

Self-Report Questionnaires

Constantina Demetriou,¹ Bilge Uzun Ozer,² and Cecilia A. Essau¹

¹University of Roehampton, U.K. and ²Cumhuriyet University, Turkey

Self-report questionnaires are frequently needed in clinical psychology to identify specific symptoms or signs of psychological disorders. They are also used for understanding the prevalence of the disorders. The general format of most self-report questionnaires is Likert-style responses to items in terms of subjective experience, or frequency of specific symptoms of psychological disorders or degree of impairment. Self-report instruments may not be sufficient for determining an individual's specific diagnosis, but are necessary to assess their experience. They can be completed in hand or online with or without the assessor's presence.

In order to compare responses across different diagnoses, questionnaires are frequently used to assess symptoms before and after treatment. Self-report instruments are also typically administered as part of a comprehensive assessment. The responses can be used to assist a clinician in the initial evaluation by providing a guide as to the probability of a particular problem and as a tool for quantifying the individual's presenting symptoms.

Advantages and Disadvantages

One of the main advantages of the self-report questionnaire is that it can be administered to a large sample of people quickly without much effort or financial cost. As self-report questionnaires enable the collection of a large amount of quantitative data, generalization of the findings is possible, especially when the sample is collected randomly. For this reason, self-report

questionnaires are often used as screening instruments in the first stage of a two-stage process in epidemiological studies. Furthermore, because the respondents themselves are much closer to the issues in question than other individuals, the information they give in self-report questionnaires tends to be more accurate. Others are limited to reporting only the obvious side of respondent's experience through their behavior and verbal responses.

Nevertheless, some of the limitations of using self-report questionnaires should be considered. The main disadvantage of self-report questionnaires might be the possibility of providing invalid answers. While responding to the items, respondents may not answer truthfully, especially on sensitive questions. This phenomenon is known as *social desirability bias*, in that they may respond in a socially acceptable way. There are also some issues that affect the validity and reliability of the questionnaires. One of the issues is a *response bias*, which is an individual's tendency to respond in a certain way regardless of the question. For example, individuals may be more likely to respond "yes" regardless of the content of the question—known as an acquiescent response bias—or more likely to respond "no"—a nonacquiescent bias. Another problem in using self-report questionnaires might be the clarity of the items, which brings the risk of obtaining different interpretations of questions. Moreover, in the case of the highly structured questionnaires, the structure may force participants to answer in a way that does not match their views. For example, the structure may reflect the preconceptions of the researcher. On the other hand, open-ended questions may lead to subjectivity that leads to more complex analysis. The presence of the researcher at the completion of the questionnaire may affect answers as well. For example, subjects may change their behavior or demonstrate an improvement in their outcome because

they know they are being observed. More specifically, in the presence of the examiner, a responder may not feel comfortable selecting the extreme choices. Another disadvantage can be the lack of flexibility, especially with fixed-choice questions. Asking the participants to rate a statement gives them limited ability to express themselves and their feelings.

Designing a Self-Report Questionnaire

Designing the questionnaire is a crucial procedure and several steps are needed. The first step involves deciding on the topic(s) to be covered. Self-report questionnaires are often divided into subsections representing different topics or constructs. In this respect, to cover the topic to be examined, the questionnaire should include items covering the range of the topic (sometimes referred to as content validity).

The second step involves constructing specific items to be included in the self-report questionnaire. All the items in a self-report questionnaire must be clear to understand. Self-report questionnaires can contain open and closed questions. Open questions invite the participants to respond in their own words, which provide qualitative data. On the other hand, closed questions, providing quantitative data, limit the participants to certain answer choices. Moreover, the self-report instruments may contain fixed-choice questions, where the respondent has to make a fixed-choice answer, usually yes or no. It is suggested to start with easy, simple, and clear questions written in everyday language. In a clinical setting where the researcher may want to obtain information in sensitive areas such as suicide and self-harm, the questions should be nonthreatening. Double-barreled questions that include more than one construct in a single question are problematic since they can confuse participants; for example, agreeing with the one part of a question but disagreeing with another (de Leeuw, Hox, & Dillman, 2008).

The third step in designing self-report questionnaires involves selecting a response scale. There is a broad variety of scale types, such as a Guttman scale, a Thurstone scale, and ranking. One of the most common rating scales is the *Likert scale*, in which the participants decide how strongly they agree or disagree with a statement. In constructing a response scale there are a number of points to consider. One is the number of scale points, which can range from two upwards. Scales with two options are known as dichotomous, whereas scales with three or more choices are known as multiple choice. Dichotomous scales can limit a person's ability to express uncertainty or varying degrees of a characteristic (e.g., a false choice between fully absent and fully present for a symptom). However, some participants may use in-between rating options to minimize their response (e.g., endorsing "some of the time" when the symptom is present "most of the time"); a dichotomous item forces the individual to state that something is or is not present.

