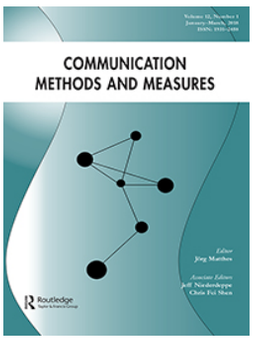

Routledge
Taylor & Francis Group

Communication Methods and Measures

ISSN: 1931-2458 (Print) 1931-2466 (Online) Journal homepage: http://www.tandfonline.com/loi/hcms20

Ten Steps in Scale Development and Reporting: A Guide for Researchers

Serena Carpenter

To cite this article: Serena Carpenter (2018) Ten Steps in Scale Development and Reporting: A Guide for Researchers, Communication Methods and Measures, 12:1, 25-44, DOI: 10.1080/19312458.2017.1396583

To link to this article: https://doi.org/10.1080/19312458.2017.1396583

COMMUNICATION METHODS AND MEASURES
2018, VOL. 12, NO. 1, 25–44
https://doi.org/10.1080/19312458.2017.1396583

Routledge
Taylor & Francis Group

Ten Steps in Scale Development and Reporting: A Guide for Researchers

Serena Carpenter

School of Journalism, Michigan State University, East Lansing, MI, USA

ABSTRACT

Scale development involves numerous theoretical, methodological, and statistical competencies. Despite the central role that scales play in our predictions, scholars often apply measurement building procedures that are inconsistent with best practices. The defaults in statistical programs, inadequate training, and numerous evaluation points can lead to improper practices. Based on a quantitative content analysis of communication journal articles, scholars have improved very little in the communication of their scale development decisions and practices. To address these reoccurring issues, this article breaks down and recommends 10 steps to follow in the scale development process for researchers unfamiliar with the process. Furthermore, the present research makes a unique contribution by overviewing procedures scholars should employ to develop their dimensions and corresponding items. The overarching objective is to encourage the adoption of scale development best practices that yield stronger concepts, and in the long run, a more stable foundation of knowledge.

Social scientific terms such as interactivity, source expertise, and media credibility organize our thinking about research. Theoreticians construct abstract measures based on the collective scientific community's interpretation of a latent term. The hallmark of the quantitative approach is that concepts have been grounded in systematic procedures that enable scholars to apply their measures in similar or varying settings to determine their usefulness. Usefulness of a particular concept is often determined by the concept's ability to predict phenomena and make claims of scientific knowledge, but that knowledge may be imprecise if scholars are not aware of proper scale development techniques and reporting procedures. The linking of measurement indicators to a concept is a complex process. Unfortunately, literature on measurement theory and practice to guide communication scholars is not emphasized.

The purpose of the present article is to be constructive by informing readers about the misuses of Exploratory Factor Analysis (EFA) in order to discourage such mistakes in the future. A lack of awareness regarding the appropriate procedures has prompted leading scale methodologists to argue that most measures are seriously flawed even in reputable journals (Conway & Huffcutt, 2003; Kline, 2013; McCrosky & Young, 1979). Furthermore, three separate content analyses of communication journal articles published from 1978–2009 found that statistical and methodological decisions associated with scale development were poor, which provides evidence concerning the existence of questionable measures in the field of communication (Morrison, 2009; Park, Dailey, & Lemus, 2002; Wimmer & Haynes, 1978).

