst) to the right (best) (Streiner & Norman, 1989). In some studies, horizontal scales are orientated from right to left; however, many investigators use a vertical VAS (Scott, 1976).

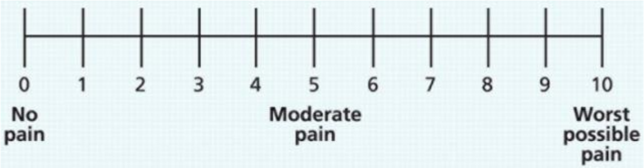


Figure 7.2: Example of VAS for measuring pain

The question now is: “Should VAS fall under which type of questionnaire?” Based on empirical considerations, VAS should fall under type 3 questionnaire which aims to measure latent variables. Previously, we learned that a few questions may have to be developed to represent a specific domain. So, VAS is the simplest form of question that can represent a specific domain (e.g: pain). Thus, the basis for validating VAS is not the same as that for validating a set of questions that represent specific domains as previously discussed. As for quantitative approach to validate a questionnaire that measures latent variables, normally the statistical techniques that we should be using will be exploratory factor analysis and/or structural equation modelling and/or RASCH analysis.

However, in order to validate the VAS, a list of suitable statistical techniques that can possibly be used includes known group comparison (independent or paired study design), convergent validity and predictive validity; and the choice of which statistical technique to use shall depend on the situation. Apart from its simplicity, another major strength of VAS is that its measurement will fall under the interval scale.

Therefore, besides having an element of continuum, it also has an element of fixed interval with the same size. Therefore, statistics such as mean and standard deviation can easily be computed from the VAS.

7.2 Sample size requirement

Sample size estimation is determined by three components: type I error (usually is fixed at 0.05), power (usually is fixed at 80.0%) and effect size. Thus, the effect sizes will need to be estimated beforehand by using different formula which shall depend on the study objectives.

Different validation studies may apply different statistical techniques. Therefore, researchers will need to first calculate or estimate the sample size requirement that is based on the statistical tests for which the questionnaire is being validated. There are many useful published guidelines on the estimation of sample size requirement for a wide variety of research purposes; and all of these are already widely available in the existing literature such as intra-correlation, kappa agreement, correlation test, sensitivity and specificity, multiple linear regression and analysis of covariance, logistic regression and exploratory factor analysis (Bujang et al., 2012; Bujang and Baharum, 2017a; Bujang and Baharum, 2017b; Bujang and Baharum, 2016; Bujang and Adnan, 2016, Bujang et al., 2017; Bujang et al., 2018; Bujang et al., 2019).

7.3 Questionnaire translation

If it is necessary to translate a newly-developed questionnaire into other language(s), proper translation technique (forward and backward translation) will need to be applied (Bujang et al., 2016; Bujang et al., 2017). In addition, questionnaire translation is also necessary when the target respondents are multi-racial and speak different languages like in Malaysia. Say, if the questionnaire has initially been developed in the Malay language. Then, the researcher may need to translate the questionnaire into Mandarin and Tamil for Chinese and Indian population respectively. It can be a very challenging task for translating a questionnaire which is designed to measure latent variables. This is because different ethnicities may perceive the latent variables differently due to different cultural practices and lifestyles (Omar et al., 2010; Adam et al., 2010). Therefore, the translators should also emphasize on the need for cross-cultural adaptation of the different languages since a direct word-by-word translation might not work well (Guilleman et al., 1993).

7.4 Number of items

Questionnaires with too few questions may not be sufficient to acquire sufficient information while questionnaires with too many items can over-burden the subjects which will potentially lead to missing values. In a questionnaire development, there is no limit to the number of questions. Preferably, the recommended number of items should be less than 100. However, it is dependent on the purpose of the research itself. For example, a questionnaire for patient registry questionnaire may collect more than 100 variables (NHAM, 2015). However, the researcher will have to consider the feasibility of capturing all the variables and the time needed for data collection and data entry. Otherwise, the researcher may have to simplify the form by reducing the total number of questions and/or items in order to make allowances for the constraints imposed by manpower (such as the total number of staff available) and financial resources (such as financial budget).