Additionally, response scales can be either unipolar or bipolar. A unipolar scale has only one concept (e.g., "no pain" at the one end and "unbearable pain" at the other), whereas a bipolar scale has opposite descriptions at each end of the scale (e.g., from "agree" to "disagree"). There is a possibility that a bipolar scale may also have a midpoint. It allows respondents to legitimately express "no opinion" if they have none. However, a neutral midpoint may allow individuals to avoid expressing an opinion even when they have one. The last consideration in designing self-report questionnaire is *anchoring*, which refers to identifying the points of the scales in words and/or numbers. Explicitly defining the rating options increases the reliability of how respondents use the scale. In this respect, the researcher should always consider the measurement assumptions of the scale, including the scale of measurement (e.g., nominal, ordinal, interval, or ratio) and the specificity of the anchors.

After constructing the questions and the response format, the subsequent step is to list the questions in a coherent manner. The layout

of a questionnaire is also important as good layout encourages respondent engagement and minimizes possible irritation. In this respect, a questionnaire should have a good layout with simple instructions. Physical attractiveness of the questionnaire is another important aspect. Questionnaires should be designed to appear as short as possible, should not be overcrowded, and should have adequate space for open-ended responses. After completing all the aforementioned steps, a pilot study is necessary to ensure that the items used for the instrument are understandable. The pilot study allows the developer to make revisions if needed.

Reliability

Self-report questionnaires should be tested for reliability, which is the repeatability or dependability of the measurement. A questionnaire yielding the same results in repeated usage in similar circumstances can be accepted as reliable (Nunnally & Bernstein, 1994). There are several methods to understand the scale's reliability. *Internal consistency* refers to the homogeneity of items. The internal consistency of multi-item scales can be assessed by coefficient alpha, which is the mean of all split-half coefficients (Cronbach, 1951). For a reliable questionnaire, an internal consistency score of .7 is recommended as a minimum (Nunnally & Bernstein, 1994). Reliability can also be assessed by the *test-retest* method, in that the participants are asked to respond to the items of the instrument twice at different times. Test-retest reliability addresses consistency across time, which is obtained by calculating a correlation coefficient for the two assessment periods. As the test-retest coefficient is not sensitive to specific error, *equivalent/parallel forms reliability* can be a solution. It examines the reliability across different but parallel versions of the same instrument.

Validity

Validity is the extent to which an instrument measures what it is supposed to measure. It is a multifaceted concept determined by relations

with other variables (Silva, 1993). Validity for self-report questionnaires is established by correlating the scores with a similar instrument. There are four types of validity, termed criterion validity, content validity, construct validity, and face validity. *Face validity* refers to whether the items appear to measure the intended construct. For example, a measure of anxiety has face validity if it includes items asking how anxious the person feels. *Content validity* indicates the extent to which a questionnaire examines all the targeted aspects of a construct. If a depression scale, for example, examines only the emotional aspect of depression but not the behavioral, it may lack content validity. The third type, *criterion validity*, assesses how well the instrument correlates with an established criterion. In other words it is assessing validity by comparing how well a variable predicts an outcome based on other variables. Last, *construct validity* is the degree to which an instrument takes the hypothetical qualities it was designed to measure. It examines the extent of the pattern of relationship between measures of two constructs consistent with theoretical expectation.

There are several factors that might contribute to a lack of validity for the use of particular questionnaires. First of all, the respondent must be cognitively capable of answering the question. Brener, Billy, and Grady (2003) suggested four cognitive processes influencing the response of participants including comprehension, retrieval, decision making, and response generation. In this respect, participants first engage in a comprehension process that determines how the question is encoded in the person's memory. Then retrieval cues are used to search memory in the subsequent retrieval stage. Any retrieved information is evaluated during the decision-making stage and finally, if the retrieved information is deemed passable for the purposes of answering the question, then a response will ensue. However, if this information is judged inadequate, additional retrieval attempts will be made. Breakdowns at any of these stages could disrupt the person's ability

to answer a question. In addition, other such issues as reading ability, attention problems, memory problems, and an individual's clinical problems may influence a respondent's ability to recall their attitudes or their behaviors. Second, validity problems also arise from characteristics of the external, social environment instead of internal processing. For example, social desirability, desire of attention, and lack of confidentiality, anonymity, or privacy are some factors postulated to be influential. More specifically, social desirability has an impact on situational biases while both desire of attention and lack of confidentiality can lead to response bias (Brener et al., 2003). Third, the person may lack introspective ability, even if the person is trying to be honest and accurate. A person may not recognize a particular problem or quality, even though others would attest to its presence. So, even if the response is accurate, there may be a lack of validity. Finally, there are other practical reasons why questionnaires may not be valid. Participants may vary regarding their understanding or interpretation of particular items. Some people are extreme responders while others interpret and use the scales differently. There may be a problem when the questionnaire includes ordinal data due to the fact that these kinds of data provide information only about the order in which questions are ranked but not the distance between them.

Considering both reliability and validity, evidence from previous research has indicated that validity takes priority. In terms of research, establishment of validity of a self-report questionnaire is of primary importance, as it characterizes theoretical practicality, but it needs to be supported by recommended levels of consistency and appropriate external reliability.