Scholars learn formal measurement development expectations and our dedication to measurement quality through the observation of what scientists do. Scale development is not often taught in doctoral programs, which likely means that scholars learn by imitating procedures communicated in

CONTACT Serena Carpenter carp@msu.edu School of Journalism, Michigan State University, Room 344, 404 Wilson Road, East Lansing, MI 48824.

research journals (Conway & Huffcutt, 2003; Ford, MacCullum, & Tait, 1986; Park et al., 2002). This article examined the practices of scholars that applied procedures to develop their scales in top communication journals ranked by Thomson Reuter representing a ten-year period from 2005–2015 to determine whether the findings from previous content analyses hold in a more recent data set. It did not evaluate the work of scholars who created scales without the use of factor analysis procedures. The present study also extended previous content analysis work through the examination of several other variables essential to the measurement building process (e.g., the scale item generation and assessment procedures, sample characteristics, appropriateness of FA, item deletion process, factor cut-off levels, etc.). Additionally, communication and media researchers' concept explication choices concerning how they built their scales decisions *prior* to the application of statistics and methods were reviewed. Based on the present (n = 600) and previous content analysis results that continue to demonstrate that authors do not abide by scale development best practices, this article enacts a narrative intended to support researchers by breaking down the scale development process into ten manageable steps (see Table 1).

Scale development concepts

Scales try to capture not directly observable latent concepts with a group of concrete statements. Scales are "collections of items combined into a composite score intended to reveal levels of

Table 1. 10 steps in scale development and reporting.

1. Research the intended meaning and breadth of the theoretical concept
    a. Select appropriate conceptual labels
    b. Select conceptual definitions
    c. Identify potential dimensions and items
    d. Conduct qualitative research to generate dimensions and items
    i. Use feedback to refine scale
    i. Expert feedback, pre-tests, cognitive interviews, or pilot tests can be employed to evaluate item wording, item validity, questionnaire design, and model structure
2. Determine sampling procedure (5:1)
3. Examine data quality
4. Verify the factorability of the data
    a. Bartlett's Test of Sphericity (≤.05)
    b. Kaiser-Meyer-Olkin test of sampling adequacy (≥.60)
    c. Inspect correlation matrix (≥.30)
5. Conduct Common Factor Analysis
6. Select factor extraction method
    a. Principal Factors Analysis
    b. Maximum Likelihood
7. Determine number of factors
    a. Theoretical convergence and parsimony
    b. Scree test
    c. Parallel Analysis (PA)
    d. Minimum Average Partials (MAP)
8. Rotate factors
    a. Oblique rotation (Direct Oblimin, Promax)
9. Evaluate items based on a priori criteria
    a. Theoretical convergence
    b. Parsimony
    c. Weak loadings (≥.32)
    d. Cross loadings
    e. Inter-item correlations
    f. At least three-item factors
    g. Communalities of items (≥.40)
10. Present results
    a. Scale and subscale naming logic, conceptual definitions, sample size logic, methods for determining factor numbers, Bartlett's test of sphericity, Kaiser-Meyer-Olkin test of sampling adequacy results, factor extraction method, rotational method, strategies for deciding on items, eigenvalues for all factors, pattern matrix, computer program package, communalities for each variable, descriptive statistics, subscale reliabilities, and percentage of variance accounted for by each factor.

theoretical variables not readily observable by direct means (DeVellis, 2012, p. 11)." Scholars are not able to observe the direct relationship among items, but they can determine if they are sufficiently intercorrelated with one another (DeVellis, 2012). If a scientific construct is truly abstract, subscales should comprise of at least three variables to capture the true central of the concept and to ensure content validity (Viswanathan, 2010). Multiple empirical items protect against the influence of culture, biases, and item order (Morrison, 2009).

Scale dimensions (i.e., subscales, factors)

Researchers also need to investigate whether the latent variable is a unidimensional or a multidimensional measure. If it is multidimensional, the scale will need to be eventually split into subscales that represent one composite scale. In fact, 70% of measures published in a *Psychological Assessment* sample included subscales (Clark & Watson, 1995). A literature review is necessary to map the dimensional structure of the construct because researchers need to craft items that reflect their theoretical understanding of each dimension. As abstractness (and breadth) increases, one can expect the construct to be comprised of more than one dimension. For example, a literature review of *graduate student mentoring support* may reveal three separate types (or dimensions) of faculty mentoring support: psychosocial, research, and career.