7.5 Modification of available questionnaire

Making modifications to an existing validated questionnaire or selecting questions from a few validated questionnaires and combining into one are not considered as questionnaire development even if some of the original questions have been amended. The owner of the questionnaire is still the original author(s) who was involved in the initial development of the questionnaire and the copyright licence is being granted to the publisher in most circumstances. Instead of making amendments on readily available validated questionnaires, it is recommended to develop a new questionnaire. In a scenario where the researcher intends to adopt a validated questionnaire from a foreign language for use in the local setting, a proper validation process will need to be conducted such as: to evaluate the content validity, to do forward and backward translation (from foreign language to local language and vice versa), to assess face validity, to conduct a pilot study mainly to assess the reliability of the questionnaire which will then be followed by a field test to assess the validity of the questionnaire.

7.6 To provide a manual for handling missing values

Missing values are common and need to be properly managed. The easiest way to handle the missing value is to declare it as missing value in the dataset. A specific code such as “9999” (something distinct and will not be confused with a valid value) can be applied to replace each and every missing values in the dataset. The other technique is to do imputation. In statistics, imputation is the process of replacing missing data with substituted values. There are many imputation techniques available such as replacing missing values with mean, mode or other specific technique. Most importantly, researchers need to provide a manual for handling missing values for a newly questionnaire that they developed. This is to prevent inconsistent handling of missing values in a questionnaire which can introduce bias in results.

Missing values in knowledge questionnaires can be problematic as it is impossible to impute the missing values in such case. As a result, the total score cannot be computed and thus becomes missing value. This also applies to questionnaires that measure latent variables since majority of this type of questionnaire have scores for overall items, domains or both the overall score and domain scores. Some studies suggest an imputation technique by using mean values but, imputation is not allowable if the extent of missing values reaches a certain limit (Kovacs, 1985). It is recommended that only 10% to 20% missing values out of the total number of questions can be imputed. Thus, researchers need to decide on the rule of handling missing values so that there is a standardized practice in administration of the same questionnaire.

Questionnaires for diagnostic research are usually based on facts or scientific knowledge. Each question assists the researcher to discriminate subjects with or without disease. Therefore, an imputation is discouraged to avoid bias in screening or diagnosing the subjects. In this type of questionnaire, it is often recommended for including a smaller total number of items so that missing values can be minimised. In general, missing values is a challenging issue but needs to be properly controlled. To avoid insufficient sample size due to missing data, it is advisable to account for 20% to 30% non-response rate when recruiting participants.

7.7 To state possible limitations of study

Developing a new questionnaire from scratch can often be highly complicated. The researcher may try to develop many relevant questions and to provide as much evidence as possible for establishing the validity of the newly-developed questionnaire. However, it is still possible to identify many study limitations which may occur at any of the stages of this research study: such as in determining the scope of study, in selecting a suitable study design, in choosing the sampling technique, and in establishing the subject eligibility criteria. Therefore, it is necessary to identify and report any limitations for questionnaire development study. By putting them into proper perspective, all these limitations will provide a fair appraisal for the overall utility and versatility of the questionnaire, and a valid justification for the degree of generalisability of the questionnaire for the target population, and also a list of useful recommendations for future research.

Chapter 8

Summary

In conclusion, this book has provided a lengthy discussion on the entire questionnaire development process for four main types of questionnaires, namely: (i)one that is designed to collect facts or information, (ii)to measure knowledge, (iii)to measure latent variables and (iv)for diagnostic research. Moreover, proper steps for developing a new questionnaire based on each of the four categories of questionnaire are also proposed. Each type of questionnaire will require a different approach for designing and developing the questions and also ultimately testing the reliability and validity of the questionnaire. However, discussion on the statistical treatment such as the process of statistical analysis and interpretation are not discussed in detail in this book since its emphasis is on the entire process of questionnaire development (as well as all the major processes involved in questionnaire development).

In addition, this book proposes a simple checklist (Table 8.1) that can serve as a guide to perform critical appraisal of a questionnaire development process and also as a guide for the researcher to plan and conduct a questionnaire development study. There are a total of 21 checklist items that cover the entire process of questionnaire development to enable the researcher to achieve at least the bare minimum standard requirements for a questionnaire development study.

Table 8.1: A detailed checklist that provides a step-by-step guide for conducting a questionnaire development study

No	Steps	Remark(s) (Y/N/Uncertain)
1	The title should emphasize on "Questionnaire development study…"	
	Introduction/background of study	
2	The underlying rationale and basis for developing a new questionnaire is clearly described.	
3	An extensive review of the existing literature and/or discussion on the background subject matter is provided. (*Especially for Type 3 questionnaire – Questionnaire designed to measure latent variable(s), discussion needs to highlight any related theory and/or conceptual framework and to place an emphasis on previous studies for development of a similar type of questionnaire if available*)	
4	The objective of study includes a determination of the reliability and validity of the newly developed questionnaire.	
	Methods	
5	The process involved in the development of the items or questions is clearly described, and all decisions related to it are well-justified.	
6	The measurement scale(s) and/or type of response is/are appropriate.	
7	The process for the content validity assessment has been adequately described.	

