On Meaning and Measurement: A Review of Content Validity in IS

Emergent Research Forum Paper

Jackie London
Clemson University
jlondon@clemson.edu

Kevin D Matthews
Clemson University
kmatth@clemson.edu

Varun Grover
Clemson University
vgrover@clemson.edu

Abstract

Content validity is essential for developing accurate and consistent psychometric measures, which are necessary for the progress of theory and thus the health of an academic field and yet the field is lacking a systematic analysis of its engagement with these practices. In this research we turn to two of the field's top journal to provide a view of the best practices with regard to establishing content validity. Specifically, we assess the extent to which the content domain of a construct is established and defined, and the extent to which instruments are developed and judged according to the relevance and representativeness of the items. Additionally, develop a framework for systematically and quantitatively assessing content validity.

Keywords

Content validity, methodologies, introspective.

Introduction

Content validity is essential to the development of accurate and consistent psychometric measures. In academic research, the accumulation of knowledge is slowed or tainted when a researcher's instrument lacks validity and is thus unable to accurately assess the focal phenomenon. Though there are many ways in which an instrument may err, content validity is uniquely pernicious because, as a qualitative measure, it is immune to many of the statistical safeguards that alert researchers to anomalies in their results. Content validity is a judgment of the degree to which an instrument measures what its authors claim it measures, and as such, is entirely dependent on a researcher's ability to fully and unambiguously represent the content domain of a phenomenon within a set of questions, tasks or prompts. When this translation process is faulty, a gap forms between what an instrument measures and what the researcher intends to measure, and the consequences may be severe.

Scholars have expressed concern that the field has not paid enough attention to the process of accurately and precisely conceptualizing phenomena. Straub (1989) found that 4% of studies employ practices to ensure the content validity of their instruments. In a follow-up study, Boudreau et al. (2001) found an alarmingly high proportion (75%) of new measurement instruments were exempted from any type of content validation. Later, Petter et al. (2007) attributed a high rate of model misspecification (70%) to poor conceptualization of constructs causing an incomplete understanding of the phenomenon likely omitting important aspects of the content domain. Despite an increasing appreciation for the value of content validity, MacKenzie et al. (2011) were surprised to find "that so many researchers either neglect this step in the process or fail to properly implement it."

A continued disregard of the content validation process weakens the foundations of IS research and lessens the value of IS research. Because positivist research begins with the conceptualization of abstract phenomena, any missteps propagate and are amplified throughout the research project. MacKenzie (2003) identified three consequences of weak conceptualization: contaminated measures; model misspecification and weak rationale for hypotheses. These consequences then infect the statistical analyses leading to low construct, statistical conclusion and internal validity. Straub is stark in his
On Meaning and Measurement

Despite a wealth of available resources for guiding instrument development efforts and strengthening the content validity of measures (Lewis et al. 2005; MacKenzie et al. 2011; Malhotra and Grover 1998) the field is lacking a systematic analysis of its engagement with these practices. In this research, we investigate the content validity practices employed by empirical studies published in two of the field’s top journals, and quantify the extent to which new constructs are thoroughly conceptualized and their measurement instruments are relevant and representative of the specified domain. Two research questions guide our exploration of this issue: 1) What practices do researchers employ to ensure the content validity of their instruments? 2) Are we as a field successful in developing accurate instruments for measuring local phenomena?

Literature Review of Measurement Development

Social and behavioral scientists often find themselves explaining changes in perceptions, attitudes, intentions and motivation and thus do not have the benefit of precision in either the definition or the measurement of phenomena. These concepts are abstractions that cannot be weighed, prodded or measured, only queried. This means that though the researcher may think they have been clear in their asking and the respondent may think they understand what is being asked, the error term is ever-present in the interaction. Thus, scientific inquiry in the management sciences has an inherent fuzziness that hinges on the conceptualization and operationalization of abstract phenomena--constructs.

Constructs are so called because they are linguistic constructions of abstract concepts, which are "designed for a special scientific purpose, generally to organize knowledge and direct research in an attempt to describe or explain some aspect of nature" (Peter 1981, p. 134). As such, constructs are a subjective and yet essential component of theory and theorizing (Bacharach 1989). Over time many scholars have suggested various methodologies for operationalizing constructs (Churchill 1979; Hinkin 1998; Lewis et al. 2005; MacKenzie et al. 2011; Malhotra and Grover 1998). The focus of these guides is to limit the subjective nature—or increase the validity—of both the constructs that structure management theory and the instruments that measure them. Common among many procedures is a series of steps that proceed in three phases: conceptualization, instrument development, and instrument assessment.