Exploratory factor analysis

Following the literature review, conceptual definition proposal, and exploratory methodological work to identify dimensions and items, exploratory factor analysis (EFA) is the most often applied approach in evaluating proposed scales. Factor analysis is a group of structure analyzing procedures used to identify correlations among observable variables to aid in the data reduction of variables related to each dimension (i.e., factor) of the construct (Norris & Lecavalier, 2010). Essentially, EFA explores the data and provides guidance on factor number. In a confirmatory factor analysis (CFA), researchers specify factor number and the associated variables with each factor prior to conducting one. EFA is recommended over CFA for scale development due to the possibility that researchers are incorrect regarding their assumptions about the construct's dimensionality and to also ensure item quality. A CFA should be conducted on a separate sample to *confirm* the structure of the proposed scale resulting from an EFA (Costello & Osborne, 2005; Ford et al., 1986; Haig, 2005; Kline, 2013; Pett, Lackey, & Sullivan, 2003; Preacher & MacCallum, 2003; Worthington & Whittaker, 2006). Scholars should never assume the rigor of published scales. All published scales should be submitted to a confirmatory factor analysis to validate the dimensional structure of a measure in order to prevent large bodies of literature being built on invalid scales (Levine, Hullett, Turner, & Lapinski, 2006). The complexity of the *scale development* and *scale validation* process can result in several missteps, but *scale development* is the informational focus of this article.

Common problems and issues

Today, factor analysis decisions are becoming an even more prominent issue as the number of communication scholars using factor analysis is increasing (Park et al., 2002; Ye & Ki, 2012). Preacher and MacCallum (2003) argued the most important decisions in scale development include deciding between common factor analysis and PCA, the number of dimensions (i.e., factors) to retain, and the rotational method (oblique vs. orthogonal). The default functions in many statistical analysis programs such as SPSS or SAS lead many researchers to incorrectly utilize techniques such as principal components analysis, Varimax rotation, and eigenvalues greater than one (Conway & Huffcutt, 2003; Park et al., 2002; Reise, Waller, & Comrey, 2000). Kaiser (1970) referred to this three-pronged approach to factor analysis as the *Little Jiffy*. The *Little Jiffy* approach may be the norm in communication research; however, it does not yield precise results (Costello & Osborne, 2005).

Content analysis results of journal articles outside and inside the field of communication have focused on the choices of researchers finding that factor analysis is one of the most misunderstood procedures in the social sciences. Exploratory factor analyses have been critiqued in the fields of developmental disabilities, organizational research, counseling psychology, and psychology. The findings overwhelmingly show that researchers improperly build the structure of their scales by using inappropriate statistics and methods such as principal components analysis, eigenvalues greater than one, and orthogonal rotation (Conway & Huffcutt, 2003; Fabrigar, Wegener, MacCallum, & Strahan, 1999; Ford et al., 1986; Henson & Roberts, 2006; Norris & Lecavalier, 2010; Worthington & Whittaker, 2006). In the field of communication, three content analysis studies have been found examining the factor analytic practices of scholars finding questionable scale development procedures as well (Morrison, 2009; Park et al., 2002; Wimmer & Haynes, 1978). Furthermore, two studies found that scale development authors rarely included critical pieces of information for readers to evaluate the quality of scales (Morrison, 2009; Park et al., 2002).

Research questions

To verify the continued improper use of scale development procedures, a broad sample of communication journals was selected for a systematic quantitative content analysis. Scale development is a complex process that presents many options to scholars requiring several methodological and statistical competencies.

RQ1: To what extent will communication researchers apply improper scale development procedures in communication journals?

Second, the present research makes a unique contribution to literature by recording how authors identified dimensions and generated items for their proposed scales. Measurement model building practices should be evaluated in terms of not only methods and statistics, but concept explication best practices as well. As a result, choices prior to the application of statistical and methodological methods were reviewed to explore communication scholars' commitment to the development of valid measures.

RQ2: How do journal authors report how they generated and assessed dimensions and items for their proposed scales in communication journals?

