Enhancing Questionnaire Design, Development and Testing through Standardized Approach

Sendil Mourougan[1], Dr. K. Sethuraman[2]

[1](Department of Business Administration, Annamalai university, India)
[2](Department of Business Administration, Annamalai university, India)

Abstract: *The aim of this paper is to raise awareness regarding the issues in questionnaire development and to provide strategies to enable researchers to design and develop their own measure and evaluate the quality of existing measures. The theoretical and methodological issues associated with questionnaire design and development is evaluated. The range of available scales is presented and strategies to enable question generation and development are discussed. Issues of reliability and validity are explored using item analysis and exploratory factor analysis.*

Keywords: *Questionnaire Design, Questionnaire Development, Questionnaire Testing, Validity, Realiability*

I. Introduction

The use of questionnaires as a method of data collection both nationally and internationally has increased in recent years. Questionnaires are used to enable the collection of information in a standardized manner which, when gathered from a representative sample of a defined population, allows the inference of results to the wider population. This is important when we want to evaluate the effectiveness of a system. Central to the understanding of results derived from questionnaires are the issues of reliability and validity which underpin questionnaire development from item generation to subsequent result analysis. This paper provides an analysis of key methodological issues from item generation to result evaluation. It assists researchers to develop their own measure or evaluate the work of others. Issues of reliability and validity are explored using item analysis and exploratory factor analytic techniques.

II. Questionnaire Design

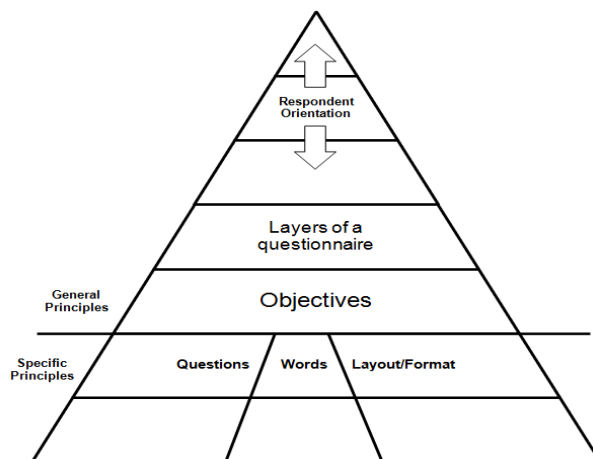**2.1 Framework for Questionnaire Design**

The overall framework for questionnaire design is shown in Figure 1. It is represented by a triangle, or pyramid, with general principles at the top and specific principles at the bottom. At the apex of this pyramid is the concept of respondent orientation, and at the base, specific principles of question wording and graphic design. The idea which this representation is intended to convey is that there are a small number of general principles of questionnaire design which broaden out into a larger number of specific principles. The pyramid is deliberately divided into general and specific principles to illustrate the contention that much of what is written about questionnaire design starts at the level of what the researcher has defined as specific principles.

2. 2 General Principles

General Principle-1: Respondent Orientation

If there is a single, fundamental principle of questionnaire design, it is that the respondent defines what can be done: the types of questions that can be asked; the types of words that can be used; the concepts that can be explored; the methodology that can be employed. To find out what is in respondents' minds, interviewer should ask them questions they can truthfully answer about their physical environment, their consciousness, their knowledge and their past behavior. Attitude and opinion questions play only a minor role in questionnaire design. There is no evidence that attitudes are good predictors of behavior. As a general principle, attitude questions should be replaced with questions about respondents' environment, consciousness, knowledge and behavior. Information on respondents' environment, consciousness and knowledge can be used to weight their opinions to give a more realistic perspective on the views of the population sampled.

By environment, it is meant that the physical aspects of respondents' lives over which they have little control, factors such as age, sex, socio-economic status, race, locale and mobility. Respondent consciousness, or awareness, means whether or not respondents can understand the implications of their answers; whether they can fit the pieces together to form a coherent picture. Knowledge simply means whether respondents really know what the researcher is asking them about. All questionnaires reflect their designer's view of the world, no matter how objective the researcher has attempted to be. Intellectually, good questionnaire designers understand this and attempt to maintain a detached objectivity. Let the respondent tell what he or she means, and don't impose values, perceptions or language on the respondent.

