

Review of the Steps for Development of Quantitative Research Tools

Ashok Kumar*

College of Nursing, All India Institute of Medical Sciences, Jodhpur Rajasthan, India

Abstract

The number of tools developed by researchers has increased in recent years. Still the demand and need for development of new and standardized tool is increasing to a great extent, the simple reason is either standardized tools are not available or available tools lacks reliability and validity in that setting. This article explores the steps and process which provides base by which a reliable and valid tool can be developed. Method followed was in-depth review of published research articles. This article reviews tool development procedures used in 17 articles published in leading journals from 1992 to 2007. It points out the steps of tool development viz. item generation in which theoretical basis to be used. Next step is reliability in that internal consistency, equivalence by inter-rater and stability by test-retest method is checked; validity i.e. face, content, concurrent and construct validity to be calculated. Further in construct validity factor analysis including principal component analysis, pre-analysis checks and factor extraction is to be analysed. Based on the review, the author mentioned the steps of tool development which will guide researchers to improve the tool development process.

Keywords: Tool development; Item generation; Reliability; Validity

Introduction

There are a range of scales and response styles that may be used when developing a tool. These produce different types or levels of data 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 [1].

In developing the evidence base of practice using this method of data collection, it is vital that tool design incorporates pre planned methods to establish reliability and validity. Failure to develop a tool sufficiently may lead to difficulty interpreting results, and this may impact upon clinical or educational practice. Hinkin and DeVellis state that the construction of a new tool is a highly complex process. Several steps are required to develop a multi-item tool to measure a construct. The researcher needs to:

- apply a theoretical basis to develop the items;
- design the individual items;
- conduct an item analysis to eliminate poor items (ambiguous, no variation);
- assess the reliability of tool i.e. internal consistency, stability and equivalence;
- determine the construct validity of the measure using factor analysis;
- determine the convergent validity of the measure; and
- determine the divergent validity (discriminant validity, including method effects) [2,3].

The researcher also needs to determine the level of specificity required of the scale. This will largely be determined by the research question, as the level of specificity of the scale should align with the level of specificity of the research question and the other constructs it will be compared with [3].

Item Generation: Use a Theoretical Basis

As a first step to tool development, the researcher should carefully examine the extant theory relating to the construct he or she wishes

to measure. Theory can provide a guide in terms of developing the conceptual formulations required for operationalization. Examining theory helps to establish the parameters of the construct to ensure that the content of the scale is focused on the actual domain of interest, rather than unrelated areas [3].

The researcher should develop the items from a theory of the construct (latent variable) so that they are consistent with it. If there is a theoretical basis for this construct, it can be defined and the type of relationships it has with other constructs can be predicted. The generation of items during tool development requires considerable pilot work to refine wording and content. In addition, a key strategy in item generation is to revisit the research questions frequently and to ensure that items reflect these and remain relevant. During this stage that the proposed subscales of a tool are identified and to ensure that items are representative of these. The item and factor analysis stages of the tool development process may then be used to establish if such items are indeed representative of the expected subscale or factor [4,5].

Once the researcher has generated the initial pool of items, the next step involves having a panel of subject-matter expert's review the items in terms of content adequacy. These experts should be provided with construct definitions and instructed to sort the items according to these definitions, to determine whether their sort aligns with the scale developer's conceptualisations [6].

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. The type of tool, language used and order of items may all bias response. Consideration should be given to the order in which items are

*Corresponding author: Ashok Kumar, Assistant Professor, College of Nursing, All India Institute of Medical Sciences, Jodhpur, Rajasthan, India, Tel: 8003996929; E-mail: ashokbishnoi1@gmail.com

Received October 31, 2015; Accepted December 21, 2015; Published December 30, 2015

Citation: Kumar A (2015) Review of the Steps for Development of Quantitative Research Tools. Adv Practice Nurs 1:103. doi: [10.4172/2573-0347.1000103](https://doi.org/10.4172/2573-0347.1000103)

Copyright: © 2015 Kumar A. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

presented, e.g. it is best to avoid presenting controversial or emotive items at the beginning of the tool. To engage participants and prevent boredom, demographic and/or clinical data may be presented at the end. Certain questions should be avoided, e.g. those that leads or includes double negatives or double-barrelled questions [5]. 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 items [7].