	(To state who is/are the panel(s) involved in the process of assessment of its content validity.)	
8	The result(s) obtained from content validity assessment is/are found to be satisfactory. *(To decide if further amendment of the questions will need to be made.)*	
9	The process of face validity assessment has been appropriately conducted. (*How was face validity assessed? Who are the subjects involved in this process?*)	
10	The result(s) obtained from face validity assessment is/are found to be satisfactory. *(To decide if further amendment of the questions are required to be made.)*	
11	Describe how the pilot study was conducted. (*How many subjects were involved? Who are the subjects involved in the pilot study?*)	
12	The result(s) obtained from the pilot study is/are found to be satisfactory. *(To decide if further amendments of the questions are required to be made.)*	
13	The research design (population and sample, sampling technique, sample size and statistical analysis) of the study is clearly described and is well-justified and appropriate.	

14*	At least one type of reliability testing is applied, and a suitable type of reliability testing has been selected for the newly developed questionnaire.	
15*	At least one type of validity testing is applied (besides content validity and face validity), and a suitable type of validity testing has been selected for the newly developed questionnaire.	
16	Is the approach of handling missing values appropriate?	
	Results	
17	State the total number of subjects or participants involved in the study and determine whether the number of participants is sufficient to test for the basis of a specific type of validity.	
18*	Are the results obtained from reliability assessment found to be satisfactory? *(To describe the results.)*	
19*	Are the results obtained from validity testing found to be satisfactory? *(To describe the results.)*	
	Discussion	
20	Is the newly developed questionnaire sufficiently valid and reliable to be used? *(To describe its versatility for use by the local population in term of its validity and reliability)*	
21	The limitation(s) of the study is/are stated (If available)	

Note: *Not necessary for type 1 questionnaire (i.e., questionnaire that is design to collect fact or information).

Remark note: To answer “yes” or “no” or “uncertain” and to provide the necessary justifications.

References

Adam, B., Ramli, M., Jamaiyah, H., Noor Azimah, M., & Khairani, O. (2010). Comparison of Family Environmental Scale (FES) Subscales between Malaysian Setting with the Original Dimension of FES. *Malaysian Journal of Psychiatry*, 19(1), 28–34.

Adnan TH, Mohamed Apandi M, Kamaruddin H, Salowi MA, Law KB, Haniff J, et al. (2018). Catquest-9SF questionnaire: validation of Malay and Chinese-language versions using Rasch analysis. *Health Qual Life Outcomes*, 16, 5. pmid:29304817

Ahmad NFD, Ren Jye AK, Zulkifli Z, Bujang MA. (2020). The development and validation of job satisfaction questionnaire for health workforce. *Malays J Med Sci*. 27(6), 128–143.

AERA, APA NMCE. (2014). The Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association

Berikai, P., Meyer, P.M., Kazlauskaite, R., Savoy, B., Kozik, K., & Fogelfeld, L. (2007). Gain in patients' knowledge of diabetes management targets is associated with better glycemic control. *Diabetes Care*, 30, 1587–1589.

Bollen, K.A. Latent variables in psychology and the social sciences. *Annu Rev Psychol*, 53(1), 605–634.

Bond TG, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences. Mahwah NJ: Lawrence Erlbaum Associates; 2001.

Boynton, P.M., & Greenhalgh, T. (2004). Hands-on guide to questionnaire research: selecting, designing, and developing your questionnaire. *BMJ*, 328, 1312–1315.

Bujang MA, Ab Ghani P, Soelar SA, Zulkifli NA (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) 2012*; Langkawi, Malaysia. 10–12 September 2012.

Bujang, M.A., & Baharum, N. (2016). Sample Size Guideline for Correlation Analysis. World *Journal of Social Science Research*, 3(1), 37–46.

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

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., & Baharum, N. (2017a) Guidelines of the minimum sample size requirements for Kappa agreement test. *Epidemiol, Biostat Public Health*, 14(2).

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 Orofac Sci*, 12, 1–11.