Figure 1. A Framework for Questionnaire Design

General Principle-2: Layers of a questionnaire
A questionnaire is not simply a series of questions, nor is a question merely a series of words. A questionnaire is a structure consisting of several different layers which must be simultaneously integrated into an overall entity whose properties are greater than the sum of the properties of the individual layers. The four layers of a questionnaire are: objectives, questions, words and layout.

General Principle-3: Objectives
The researcher cannot begin to formulate questions unless he knows what he wants to accomplish with his questions and words.

1.3 Specific Principles
Specific questionnaire design principles are divided into three sections, each concerned with one questionnaire layers: question design, question wording, and formatting or layout. Within each section the principles outlined become progressively more and more specific. Though the sections are shown as separate and are treated separately in the discussion that follows, this is largely a matter of convenience.
Like all aspects of questionnaire design, one element cannot be dealt with in isolation of the others. This becomes increasingly apparent as the principles outlined become more specific. Consequently, the section to which some specific aspects of questionnaire design are allocated is somewhat arbitrary. The specific principles of question design, question wording, and layout or formatting are described below:

Specific Principle-1: Questions
A good question is one that produces answers that are reliable and valid measures of something that we want to describe (Fowler 1995). A bad question is one that obscures, prohibits or distorts the communication from respondent to researcher, and vice versa.
- Use closed rather than open-ended questions wherever possible. Then at least the context is the same for all respondents. However, the pattern of responses for a closed question is critically dependent on the answer set presented; the inclusion of "other" will not compensate for the omission of an important answer, and if an unimportant answer is included, its importance is likely to be over estimated.
- The likelihood of question context effects needs to be anticipated so that as far as possible they are deliberate rather than unanticipated. This can be done by understanding the rules that govern conversations and social encounters between strangers. (Sudman, Bradburn & Schwarz 1996)
- The more specific a question the easier it will be for a respondent to understand what the researcher wants.
- Longer questions will improve recall if respondents use recall and count strategies, but not if they estimate.
- Avoid asking questions in a context that is likely to deviate strongly from the probable context in which an issue will be considered.
- If an anticipated context effect is undesirable, either change the context by omitting questions that may have an undesirable impact or by putting the crucial question in the first position.
- To eliminate undesirable response order effects, consider using open ended rather than closed questions.
- In personal interviews, the introduction given by the interviewer to a question or series of questions can be used to increase, reduce or eliminate context effects, as desired by the researcher. In self-administrated questionnaires, the same result can be obtained by including printed instructions, putting the questions together in a box on the same page or putting the questions on different pages. (Sudman, Bradburn &

Schwarz 1996)
- Use forced choice, not agree-disagree attitude questions. (Converse & Presser 1982)
- Longer questions may produce longer responses. (Sudman & Bradburn 1982)

Specific Principle-2: Words

Question wording variations generally have little impact on the stability of survey results. They become significant when they introduce or tap a different concept or reality or emotional level surrounding an issue. In general, keep questions short, simple and concrete. Avoid:
- unfamiliar or difficult words
- many information-carrying words in one question
- words that sound like something else (partial/impartial)
- broad concepts (e.g., children, the government)
- a qualifying clause at the end of a question
- two questions in one
- suggestions or leadingness
- big names
- questions that call for a lot of respondent effort (Belson 1981)
- General questions are prone to assimilation effects and specific questions to contrast effects. In either case, context effects will be much smaller, or vanish altogether if a previously-formed judgement or substantial amounts of relevant information are chronically assembled in memory. This is not something the researcher can easily manipulate.
- If response alternatives to a question are necessary, their order may be randomised unless there is a natural ordering.
- Avoid items or stimuli that receive extremely positive or negative ratings when asking for judgement about a series of items, because items that follow will be strongly affected.
- Even when the wording of the question and the ordering of the response alternatives are identical, a visual presentation mode is likely to result in a different pattern of response order effects than will an auditory presentation mode. (Sudman, Bradburn & Schwarz 1996)
- A no opinion option should always be offered.
- When measuring attitudes you should omit the middle alternative and measure intensity.