To allow respondents to expand upon answers and provide more in-depth responses, free text response or open statements/ questions may be included. Respondents may welcome this opportunity. However, whilst this approach can provide the inter-viewer with rich data, such material can be difficult to analyse and interpret. However, these problems may be outweighed by the benefits of including this option and can be especially useful in the early development of a tool. Free text comments can inform future tool development by identifying poorly constructed items or new items for future inclusion [8].

Reliability

Reliability is an essential issue in scale development and refers to the amount of variance attributable to the true score of the latent construct. Reliability refers to the repeatability, stability or consistency of a tool. One form of reliability, internal consistency, is determined by calculating coefficient alpha/ Cronbach’s alpha statistic. This statistic uses inter-item correlations to determine whether constituent items are measuring the same domain. Internal consistency refers to the homogeneity of items within a scale [9]. This coefficient should be as high as possible. If not, items contributing to low reliability (low item to total correlations) need to be dropped and new items developed. If the items show good internal consistency, Cronbach’s alfa should exceed 0.70 for a developing tool or 0.80 for a more established tool. It is usual to report the Cronbach’s alpha statistic for the separate domains within a tool rather for the entire tool. Reliability is a necessary pre-condition for validity [7] (Table 1).

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 tool or domain [4]. This score can be biased, especially in small sample sizes,

as the item itself is included in the total score. 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 tool or domain prior to the correlation. Kline recommends deleting any tool item with a corrected item-total correlation of <0.30. Item analysis using inter-item correlations will also identify those items that are too similar. High inter-item correlations (>0.80) suggest that these are indeed repetitions of each other (sometimes referred to as bloated specifics) and are in essence asking the same question [10].

Test-retest reliability can assess stability of a measure over time and this should be included in the process of any tool development. This is of particular importance if the intended use of the measure is to assess change over time or responsiveness [11].

Equivalence in the context of reliability assessment primarily concerns the degree to which two or more independent observers or coders agree about scoring. For this the inter-rater reliability to be calculated by Cohen’s kappa and values of 0.75 or higher are considered very good [12].

Validity

Validity refers to whether a tool is measuring what it purports to. While this can be difficult to establish, demonstrating the validity of a developing measure is vital [7]. There are several different types of validity. Content validity (or face validity) refers to expert opinion concerning whether the scale items represent the proposed domains or concepts the tool is intended to measure. This is an initial step in establishing validity, but is not sufficient by itself. Convergent (or concurrent) and discriminant validity must also demonstrated by correlating the measure with related and/or dissimilar measures [5].

The researcher would obtain measures of the scale from a sample from whom he or she also obtained measures of constructs the scale should be related to, including alternative measures of the construct of interest (convergent validity), and of constructs the scale should not be related to (divergent validity). These relationships are calculated by correlation coefficients (e.g. a Pearson product moment correlation coefficient). When developing a tool it is, therefore, important to include, within the research design, additional established measures with proven validity against which to test the developing tool. Construct

Tool development	Key issues	Decision aids
Piloting the tool	Spread of responses across options	High endorsement of a single option is problematic.
Item analysis	Initial analysis	An item should be considered for removal if ≥80%, ≤20% of responses endorsed one response.
	Clarity and relevance of items	Items with an inter-item correlation of <0.3 or >0.7 should be considered for removal.
	Items deemed theoretically important Is your measure affected by social desirability bias?	Items with a poor Cronbach’s a, i.e. <0.7 should be considered for removal
		Researcher’s interpretation of patient comments. Alternatively, if respondents fail to complete an item it suggests that the item may lack clarity. Items should be retained if they are deemed to be theoretically important even if they do not meet the above criteria. Explore the relationship between item and scale total with measure that captures this response tendency, e.g. Marlowe-Crown Social Desirability Index.
Reliability	Internal consistency	Corrected inter-item correlations
	Test-retest	Item-total correlation
	Inter-observer	Cronbach alpha
		Coefficient correlations
Validity	Face or content	Do the items sufficiently represent different hypothesized domains?
	Concurrent or discriminant	Do subscale scores correlate with existing, validated measures presented concurrently?
	Predictive	Do subscale scores predict hypothesis reports on existing, validated measures presented longitudinally?