Bujang, M.A., Ismail, M., Hatta, N.K.B., Othman, S.H., Baharum, N.A, & 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, M.A., Sa'at, N., & Tg Abu Bakar Sidik, T.M.I. (2017). Determination of Minimum Sample Size Requirement for Multiple Linear Regression and Analysis of Covariance Based on Experimental and Non-Experimental Study. *Epidemiol, Biostat Public Health*, 14(3).

Bujang, M.A., Sa'at, N., Tg Abu Bakar Sidik, T.M.I., & Lim, C.J. (2018) Sample size guidelines for logistic regression from observational studies with large population: emphasis on the accuracy between statistics and parameters based on real life clinical data. *Malays J Med Sci*, 25(4), 122–130.

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 MA, Adnan TH, Mohd Hatta NKB, Ismail M, Lim CJ. (2018). A Revised Version of Diabetes Quality of Life Instrument Maintaining Domains for Satisfaction, Impact, and Worry. *Diabetes Res*. 5804687.

Bujang MA, Ghani PA, Soelar SA, Zulkifli NA, Omar ED (2019). Invalid skewed responses contribute to invalid factor solution in exploratory factor analysis: a validation approach using real-life data. *J Behav Health*, 8(4), 152–160. doi: 10.5455/jbh.20190628084939.

Brown JD. (1996). Testing in language programs. Upper Saddle River, NJ: Prentice Hall Regents.

Brown JD. (2000). What is construct validity? TALT, Testing and evaluation SIG Newsletter, 4 (2), 7-10.

Chung, F., Yegneswaran, B., Liao, P., Chung, S.A., Vairavanathan, S., Islam, S., et al. (2008). STOP questionnaire: a tool to screen patients for obstructive sleep apnea. *Anesthesiology*, 108, 812–821.

Cicchetti, & Domenic V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302

Dauphin AP et al. (1999). Bias and Precision in Visual Analogue Scales: A Randomized Controlled Trial. *American Journal of Epidemiology*, 150(10), 1117-1127

Davidson M. (2014) Known-Groups Validity. In: Michalos A.C. (eds) Encyclopedia of Quality of Life and Well-Being Research. Springer, Dordrecht.

EuroQol Group. EuroQol EQ5D user guide. Rotterdam: Rotterdam Centre for Health Policy and Law, Erasmus University, 1996.

Farney RJ, Walker BS, Farney RM, Snow GL, Walker JM. (2011). The STOP-Bang equivalent model and prediction of severity of obstructive sleep apnea: relation to polysomnographic measurements of the apnea/hypopnea index. J Clin Sleep Med, 7 (5), 459-465

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.

Gravetter, Frederick J.; Forzano, Lori-Ann B. (2012). *Research Methods for the Behavioral Sciences* (4th ed.). Belmont, Calif.: Wadsworth.

Gulabani, M.l., John, M., & Isaac, R. (2008). Knowledge of diabetes, its treatment and complications amongst diabetic patients in a tertiary care hospital. *Indian J Community Med*, 33, 204–206.

Gould D. et al. (2001) Visual Analogue Scale (VAS). *Journal of Clinical Nursing*, 10, 697-706

Herzberg, F., Mausner, B. & Snyderman, B.B. (1967), The Motivation to Work, 2nd ed., Wiley, New York, NY.

Holden, Ronald B. (2010). *"Face validity"*. In Weiner, Irving B.; Craighead, W. Edward (eds.). The Corsini Encyclopedia of Psychology (4th ed.). Hoboken, New Jersey: Wiley.

Jacobson, A.M., De Groot, M., & Samson, J.A. (1994). The evaluation of two measures of quality of life in patients with type I and type II diabetes. *Diabetes Care*, 17(4), 267–274.

Kissi, J., Dai B., Dogbe, CSK. (2020). Predictive factors of physicians' satisfaction with telemedicine services acceptance. *Health Informatics Journal*. 2020, 26(3), 1866–1880

Kline, P. (1993). Handbook of Psychological Testing. 2nd Eds, Routledge, London

Kovacs, M. (1985). The Children's Depression Inventory. *Psychopharmacology Bulletin*, 21, 995–998.

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 patients. *American Journal of Psychiatry*, 156, 1522–1528.

Lovibond, S.H., & Lovibond, P.F. (1995). Manual for the depression anxiety stress scales. Psychology Foundation, Sydney.

McColl, E., & Thomas, R. (2000). The use and design of questionnaires. Royal College of General Practitioners, London.