Specific Principle-3: Layout

Questionnaires should be designed to make the task of reading questions, following instructions and recording answers as easy as possible for interviewers and respondents. (Fowler 1995)
- Questions should follow a logical sequence. Commonly this is in the form of a downward funnel; general and non threatening questions first, followed by more specific, more personal ones. However, other pathways are possible and sometimes more desirable.
- Formatting should meet the respondent's needs first, the interviewer's needs second and the researcher's or data processor's needs last.
- Formatting and graphic design are particularly important for self-completion questionnaires, to motivate and guide respondents through a questionnaire and achieve good information organization.
- Present information in a format that respondents are accustomed to reading.
- Present only the most relevant information using graphical design features and composition.
- Pique respondent's interest early in the questionnaire.
- Dominantly feature questions over additional explanatory information.
- Include in each question all of the relevant information necessary for respondents to answer it, rather than specifying information in a subsequent instruction.
- Vertically align the questions and response categories.
- If incorporating needed information into a question makes it too complicated to understand, then provide accompanying instructions at the place where they are needed.
- Utilize single-formats rather than multi-task formats.
- Utilize single-question formats rather than matrix-question formats.
- Make headings and instructions at the top of the page more prominent than those in the middle of the page.
- Provide directions in a natural reading format and utilize graphical design features to make the directions more salient.
- Utilize graphical design techniques to establish a clear path through the questionnaire for the respondent to follow.

- Avoid using the same design feature to request different respondent actions.
- Utilize variability in design features judiciously.
- Visually emphasize information the respondent needs to see and de-emphasize information which the respondent does not need to see.
- Utilize graphical layout of questions on the page to distinguish among different types of question structures; maintain consistency within types.
- Provide descriptive captions either above, beneath, or to the right of blank answer spaces and utilize appropriate signs or symbols whenever numbers are requested.
- Utilize dominant graphical markings to provide the most important information needed by respondents to guide them through the answering process.
- Avoid the separation of questions through the use of lines and rectangles, in favour of an open format.
- Structure and organize the questionnaire in such a way that it, first, makes sense to respondents and, second, avoids leaving the choice of the order in which questions get answered up to the respondents. (Jenkins & Dillman 1973)
- Questionnaire cover design (see: Dillman & Dillman 1995; Gendall 1996)
- Covering letters (see: Gendall *et al* 1995)

III. Questionnaire Development

3.1 Measurement

Researchers use questionnaires to measure knowledge, attitudes, emotion, cognition, intention or behaviour. This approach captures the self-reported observations of the individual and is commonly used to measure participant perceptions. When developing a questionnaire, items or questions are generated that require the respondent to respond to a series of questions or statements. Participant responses are then converted into numerical form and statistically analysed. These items must reliably operationalize the key concepts detailed within specific research questions and must, in turn, be relevant and acceptable to the target group. The main benefits of such a method of data collection are that questionnaires are usually relatively quick to complete, are relatively economical and are usually easy to analyse (Bowling 1997). This approach to data generation is not without criticism. It assumes that the researcher and respondents share underlying assumptions about language and interpret statement wording in a similar manner. Closed questions which are commonly used may restrict the depth of participant response (Bowling 1997) and thus the quality of data collected may be diminished or incomplete. Questionnaire-based methods are, therefore, not the method of choice where little is known about a subject or topic area. In such an instance, qualitative methods may be more appropriate.

3.2 Scales

There are a range of scales and response styles that may be used when developing a questionnaire. These produce different types or levels of data (see Table 1) and this will influence the analysis options. Therefore, when developing a new measure, it is important to be clear which scale and response format to use. Frequency scales may be used when it is important to establish how often a target behaviour or event has occurred.

Table 1. Stages in questionnaire development: item generation and scale construction

Questionnaire Development	Key issues
What will the questionnaire measure?	Knowledge Attitude/beliefs/intention Cognition Emotion Behaviour
What types of scale can be used?	Frequency Thurstone Rasch Guttman Mokken Likert type Multiple choice
How do I generate items from my questionnaire?	Ensure relevance of items? Wording issues Which response format is best? Which types of questions are possible? Free text options? Does your measure have subscales? Questionnaire layout

Within research Likert-type or frequency scales are most commonly used. These scales use fixed choice response formats and are designed to measure attitudes or opinions (Bowling 1997, Burns & Grove 1997). These ordinal scales measure levels of agreement/disagreement. A Likert-type scale assumes that the strength/intensity of experience is linear, i.e. on a continuum from strongly agree to strongly disagree, and makes the assumption that attitudes can be measured. Respondents may be offered a choice of five to seven or even nine precoded responses with the neutral point being neither agree nor disagree.