Conceptualization is the process by which abstract phenomena are concretized. Through conceptualization the construct is described and defined in a way that both delineates the essential aspects of the concept as well as distinguishes the focal construct from similar or related concepts. The result of conceptualization will provide the basis for instrument development. During the instrument development phase, researchers generate several items or prompts which, together, fully and accurately probe the construct’s content domain. Once the items are created, a pretest should be completed where a panel of subject experts and researchers should review the resultant collection of items to assess the quality and appropriateness of each item and to suggest the exclusion of weak items or the inclusion of additional items. The full instrument is then presented to a pilot sample of potential subjects for data collection and additional comment. The data collected during the pilot test will provide statistical support for the psychometric properties of the instrument and may offer direction for further refinement through the identification of redundant or problem items. As a whole, the construct validation process is intended to increase the “epistemic correlation” between a construct and its measure such that variations in measurement are closely related to variations in the construct itself (Peter 1981).

In psychological assessment, measurement instruments are evaluated based on their validity and reliability. Reliability refers to the consistency with which variations in the measure correspond to variations in the construct. Validity is generally understood as an assessment of the accuracy with which the instrument measures the construct, and is broadly defined as “the extent to which an operationalization measures the concept it is supposed to measure” (Bagozzi et al. 1991, p. 421). Validity is an assessment which contains three distinct sub-evaluations of psychometric quality: content, construct and criterion (Cronbach and Meehl 1955). Content validity is a qualitative assessment of the construct’s conceptualization and operationalization. Construct and criterion validity are quantitative assessments of the measurement instrument with construct validity assessing the extent to which the instrument measures a structurally-consistent, unitary and distinct construct, and criterion validity assessing the
extent to which the focal construct correlates with antecedents, consequents and related constructs in a manner that is consistent with theory. Each facet of validity is further subdivided into additional sub-validities, and poor validity in any dimension will threaten the instrument’s overall validity (Lewis et al. 2005). Thus, researchers must ensure the face, content, nomological, concurrent, predicative, factorial, convergent and discriminant validity of their instruments (see Table 1).

<table>
<thead>
<tr>
<th>Facet</th>
<th>Facet</th>
<th>Definition</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>Face</td>
<td>Degree to which a member of the sample frame believe an instrument is appropriate for the measure of a given concept.</td>
<td>Qualitative</td>
</tr>
<tr>
<td>Content</td>
<td>Content</td>
<td>Degree to which the content domain of the construct is fully and accurately identified and adequately sampled by the test.</td>
<td></td>
</tr>
<tr>
<td>Criterion</td>
<td>Concurrent</td>
<td>Degree to which scores for the focal construct are correlated with scores for another variable.</td>
<td></td>
</tr>
<tr>
<td>Criterion</td>
<td>Predictive</td>
<td>Degree to which the instrument is capable of predicting the outcome of another variable.</td>
<td></td>
</tr>
<tr>
<td>Criterion</td>
<td>Nomological</td>
<td>Degree to which a measure is able to predict the values of other variables in a network of related constructs.</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Construct</td>
<td>Convergent</td>
<td>Degree to which multiple items purporting to measure the same construct are in agreement with one another.</td>
<td></td>
</tr>
<tr>
<td>Construct</td>
<td>Discriminant</td>
<td>Degree to which measures of different concepts are distinct.</td>
<td></td>
</tr>
<tr>
<td>Construct</td>
<td>Factorial</td>
<td>Degree to which the factor analytic structure of the items is consistent with the construct’s hypothesized structure.</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Validities in Psychological Research

**Content Validity**

Content validity is an assessment of a scholar’s ability to concretize the abstract for the purpose of measurement. Though many definitions of content validity have been offered (Cronbach and Meehl 1955; Haynes et al. 1995; Lawshe 1975; Lewis et al. 2005; Lynn 1986), all revolve around issues related to the selection of items from a specified content domain. Therefore, we embrace a version of Lawshe’s (1975) definition and define content validity as the degree to which the content domain of the construct is fully and accurately identified and the test adequately samples that universe. First the researcher must take care to fully and accurately conceptualize the proposed construct. This will involve the elucidation of both a content domain and a conceptual definition of the construct. These resources will then provide guidance for the selection of test items to measure the construct. To the extent that the researcher successfully performs these actions, the resultant instrument will possess the characteristics of a content valid measure: a defined domain, domain-relevant items, and items representative of the domain (Sireci 1998).

Within the instrument development framework, content validation is a sub-process which should follow a series of iterative and cumulative steps. First, researchers must conduct a review of prior academic research and an investigation of the phenomenon “in the wild.” The literature review will ensure that prior measurement efforts are incorporated into the new measure. Also, the prior research will help delineate the new construct from related concepts and help position the construct within a nomological network of related constructs. As an applied field, many IS constructs have a practice-oriented focus. In that case, researchers should additionally turn to practitioners and experts to fill in any epistemic gaps. The purpose of this initial step is to establish the phenomenon’s conceptual boundaries and structure (e.g. dimensions) which will allow for the building of a construct definition that efficiently and precisely communicates the researcher’s understanding of both the content and the boundaries of phenomenon (Suddaby 2010).