Wan-Ahmad, W.A., & Sim, K.H., (Eds). The Annual Report of the NCVD-ACS Registry 2011-2013 Malaysia. National Heart Association of Malaysia (NHAM), Malaysia.

Netzer, N.C., Stoohs, R.A., Netzer, C.M., Clark, K., & Strohl, K.P. (1999). Using the Berlin questionnaire to identify patients at risk for the sleep apnea syndrome. *Ann Intern Med*, 131, 485–489.

Nevo, B. (1985). Face validity revisited. Journal of Educational Measurement, 22(4), 287-293.

Nunnally, J. C. (1967). Psychometric theory. New York, NY: McGraw-Hill.

Omar, K., Musa, R., Haniff, J., Azimah, N., Bujang, A., & Mohamed Fadhlullah, F. (2010). Reliability of Bahasa Malaysia version of family environment scale and its measurement issues. *ASEAN Journal of Psychiatry*, 11(1), 1–12.

Oxford Dictionaries. Oxford University Press: 2018. https://en.oxforddictionaries.com/definition/questionnaire. (assessed 24 October 2018).

Peduzzi, P., Concato, J., Feinstein, A.R., & Holford, T.R. (1995). Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol*, 48, 1503–1510.

Polit DF, Beck CT (2006) The content validity index: are you sure you know what's being reported? Critique and recommendations. *Research in nursing & health*, 29(5): 489–497.

Rattray, J., & Jones, C.M. (2007). Essential elements of questionnaire design and development. *J Clin Nurs*, 16, 234–243.

Rubio DM, Berg-Weger M, Tebb SS, Lee ES, Rauch S. Objectifying content validity: Conducting a content validity study in social work research. Social Work Research. 2003;27(2):94–104.

Secolsky, C. (1987). On the direct measurement of face validity: A comment on Nevo.Journal of Education Measurement, 24(1), 82–83

Scott J, Huskisson EC. (1976). Graphic representation of pain. *Pain*, 2, 175-84

Streiner DL, Norman GR. (1989). Health measurement scales: a practical guide to their development and use. New York; Oxford University Press.

Tennant, A., McKenna, S.P., & Hagell, P. (2004). Application of Rasch analysis in the development of quality of life measures. *Value Health*, 7(Suppl 1), S22–S26.

Thriemer, K., Katuala, Y., Batoko, B., Alworonga, J.P., Devlieger, H., Van Geet, C., et. al. (2013). Antibiotic prescribing in DR Congo: a knowledge, attitude and practice survey among medical doctors and students. *PLoS One*, 8, e55495.

Tiong, X.T., Abdullah, N.S.S., & Bujang, M.A. (2018). Validation of the Kessler's psychological distress scale (K10 & K6) in a Malaysian population. *ASEAN J Psychiatry*, 19.

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. Hardwood Academic, Switzerland.

The WHOQOL Group (1998). Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychol Med*, 28(3), 551-558.

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

Wilkinson, D., & Birmingham, P. (2003). Using research instruments a guide for researchers, Routledge Falmer, London, UK.

Wirtz, V.J., Taxis, K., & Dreser, A. (2009). Pharmacy customers' knowledge of side effects of purchased medicines in Mexico. *Trop Med Int Health*, 14(1), 93-100.

World Health Organization (WHO). http://www.who.int/. (accessed 31 Oct 2018)

Yaghmale F. (2003). Content validity and its estimation. *Journal of Medical Education*, 3, 25-27.

Young, T., Peppard, P.E., & Gottlieb, D.J. (2002). Epidemiology of obstructive sleep apnea: a population health perspective. *Am J Respir Crit Care Med*, 165, 1217–1239.

Appendices

Appendix 1:

A conceptual framework for job satisfaction questionnaire (JSQ) derived from Ahmad et al., (2020)

Appendix 2

Example of presentation to support evidence for knowledge questions

Appendix 1

A conceptual framework for job satisfaction questionnaire (JSQ) derived from Ahmad et al., (2020)

Appendix 1 is referring to a conceptual framework for job satisfaction. Based on this model, there are seven domains contributing to job satisfaction such as leadership, reward and recognition, empowerment, training and development, working hours, communication and working condition.