Method

Descriptives

Articles that concentrated on survey research (79.0%) or experimental research (20.7%) were examined for this content analysis. Undergraduate students were targeted as respondents in a notable proportion (34.5%, n = 207) of the journal articles. If reported, response rates of surveys ranged from 5.0–99.0% with 6.3% (n = 38) of the studies reporting reaching less than a 25% response rate, 9.3% (n = 56) reporting between 26–50%, 8.0% (n = 48) reporting 51–75% response rate, and 4.8% (n = 29) with more than a 75% response rate. Notably, the majority of scholars (83.7%, n = 502) provided reliability levels for their scales and provided all scale items for readers (69.7%, n = 418). If reported, SPSS was the most used software used (7.0%, n = 42). In this study, the total explained variance by a scale was reported in 52.0% (n = 312) of the articles. The overall scale should account for the maximum amount of variance while not including items or dimensions that explain little variance. A scale should explain at least 50% of variance, but 75–90% is preferred (Beaver, et al., 2013; Streiner, 1994).

Sampling procedure

A quantitative content analysis of leading communication journals was selected for the purposes of describing the current state of scale development practices in the communication field. The 68 communication journals were identified based on the list of rankings from the Thomson Reuters' Journal Citation Reports. The unit of analysis was the journal article that included *exploratory factor analysis* or *principal components* to develop a latent measure. Authors of articles that included only a citation of a research study using factor analysis, only a confirmatory factor analysis using structural equation modeling, the use of factor analysis to develop a content analysis measure, or a mention, rather than the application, of the keyword search terms within the manuscript were removed from examination. This process resulted in a total of 1,318 journal articles from the 68 journals for the 10-year period.

A stratified random sampling procedure was employed because observations revealed that authors in the first 20 journals were more likely to employ scale development procedures. As a result, the articles were grouped based on their ranking: (1) 1–20, (2) 21–40, and (3) 41–68. A random sample of 600 articles (45.5% of the population) was selected: 300 articles from the top 20 journals, followed by 150 for the other 2 ranking categorizations from a total of 48 communication journals on the list. Several journals were not represented in the random selection process because the journals included no or very few studies using factor analysis.

Operational definitions

Scale development theoretical/methodological decision measures

Scale item and development measures included the reporting of relying on a literature review and/or theory, focus groups, interviews, cognitive interviews, q-sorts, pre-tests or pilot tests, experts, and other to create and assess items for scale development purposes. These variables were primarily assessed by examining the measures section within the method section. For example, literature review was coded as present if the authors cited other research that informed the creation of items for their proposed scale. All variables were treated as separate variables and dummy coded with *present* coded as 1 and *absent* as 0.

Major scale development analysis variables

The categories were based on previous content analyses and best practice recommendations from scale methodologists. The sample size used to conduct an EFA or PCA was broken down into categories including no report (see Table 2). The appropriateness of FA was examined by coding the presence or absence of Bartlett's test of sphericity and KMO. The methods of reduction included principal component analysis (PCA), common factor analysis (principal axis and/or maximum likelihood), multiple methods (PCA and common factor analysis), other, or no report. Oblique and orthogonal rotation method categories included oblique, orthogonal, both orthogonal and oblique, or no report. Rotation method type consisted of Promax, Direct Oblimin; Direct Quartimin, Varimax; Equamax; Quartimax, multiple rotation methods, other, and no report. The factor number determination options included eigenvalue greater than one, scree test, parallel analysis, minimum average partial, chi-square statistic, a priori number of factors retained (e.g., literature review suggested four factors), percentage of variance accounted for per factor, other approach, and no report. The scale item selection criteria categories included coding the presence and absence of minimal significant loadings (e.g., loadings above .40), cross loadings (i.e., significant loadings on two or more factors), item number per factor (e.g., minimum of 3 items per factor), inter-item correlations, theoretical convergence, communalities of variables, percentage of variance explained by a subscale, redundancy of wording or meaning across items, other item deletion criteria, and no report.