There is no assumption made that equal intervals exist between the points on the scale; however, they can indicate the relative ordering of an individual's response to an item. While this is perhaps too simplistic, until an alternative model is developed, it is a relatively easy and appropriate method to use (Oppenheim 1992). Some controversy exists as to whether a neutral point should be offered. If this option is removed, this forces the respondent to choose a response, which may lead to respondent irritation and increase non-response bias (Burns & Grove 1997).

It is acceptable to treat scores from this type of response format as interval data to allow the use of common parametric tests (Ferguson & Cox 1993, Polgar & Thomas 1995, Bowling 1997, Burns &Grove 1997). As with any data set, subsequent statistical analysis should be determined by the normality of distribution of the data and whether the data meets the underlying assumptions of the proposed statistical test.

It would be unusual to develop a questionnaire that relied upon a single-item response, and multi-item scales are generally used in preference to single-item scales to avoid bias, misinterpretation and reduce measurement error (Bowling 1997, Burns & Grove 1997). Such questionnaires have a number of subscales that 'tap' into the main construct being measured.

3.3 Question Generation, Working and Order

The generation of questions during questionnaire development requires considerable pilot work to refine wording and content. To assure face or content validity, items can be generated from a number of sources including consultation with experts in the field, proposed respondents and review of associated literature (Priest et al. 1995, Bowling 1997; see Table 1). In addition, a key strategy in question generation is to revisit the research questions frequently and to ensure that items reflect these and remain relevant (Oppenheim 1992, Bowling 1997). It is during this stage that the proposed subscales of a questionnaire are identified (Ferguson & Cox 1993) and to ensure that questions are representative of these. The question and factor analysis stages of the questionnaire development process may then be used to establish if such items are indeed representative of the expected subscale or factor.

The type of question, language used and order of items may all bias response. Consideration should be given to the order in which questions are presented, e.g. it is best to avoid presenting controversial or emotive questions at the beginning of the questionnaire. To engage participants and prevent boredom, data may be presented at the end. Certain questions should be avoided, e.g. those that lead or include double negatives or double-barreled questions (Bowling 1997). A mixture of both positively and negatively worded items may minimize the danger of acquiescent response bias, i.e. the tendency for respondents to agree with a statement, or respond in the same way to questions.

To allow respondents to expand upon answers and provide more in-depth responses, free text response or open questions may be included. Respondents may welcome this opportunity. However, while this approach can provide the interviewer with rich data, such material can be difficult to analyse and interpret (Polgar & Thomas 1995). However, these problems may be outweighed by the benefits of including this option and can be especially useful in the early development of a questionnaire. Free text comments can inform future questionnaire development by identifying poorly constructed questions or new questions for future inclusion.

IV. Questionnaire Testing

4.1 Pilot Testing

It is important to ensure that sufficient pilot work is carried out during the development of a new measure. This will identify questions that lack clarity or that may not be appropriate for, or discriminate between, respondents. Ideally, the questionnaire should be piloted on a smaller sample of intended respondents, but with a sample size sufficient to perform systematic appraisal of its performance. Item analysis is one way to pilot a questionnaire. This provides a range of simple heuristics on item retention or deletion, see Table 2.

Table 2 Stages in questionnaire development: piloting the questionnaire: item analysis

Questionnaire Development	Key Issues
Piloting the questionnaire: **Item analysis**	Spread of responses across options: Initial psychometric analysis: Clarity and relevance of items: Items deemed theoretically important: Is your measure affected by social desirability bias?
Reliability	Internal consistency

	Test-retest
	Inter-observer
Validity	Face or content
	Concurrent or discriminant
	Predictive

High endorsement of an option within a particular item suggests poor discriminatory power or the redundancy of an item that requires deletion (Priest et al. 1995). Alternatively, a Cronbach's alpha < 0.70 may suggest that items in a questionnaire or subscale are poorly grouped.