Appendix 2

Example of presentation to support evidence for knowledge questions

No.	Soalan-soalan/Questions	Rujukan-rujukan/References
1.	Apakah itu plak gigi?/ *What is dental plaque*?	Newman, M. G., Takei, H. H., Klokkevold, P. R., & Carranza, F. A. (2019). Newman and Carranza’s Clinical periodontology. In *Newman and Carranza's clinical periodontology* (pp. 119-150). Philadelphia, PA: Elsevier.

		Lang, N. P., & Karring, T. (2008). Oral Biofilms and Calculus. In *Clinical periodontology and implant dentistry* (5th ed., Vol. 1, pp. 187-196). Oxford, UK: Blackwell Munksgaard.
		.
2.	Plak boleh menyebabkan apa?/ *What can plaque cause?*	Newman, M., Takei, H., Klokkevold, P., & Carranza, F. (2019). Newman and carranza's Clinical periodontology. In *Newman and Carranza's clinical periodontology* (p. 122). Philadelphia, PA: Elsevier.

		Lang, N., & Karring, T. (2008). Microbiology. In *Clinical periodontology and implant dentistry* (5th ed., Vol. 1, p. 208). Oxford, UK: Blackwell Munksgaard.
		.
3.	Di antara berikut, yang manakah merupakan tanda-tanda awal penyakit gusi?/ Which of the following are early signs of gum disease?	Newman, M., Takei, H., Klokkevold, P., & Carranza, F. (2019). Clinical Features of Gingivitis. In *Newman and Carranza's clinical periodontology* (13th ed., pp. 248-255). Philadelphia, PA: Elsevier.

		Lang, N., & Karring, T. (2008). Plaque-Induced Gingival Diseases. In *Clinical periodontology and implant dentistry* (5th ed., Vol. 1, pp. 405-407). Oxford, UK: Blackwell Munksgaard.
		.
4.	Adakah penyakit gusi dapat dirawat? / *Can gum disease be treated?*	Newman, M., Takei, H., Klokkevold, P., & Carranza, F. (2019). The Treatment Plan. In *Newman and Carranza's clinical periodontology* (13th ed., pp. 426-430). Philadelphia, PA: Elsevier.

		Lang, N., & Karring, T. (2008). Treatment of plaque-induced gingival diseases. In *Clinical periodontology and implant dentistry* (5th ed., Vol. 1, p. 414). Oxford, UK: Blackwell Munksgaard.
		.
5	Adakah penyakit gusi boleh berjangkit? / *Is gum disease contagious?*	Newman, M., Takei, H., Klokkevold, P., & Carranza, F. (2019). Biofilm and Periodontal Microbiology. In *Newman and Carranza's clinical periodontology* (13th ed., pp. 131-133). Philadelphia, PA: Elsevier.

		Lang, N., & Karring, T. (2008). Periodontal Pathology. In *Clinical periodontology and implant dentistry* (5th ed., Vol. 1, pp. 466-468). Oxford, UK: Blackwell Munksgaard.
		.
6.	Adakah penyakit gusi boleh menyebabkan gigi longgar? / *Can gum disease lead to loosened/mobility of teeth?*	Newman, M., Takei, H., Klokkevold, P., & Carranza, F. (2019). Tooth Mobility. In *Newman and Carranza's clinical periodontology* (13th ed., pp. 391-392). Philadelphia, PA: Elsevier.

		Lang, N., & Karring, T. (2008). Tooth mobility crown excursion/root displacement. In *Clinical periodontology and implant dentistry* (5th ed., Vol. 1, pp. 1125-1137). Oxford, UK: Blackwell Munksgaard.
		.
7.	Pada peringkat umur berapakah seseorang itu mudah mendapat penyakit gusi? / *At what age, a person is prone to have gum disease?*	Newman, M., Takei, H., Klokkevold, P., & Carranza, F. (2019). Gingival Disease in Childhood. In *Newman and Carranza's clinical periodontology* (13th ed., pp. 277-

		286). Philadelphia, PA: Elsevier. Newman, M., Takei, H., Klokkevold, P., & Carranza, F. (2019). Aging and the Periodontium. In *Newman and Carranza's clinical periodontology* (13th ed., pp. 50-54). Philadelphia, PA: Elsevier. Lang, N., & Karring, T. (2008). Supportive Periodontal Therapy (SPT). In *Clinical periodontology and implant dentistry*(5th ed., Vol. 2, p. 1307). Oxford, UK:

		Blackwell Munksgaard.
		.
8.	Adakah seseorang itu lebih mudah mendapat penyakit gusi semasa mengandung?/ *Is it easier to get gum disease during pregnancy?*	Newman, M., Takei, H., Klokkevold, P., & Carranza, F. (2019). Periodontal Therapy in the Female Patient. In *Newman and Carranza's clinical periodontology* (13th ed., pp. 467-472). Philadelphia, PA: Elsevier. Lang, N., & Karring, T. (2008). Periodontal Disease as a Risk for Systemic Disease. In *Clinical periodontology and implant dentistry* (5th ed., Vol. 1, pp.