Table 2. Summary information of practices in the use of exploratory factor analysis in scale development (n = 600).

Characteristic	Frequency	%
Sample size*		
< 100	54	9.0
101–200	108	18.0
201–300	94	15.7
301–400	71	11.8
401–500	56	9.3
500 or larger	205	34.2
Ratio logic (e.g., 5:1; 10:1)	1	0.2
Factorability of data**		
Kaiser-Meyer Olkin (KMO)	11	8.9
Bartlett's test of sphericity	57	9.5
Type of analysis*		
Principal components analysis (PCA)	266	44.3
Common factor analysis	73	12.2
Multiple methods	13	2.2
Factor rotation method*		
Orthogonal	214	35.7
Varimax	205	34.2
Oblique	91	15.2
Direct Oblimin	27	4.5
Promax	24	4.0
Quartimax	3	0.5
Both orthogonal and oblique	8	1.3
Factor number determination criteria**		
Eigenvalue < 1	142	23.7
Scree test	45	7.5
A priori number of factors retained	20	3.3
Parallel analysis (PA)	9	1.9
Percentage of variance per factor	8	1.3
Minimum average partial (MAP)	2	0.3
Other number retention criteria	2	0.3
Item deletion or retention criteria**		
Factor loading magnitudes	154	25.7
Cross loadings	51	8.5
Inter-item correlations	35	5.8
Theoretical convergence	32	5.3
Factor number minimum	8	1.3
Percentage of variance	7	1.2
Communalities of variables	7	1.2
Item redundancy	6	1.1
Other criteria	7	1.2

Note. Variables do not add up to 100% because *no reports** were not included and some variables were absence/presence variables.**

Intercoder reliability

Reliability analyses of protocols are necessary when numbers represent the need to interpret meanings of the text. Riffe, Lacy, and Fico's (2014) sampling procedure was used to compute the test sample size for intercoder reliability resulting in 87 stories. These stories were randomly selected for intercoder reliability. Intercoder reliability on this sample was a challenge because of the nonexistent and minimal presence of some variables. The author of the study practiced coding articles not within the population to develop the protocol. It became clear that some variables would appear in a very small proportion of the articles. It was determined that these variables were still important due to the objectives of this study. Following the selection of articles, the PDF search function was employed to search for the presence of articles including the individual variables to add to the intercoder reliability analysis increasing the sample size to 114 for the author and one doctoral student. In addition, Riffe, Lacy, and Fico's (2014) recommend running multiple statistics to test the reliability of measures due to the intellectual debates associated with the appropriateness of certain reliability

statistics. I employed Krippendorf's Alpha, Cohen's Kappa, and Scott's Pi for reliability analyses for the nominal level variables. Reliability for the variables ranged from .79–1.0. For reliabilities of each variable and the codebook, please contact the author. Previous content analysis authors on factor analysis practices did not report intercoder reliability for their variables (Fabrigar et al., 1999; Henson & Roberts, 2006; Morrison, 2009; Norris & Lecavalier, 2010; Park et al., 2002; Worthington & Whittaker, 2006).

Results

The intent of data collection was to assess whether communication scholars have improved their scale development practices. RQ1 asked to what extent communication scholars followed scale development procedural best practices. As shown in Table 2, the present results showed that authors rarely reported that they inspected Kaiser-Meyer Olkin (KMO) (8.9%) or Bartlett's test of sphericity (9.5%) statistics prior to conducting factor analysis. Despite a consistent recommendation to not use PCA (e.g., Conway & Huffcutt, 2003; Costello & Osborne, 2005; Ford et al., 1986; Morrison, 2009; Norris & Lecavalier, 2010), 44.3% of the articles stated using it or they did not report the type of analysis used in 41.0% of the articles in this study. This PCA finding is proportionately less than in the Park et al. (52.9%) and more than in the Morrison (40.2%) studies on practices in communication journals.