Table 1: Stages in tool development: (Rattray et al.) [1].

validity relates to how well the items in the tool 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 [1].

Factor Analysis

The purpose of an exploratory factor analysis is to analyse scores on several items to see if they can be reduced to underlying dimensions. Those items that are highly related to each other will load on one factor. The items that are measuring one construct should load on one factor and those measuring another construct should load on a different factor. Analyses that yield no clear factors or one factor (for a unidimensional scale) are problematic. Additionally, the factor analysis should explain a substantial amount of the variance in the scores. Based on these factor loadings, the researcher needs to decide which items from the scale should be retained or deleted [1,3].

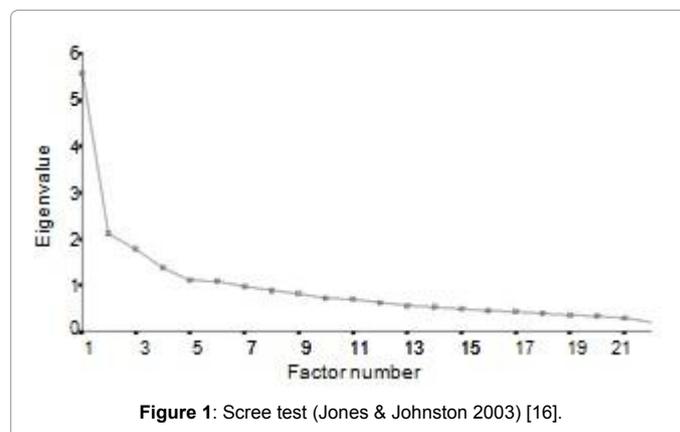
Following initial pilot work and item deletion, the tool should be administered to a sample of sufficient size to allow exploratory factor analytic techniques to be performed. Ferguson and Cox suggest that 100 respondents is the absolute minimum number to be able to undertake this analysis. However, others would suggest that this is insufficient and a rule of thumb would be five respondents per item [13].

Principal components analysis (PCA) explores the inter-relationship of variables. It provides a basis for the removal of redundant or unnecessary items in a developing measure and can identify the associated underlying concepts, domains or subscales of a questionnaire [14]. 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 tool 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 [15].

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, and to assist the interpretation of a factor solution, rotation of factors (components) is often performed. Factor rotation maximizes the loadings of variables with a strong association with a factor, and minimizes those with a weaker one 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 [13]. However, oblimin rotation may be used, when factors are thought to have some relationship. 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 [10].

Pre-analysis Checks

Ferguson and Cox give a detailed account of the process of exploratory factor analysis and provide a set of heuristics for its stages



Tool development	Key issues	Pre-analysis checks (Ferguson and Cox) [13]
Further development:	Principal components analysis (PCA):	Stable Factor Structure
Exploratory Factor analysis	Explores the inter-relationship of variables	Minimum number of participants: 100
	Provides a basis for the removal of redundant or unnecessary items, PCA is used to identify the underlying domains or factors within a measure.	Minimum participant to variable ratio, N/p: 2:1-10:1
		Minimum variable to factor ratio, p/m: 2:1-6:1
	Prior to analysis, must propose an underlying theoretical structure.	Minimum participant to factor ratio, N/m: 2:1-6:1
	Ensure that the data set is appropriate	Sampling
	Must follow a predefined and systematic analytic Sequence	Random sampling from a population.
		Item scaling
		Likert, Mokken and frequency scales are acceptable.
		Normality of distribution/skewness and Kurtosis
		Underlying assumption is of normal distribution. Values of skewness and kurtosis should be calculated for each variable, and values out with accepted levels dealt with appropriately.
		Appropriateness of the correlation matrix Kaiser Meyer-Olkin: can the correlations between variables be accounted for by a smaller set of factors? should be >0.5.
		Bartlett Test of Sphericity: based on the chi-squared test, - a large and significant test used to indicate discoverable relationships.
Further development:	Allows the further testing of the construct	Confirmation of factor structure on an independent data set, using exploratory and confirmatory methods.
Confirmatory factor analysis	Validity of the measure	Same underlying assumptions as exploratory methods.
		Confirmatory process uses single sample and multi-sample approaches.