		480-486). Oxford, UK: Blackwell Munksgaard.
		.
9.	Adakah merokok boleh meningkatkan risiko penyakit gusi? / *Does smoking increase the risk of having gum disease?*	Newman, M., Takei, H., Klokkevold, P., & Carranza, F. (2019). Smoking and Periodontal Disease. In *Newman and Carranza's clinical periodontology* (13th ed., pp. 181-189). Philadelphia, PA: Elsevier. Lang, N., & Karring, T. (2008). Modifying Factors. In *Clinical periodontology and implant dentistry* (5th ed., Vol. 1, pp. 316-322). Oxford,

		UK: Blackwell Munksgaard.
		.
10.	Adakah seseorang yang mempunyai penyakit kencing manis lebih mudah untuk mendapat penyakit gusi? / *Does an individual with diabetes is more likely to develop gum disease?*	Newman, M., Takei, H., Klokkevold, P., & Carranza, F. (2019). Influence of Systemic Conditions. In *Newman and Carranza's clinical periodontology* (13th ed., pp. 208-212). Philadelphia, PA: Elsevier. Lang, N., & Karring, T. (2008). Periodontal Disease as a Risk for Systemic Disease. In *Clinical periodontology and implant dentistry* (5th ed., Vol. 1, pp.

		486-488). Oxford, UK: Blackwell Munksgaard.
		.
11.	Bagaimana anda boleh mencegah penyakit gusi?/ How can you prevent gum disease?	American Academy of Periodontology: Position paper: guidelines for periodontal therapy, J Periodontol 72:2001, 1624. Kinane DF: The role of interdental cleaning in effective plaque control: need for interdental cleaning in primary and secondary prevention. In Lang NP, Ättstrom R, Löe H, editors: Proceedings of the European workshop on mechanical

		plaque control, Chicago, 1998, Quintessence. Lang, N., & Karring, T. (2008). Initial Periodontal Therapy (Infection Control). In *Clinical periodontology and implant dentistry* (5th ed., Vol. 2, pp. 738-739). Oxford, UK: Blackwell Munksgaard.
		.
12.	Adakah dengan penjagaan gigi yang baik, kita boleh mengekalkan gigi di dalam mulut sepanjang hayat? / *Is it true that we can preserve our*	Newman, M., Takei, H., Klokkevold, P., & Carranza, F. (2019). Tooth Mortality. In *Newman and Carranza's clinical periodontology* (13th ed., pp. 726-

	natural teeth with proper oral healthcare for a lifetime?	731). Philadelphia, PA: Elsevier. Lang, N., & Karring, T. (2008). Supportive Care. In *Clinical periodontology and implant dentistry* (5th ed., Vol. 2, pp. 1302-1303). Oxford, UK: Blackwell Munksgaard.

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, reviewer for number of peer-reviewed journals and also one of the former and active members in 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
Tel: 082-276820 & Fax: 082-276823

Madam Shirin Tan Hui

Madam Shirin Tan Hui is a Research Pharmacist at the Institute for Clinical Research (ICR), Ministry of Health, Malaysia. She graduated with Bachelor of Pharmacy (Hons) from University of Queensland, Australia in 2011. She works as a Neonatal Intensive Care Unit (NICU) Clinical Pharmacist and Clinical Research Centre (CRC) Admin Manager for 3 years. Currently, she is working as a Research Pharmacist in CRC, Sarawak General Hospital. Since started working in 2012, she has been involved in 12 investigator-initiated research. Her ongoing research interests involve translational research, patient-centered care aiming at individualized treatment for each patient. She believes that qualitative studies including questionnaires play an important role in understanding gaps in the needs, expectations and knowledge of patients which will in turn improve management of patients and their outcomes.

Contact details
Madam Shirin Tan Hui
Pharmacist,
Clinical Research Centre, Sarawak General Hospital
Email address: shirin_hui88@yahoo.com
Tel: 082-276820 & Fax: 082-276823

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