To identify specific items that do not add to the explanatory power of the questionnaire or subscale an item-total correlation cut-off of <0.3 can be used (Ferketich 1991, Kline 1993). However, it is important when revising the questionnaire to refer constantly to the original research questions that are being addressed and retain items that are thought to reflect the underlying theoretical domains of the questionnaire despite poor psychometric analysis. Problem items may also be identified because of high levels of non-response.

4.2 Demonstrating Reliability (Cronbach's Alpha)

It is essential that the reliability of a developing questionnaire can be demonstrated. Reliability refers to the repeatability, stability or internal consistency of a questionnaire (Jack & Clarke 1998). One of the most common ways to demonstrate this uses the Cronbach's a statistic. This statistic uses interitem correlations to determine whether constituent items are measuring the same domain (Bowling 1997, Bryman & Cramer 1997, Jack & Clarke 1998). If the items show good internal consistency, Cronbach's alpha should exceed 0.70 for a developing questionnaire or 0.80 for a more established questionnaire (Bowling 1997, Bryman & Cramer 1997). It is usual to report the Cronbach's alpha statistic for the separate domains within a questionnaire rather for the entire questionnaire.

Item-total correlations can also be used to assess internal consistency. If the items are measuring the same underlying concept then each item should correlate with the total score from the questionnaire or domain (Priest et al. 1995). This score can be biased, especially in small sample sizes, as the item itself is included in the total score (Kline 1993). Therefore, to reduce this bias, a corrected item-total correlation should be calculated. This removes the score from the item from the total score from the questionnaire or domain (Bowling 1997) prior to the correlation. Kline (1993) recommends deleting any questionnaire item with a corrected item-total correlation of < 0.3. Item analysis using inter-item correlations will also identify those items that are too similar. High inter-item correlations (>0.8) suggest that these are indeed repetitions of each other (sometimes referred to as bloated specifics) and are in essence asking the same question (Ferketich 1991, Kline 1993).

Test–retest reliability can assess stability of a measure over time and this should be included in the process of any questionnaire development. This is of particular importance if the intended use of the measure is to assess change over time or responsiveness.

4.3 Demonstrating Validity
Validity refers to whether a questionnaire is measuring what it purports to (Bryman & Cramer 1997). While this can be difficult to establish, demonstrating the validity of a developing measure is vital. There are several different types of validity (Polgar & Thomas 1995, Bowling 1997, Bryman & Cramer 1997).

4.3.1 Content Validity
Content validity (or face validity) refers to expert opinion concerning whether the scale items represent the proposed domains or concepts the questionnaire is intended to measure. This is an initial step in establishing validity, but is not sufficient by itself.

4.3.2 Convergent (or) Concurrent Validity
Convergent (or concurrent) and discriminant validity must also demonstrated by correlating the measure with related and/or dissimilar measures (Bowling 1997). When developing a questionnaire it is, therefore, important to include, within the research design, additional established measures with proven validity against which to test the developing questionnaire.

4.3.3 Construct Validity (Factor Analysis)
Construct validity relates to how well the items in the questionnaire represent the underlying conceptual structure. Factor analysis is one statistical technique that can be used to determine the constructs or domains within the developing measure. This approach can, therefore, contribute to establishing construct validity.

4. 4 Further Development: Exploratory Factor Analysis (N > 100)
　　　　Following initial pilot work and item deletion, the questionnaire should be administered to a sample of sufficient size to allow exploratory factor analytic techniques to be performed. Ferguson and Cox (1993) suggest that 100 respondents is the absolute minimum number to be able to undertake this analysis. However, others would suggest that this is insufficient and a rule of thumb would be five respondents per item (Bryman & Cramer 1997). This type of analysis must follow a predefined and systematic analytic sequence (Ferguson & Cox 1993).