The Clinical Use of Self-report Questionnaires

Use of self-report questionnaires plays a central role in the practice of clinical psychology. A plethora of self-report questionnaires exists

for use in clinical practice. Self-report questionnaires can be used to help practitioners monitor and evaluate treatment progress. Some questionnaires are used to assess specific diagnostic criteria, whereas others assess emotions and behaviors more broadly.

Selecting the appropriate psychological assessment instrument is important because the clinician should consider the extent to which the instrument can answer the referral question. Not every questionnaire is appropriate for every purpose. Some of the examples frequently used in clinical psychology are provided below. Certainly, practitioners' training, experience, personal preferences, and familiarity with the related literature are important factors to consider. It is also crucial for the clinician to understand the intended purpose of the instrument (e.g., was it intended to aid in diagnosis or just to track symptoms over time?), the composition of its standardization sample (and if that matches the clinician's intended population), and the adequacy of the reliability and validity of the test.

One use of self-report tests may be to generate treatment plans (Beutler, Clarkin, & Bongar, 2000). A psychological test provides useful information regarding the planning, implementation, and evaluation of the treatment process. For example, the second edition of the *Beck Depression Inventory* (BDI-II; Beck, Steer, & Brown, 1996) is one of the most widely used instruments to measure depression. It contains 21 questions loading on a two-factor approach to depression: the affective and the somatic components. For the second version of the BDI, Dozois, Dobson, and Ahnberg (1998) carried out validity and reliability studies by gender. They found the scale to be highly reliable, with a coefficient alpha of approximately .90, and valid for use in clinical settings. This scale is sensitive to changes in symptomatology over 1 week or 1 month, and therefore can be used monitor an individual's response to treatment. In addition, it offers some advantages over clinician-rated scales, as it may take less time, does not require trained personnel,

and its administration and scoring process is standardized.

Multicultural Adaptation

With the increase of multicultural and multinational research projects and the increased use of self-report questionnaires in clinical practice, the need to adapt questionnaires for use in different languages has grown. It is well known that if self-report questionnaires are to be used across different cultures, items must not only be translated into different languages, but must also be adapted culturally to maintain the content validity of the instrument (Beaton, Bombardier, Guillemin, & Bosi Ferraz, 2000). Therefore, cross-cultural adaptation necessitates a method to compare the original source and target versions of the questionnaire. Beaton and colleagues (2000) used the term *cross-cultural adaptation* to cover a process that includes both the language and cultural alteration issues in the process of preparing the new translated questionnaire.

The process of adapting questionnaires for use in another culture has six stages. The original form of the instrument is obtained in the first stage of the process. Two translators with different profiles and backgrounds (informed/uninformed) are asked to translate the questionnaire from the original language into the target language. The translations are compared in terms of wording choice and/or conceptual consistency. The second stage is the synthesis of the translations where the two translators and an observer discuss and conduct a synthesis of the results of the translations. This synthesis process is summarized in a written report. Back translation follows this stage. Working from the translated version of the questionnaire, a totally blind translator translates the questionnaire back into the original language. If possible, it is best to obtain two of these back translations. In the fourth stage, an expert committee, ideally including health professionals, methodologists, language professionals, and the translators, are asked to achieve

cross-cultural equivalence, to consolidate all the versions of the questionnaire, and to develop a prefinal version of the instrument for field testing. The fifth stage of the process includes testing the prefinal version. In this stage, 30–40 participants complete the instrument, and the participants are interviewed regarding the clarity of the items and rating scales. Based on the feedback received from the participants, the scale is revised and piloted. This process ensures that the adapted version of the questionnaire is suitable for the new culture for which it has been adapted. After the pilot study, the validity and reliability studies aforementioned are conducted. The final step includes the submission of the reports and forms to the developer of the instrument for appraisal of the adaptation process. This procedure has been found to be the best way to adapt self-report questionnaires for use in different countries and cultures (Beaton et al., 2000).

SEE ALSO: Beck Depression Inventory (BDI); Construct Validity; Cross-Cultural Issues in Assessment; Reliability

References

- Beaton, D. E., Bombardier, C., Guillemin, F., & Bosi Ferraz, M. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25*(24), 3186–3191.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.
- Beutler, L. E., Clarkin, J. F., & Bongar, B. (2000). *Guidelines for the systematic treatment of the depressed patient*. New York: Oxford University Press.
- Brener, N. D., Billy, J. O. G., & Grady, W. R. (2003). Assessment of factors affecting the validity of self-reported health risk behavior among adolescents: Evidence from the scientific literature. *Journal of Adolescent Health, 33*(6), 436–457.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

- de Leeuw, E., Hox, J., & Dillman, D. (Eds.). (2008). *International handbook of survey methodology*. New York: Taylor & Francis.
- Dozois, D. J. A., Dobson, K. S., & Ahnberg, J. L. (1998). A psychometric evaluation of the Beck Depression Inventory-II. *Psychological Assessment, 10*, 83–89.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Silva, F. (1993). *Psychometric foundations and behavioural assessment*. Thousands Oaks, CA: Sage.

Further Reading

- de Leeuw, E. D., Hox, J., & Dillman, D. (2008b). *International handbook of survey methodology*. New York: Taylor & Francis.