Once the conceptual boundaries of the phenomenon have been set, researchers should begin to generate items that are relevant and representative of the construct’s conceptual domain. This process, called domain sampling, is motivated by the assumption that though an infinite number of items would be needed to fully cover the construct’s domain, the true measure of the construct can be estimated through a proportionate sampling of each facet of the domain. Researchers have multiple avenues to aid in the generation of items including deduction from the conceptual definition, literature search, prior measures, focus groups, practitioners, and colleagues. The focus should be on covering the content domain and generating as many items as possible irrespective of quality. Once an initial set of items is finalized, the
questions should be refined and purified through sorting and pretesting procedures. A Q-sort is useful when developing measures for constructs with multiple dimensions or that are conceptually similar to existing constructs because it adds rigor to survey development and ensures that items are conceptually linked to the proposed construct (Hinkin 1998). Pretests assess the efficacy of the instrument by submitting it to a small panel of subject matter experts for review (Hunt et al. 1982). Experts provide feedback on the instrument, item relevance, representativeness and clarity, and domain coverage.

**Methodology**

To provide a systematic review of the IS community’s use of content validity practices, a literature search was performed to identify empirical studies that develop a new instrument to measure a psychological construct. The search was limited to articles published between the years of 2001 and 2015 in two of the field’s top journals: *MIS Quarterly (MISQ)* and *Information Systems Research (ISR)*. This 15 year span has given the field’s researchers the opportunity to react to and strengthen our research methods with respect to the various articles that have called for greater attention to content validity (Boudreau et al. 2001; Lewis et al. 2005; MacKenzie et al. 2011; Straub 1989). Each article journal was examined by looking at the abstract, methodology and appendicies. If the article stated that a new measure was developed, it was flagged for inclusion. In the end, 174 articles were identified that fulfilled the above criteria. While the articles included do not represent all research completed in the IS field, we feel the sample is representative of “the best” construct development practices within our community.

Two independent coders identified and documented construct development details from each article. First, a verification of inclusion or exclusion was made. Some articles were initially identified as creating new measures based on the phrasing the article author(s) made but, on closer examination, no new psychometric measure was created. These articles were excluded from the study. Of the 174 articles, 20 were excluded leaving 154 articles. Next, a newly developed construct from each article was identified. Since the unit of analysis for the current study is one article and not one construct, and to prevent the dilution of the sample by including several constructs from the same article, only one newly developed construct was selected from each article. In this case of multiple new scales, the first construct listed in the appendix or table of measures was selected. If no such listing existed, the first listed in the correlation matrix was selected. Absent a correlation matrix, the first construct mentioned in the methods or measures section of the article was selected. After collecting data on each construct (e.g. construct name, items), two additional analyses were conducted to assess the content validity practices.

**Step 1: Development**

Once the focal construct of the article was identified, the coders collected data related to that construct only. The paper was read to determine if (1) a literature review had been conducted for the construct, (2) a definition for the construct was given, (3) if the item generation process used by the author had been detailed within the paper, and (4) if a pre-test was conducted before using the new measure. Each variable was coded as a binary value. In addition, if provided, the author’s definition of the focal construct and the items generated for the measure were recorded verbatim from the article.

**Step 2: Assessment**

To assess the content validity of the final items, we plan to develop a content validity instrument (Lawshe 1975; Lynn 1986; MacKenzie et al. 2011) to determine (1) the construct definition quality, (2) the definition clarity, (3) each item’s adequacy, (4) each item’s representativeness, (5) each item’s clarity, (6) the instrument’s comprehensiveness, and (7) the instrument’s bias. Each construct will be assessed by a panel of raters who are knowledgeable of issues related to the conceptualization of constructs and the development of instruments. The results will be used to calculate a content validity index for each construct. At the time of submission, this step has not begun.

**Expected Results and Contributions**

The goal of this research is to provide a systematic analysis of the IS communities engagement with the practices identified as best practices for ensuring the content validity of measures. Based on the problems
hinted at by prior work (Boudreau et al. 2001; MacKenzie et al. 2011; Petter et al. 2007; Straub 1989) and the preliminary data collected, it is expected that this study will reveal that the field is indeed inadequately addressing content validity. However, it is our expectation that these problems are not systemic and, as a result, hope to identify several exemplar efforts which will provide a framework for establishing future best practices for the field. Additionally, we hope to provide a more systematic and quantitative method for ensuring the content validity of instruments.

Content validity is a primary support in the bridge from constructs to conclusions. Poor content validity sets subsequent researchers on weak footing from which to extend and expand the stream of research. Though concerns over weak content validity and guidance for improving content validity are well known in IS, the process by which validity is established is often absent from published research. This research has set out to investigate the current state of the content validity of new measures created within our field and illuminate the specific practices that relate to the establishment of content validity.

References and Citations