Table 2: Stages in tool development: factor analysis.

of pre-analysis checks, extraction and rotation (Table 2 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 [13].

Factor Extraction

Two main methods are used to decide upon the number of emerging factors, Kaiser's criterion for those factors with an eigenvalue of >1 and the screen test. An eigenvalue is an estimate of variance explained by a factor in a data set, and a value >1 indicates greater than average variance [13]. A scree test is the graphic representation of this. Figure 1 shows the screen test that demonstrated the four-factor structure from a tool. The number of factors is identified from the break in the slope. If a straight line is fitted along the eigenvalue rubble, the number of domains within the questionnaire is revealed by the number of factors above the line. This latter method includes a degree of subjectivity in its interpretation [16].

With Principal components analysis (PCA), the removal of redundant items within a developing measure occurs within an iterative process. Agius et al. 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 [17].

Conclusion

This article emphasizes the need to adopt a logical, systematic and structured approach to tool development. Author presented a framework that supports this type of approach and have illustrated the tool development process using item analysis, factor analytic and related methods and have demonstrated strategies to determine the reliability and validity of the new and developing measure. Here with suggested the need to preplan each stage of the tool development process and provide a series of heuristic strategies to enable the researcher to achieve this.

References

1. Rattray J, Jones MC (2007) Essential elements of questionnaire design and development, *Journal of Clinical Nursing* 16: 234-243.
2. Hinkin TR (1995) A review of scale development practices in the study of organizations. *Journal of Management* 21: 967-988.
3. DeVellis RF (2003) *Scale development: Theory and application*, Thousand Oaks, CA: Sage Publications (2nd edn).
4. Priest J, McColl BA, Thomas L, Bond S (1995) Developing and refining a new measurement tool. *Nurse Researcher* 2: 69-81.
5. Bowling A (1997) *Research Methods in Health*. Open University Press, Buckingham.
6. Schriesheim, CA, Eisenbach RJ (1995) An exploratory and confirmatory factor-analytic investigation of item wording effects on the obtained factor structures of survey questionnaire methods. *Journal of Management* 21: 1177-1193.
7. Bryman A, Cramer D (1997) *Quantitative Data Analysis with SPSS for Windows*. Routledge, London.
8. Polgar S, Thomas S (1995) *Introduction to Research in the Health Sciences*. Churchill Livingstone, Melbourne.
9. Jack B, Clarke AM (1998) The purpose and use of questionnaires in research. *Professional Nurse* 14: 176-179.
10. Kline P (1993) *The Handbook of Psychological Testing*. Routledge, London.
11. Ferketich S (1991) Focus on psychometrics: aspects of item analysis. *Research in Nursing and Health* 14: 165-168.
12. Pils K, Neumann F, Meisner W, Schano W, Vavrovsky G, et al. (2003) Predictors of falls in elderly people during rehabilitation after hip fracture - who is at risk of a second one? 36: 16-22.
13. Ferguson E, Cox T (1993) Exploratory factor analysis: a user's guide. *International Journal of Selection and Assessment* 1: 84-94.
14. Oppenheim AN (1992) *Questionnaire Design, Interviewing and Attitude Measurement*. Pinter, London.
15. Anthony D (1999) *Understanding Advanced Statistics: A Guide for Nurses and Health Care Researchers*. Churchill Livingstone, Edinburgh.
16. Jones MC, Johnston DW (2003) Further Development of the SNSI. Paper presented at the Royal College of Nursing Annual International Research Conference, University of Manchester.
17. Agius RM, Blenkin H, Deary IJ, Zealley HE, Wood RA (1996) Survey of perceived stress and work demands of consultant doctors. *Occupational and Environmental Medicine* 53: 217-224.