4.5 Principal Component Analysis (PCA)
　　　　Principal components analysis (PCA) explores the interrelationship of variables. It provides a basis for the removal of redundant or unnecessary items in a developing measure (Anthony 1999) and can identify the associated underlying concepts, domains or subscales of a questionnaire. The terms of factor analysis and PCA are often used synonymously in this context. In practice, however, PCA is most commonly used. Rarely is a questionnaire uni-dimensional and PCA usually identifies the presence of one principal component that accounts for most of the variance and subsequent components that account for less and less. In the initial PCA analysis of an unrotated solution, most items should 'load', i.e. correlate with the first component. This can make interpretation of results difficult (Kline 1994), and to assist the interpretation of a factor solution, rotation of factors (components) is often performed. This should be a standard option on statistical packages, e.g. Statistical Package for Social Scientists (SPSS Inc., Chicago, IL, USA). Factor rotation maximizes the loadings of variables with a strong association with a factor, and minimizes those with a weaker one (Oppenheim 1992) and often helps make sense of the proposed factor structure. Varimax rotation, which is an orthogonal rotation (i.e. one in which the factors do not correlate), is often used, particularly if the proposed factors are thought to be independent of each other (Ferguson &Cox 1993). However, oblimin rotation may be used, when factors are thought to have some relationship, e.g. Jones and Johnston (1999). It is, therefore, vital to state a priori the number of factors you expect to emerge and to have decided which rotation method you will use ahead of any analysis.

4.6 Pre-Analysis Checks
　　　　Ferguson and Cox (1993) give a detailed account of the process of exploratory factor analysis and provide a set of heuristics for its three stages of pre-analysis checks, extraction and rotation (see Table 3 for the pre-analysis checks). These pre-analysis checks are necessary to ensure the proposed data set is appropriate for the method. The checks include determining the stability of the emerging factor structure, sampling requirements, item scaling, skewness and kurtosis of variables and the appropriateness of the correlation matrix.

Table 3 Stages in questionnaire development: Factor Analysis

Questionnaire Development	Key Issues
Further Development: Exploratory Factor Analysis	Principal Components Analysis (PCA): Explores the inter-relationship of variables Provides a basis for the removal of redundant or unnecessary items PCA is used to identify the underlying domains or factors within a measure Prior to analysis, must propose an underlying theoretical structure Ensure that the data set is appropriate Must follow a predefined and systematic analytic sequence
Further Development: Confirmatory Factor Analysis	Allows the further testing of the construct validity of the measure

4.7 Factor Extraction
　　　　Two main methods are used to decide upon the number of emerging factors, **Kaiser's criterion** for those factors with an eigen value of >1 and the scree test. An eigenvalue is an estimate of variance explained by a factor in a data set (Ferguson & Cox 1993), and a value >1 indicates greater than average variance.
　　　　With PCA, the removal of redundant items within a developing measure occurs within an iterative process. Agius et al. (1996) describe an iterative process of removing variables with general loadings (of 0.40 on more than one factor) and weak loadings (failing to load above 0.39 on any factor). This process is applied to the initial unrotated PCA before applying a varimax or oblimin rotation to interpret the structure of the solution.

V. Conclusion

　　　　This paper emphasizes the need to adopt a logical, systematic and structured approach to questionnaire development. A framework for questionnaire design has been presented. The objective is to present a simple, logical structure for approaching the task of questionnaire design, which combines a general philosophy of questionnaire design with specific principles. A framework is also presented which supports questionnaire development process using item analysis and related methods to demonstrate the reliability and validity. The need to preplan each stage of the questionnaire development process is suggested.

References

[1]. Belson WA (1981). The Design and Understanding of Survey Questions. Aldershot, England: Gower.
[2]. Converse JM (1982). Patricia Labaw, Advanced Questionnaire Design, book review in Public Opinion Quarterly, 46(2): 294-295.
[3]. Converse JM & Presser S (1986). Survey Questions: Handcrafting the Standardised Questionnaire. Newbury Park, CA: Sage.
[4]. de Vaus DA (1991). Surveys in Social Research. (Third Edition). Sydney, Australia: Allen & Unwin.
[5]. Jenkins, CR & Dillman DA (1995). Towards a Theory of Self-administered Questionnaire Design. In Survey Measurement and Process Quality, edited by Lyberg, L; P Biemer; M Collins; E De Leeuw; C Dippo; N Schwarz and D Trewin. New York: Wiley-Interscience.
[6]. Labaw PJ (1980). Advanced Questionnaire Design. Cambridge, MA: Abt Books.
[7]. Sudman S & Bradburn NM (1983). Asking Questions: A Practical Guide to Questionnaire Design. San Francisco, CA: Jossey-Bass.
[8]. Sudman S; Bradburn NM & Schwarz N (1996). Thinking About Answers: The Application of Cognitive Processes to Survey Methodology. San Francisco, CA: Jossey-Bass.