In the present research, the eigenvalue greater than one rule (23.7%) was the most often applied method used to determine the number of factors or dimensions in a model (see Table 2), which is between the 27.7% found in the Park et al. (2002) study and 16.3% found in the Morrison (2009) study. The application of the eigenvalue rule was followed by the scree test (7.5%), theory/a priori number of factors (3.3%), and/or percentage of variance accounted by individual factors (1.3%).

In this present content analysis, orthogonal rotation (35.7%) was favored over oblique rotation (15.2%) despite recommendations to not use it. Specifically, communication scholars only reported applying the orthogonal rotation Varimax (34.2%). Notably, oblique rotation was applied slightly proportionately more often in comparison to the Park, Dailey, and Lemus (57.1% (orthogonal); 10.9% (oblique) and Morrison (34.2% (orthogonal); 11.4% (oblique) studies.

In the item deletion process, communication researchers most often relied on the rules associated with cross loadings and factor loading magnitudes. In the present study, however, more than 76.2% of the articles did not state whether authors relied on a priori cutoff criteria to determine item retention or deletion.

RQ2 queried what methods communication authors employed to capture the breadth of a concept prior to the launch of their quantitative study. The results showed that authors reported primarily relying on literature and/or theory to guide in their development of items (66.8%), followed by administering a pretest or pilot test (12.2%) with subjects. Authors, however, rarely reported seeking expert guidance (4.0%), administering interviews (2.0%), conducting focus groups (2.0%), conducting cognitive interviews (0.8%), using q-sorts (0.2%), or applying other approaches (3.5%).

10 steps in scale development and reporting

The multitude of choices involved in each scale development step have probably steered many researchers to shy away from best practices. The literature is sometimes technical, which may lead many users to simply trust the defaults in their statistical software packages. The goal is to highlight 10 major steps along the scale development decision tree to make the process more accessible and to encourage more systematic applications in future research.

STEP 1: research the intended meaning and breadth of the theoretical concept

Theory and research should play the strongest role in guiding the identification of empirical attributes that represent the abstract construct (Chaffee, 1991; Clark & Watson, 1995; Cronbach & Meehl, 1955; DeVellis, 2012). Theory should pre-specify the structure and meaning of a construct. The quality of a

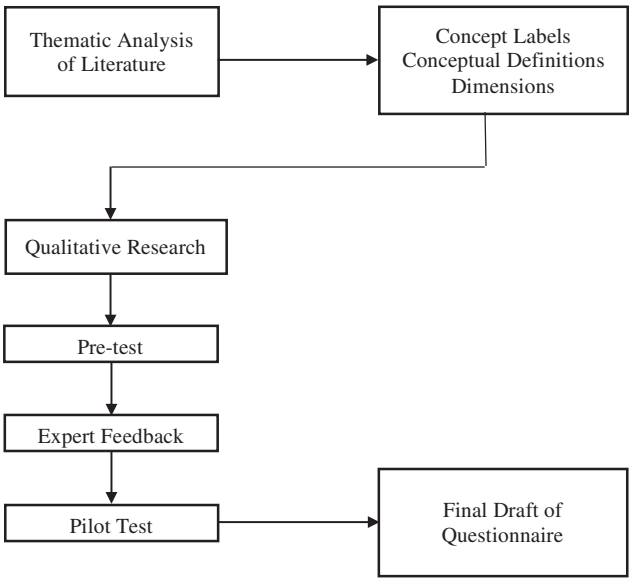

Figure 1. Research the intended meaning and breadth of the construct.

measure rests on the judgement of the researcher, which includes the selection of dimensions and wording of items to be included on the scale. Simms (2008) argued, "The apparent simplicity and efficiency of the [survey] method can be illusionary, as much time, and consideration are needed to develop measures that allow us to make reliable and valid inferences about people (p. 414)." Ill-defined constructs require a large amount of time and effort on the part of the researcher to explicate the concept (Chaffee, 1991). The durability of measures would likely withstand statistical and methodological challenges if scholars relied on multiple methods to build them. Meaningful measurement occurs when the body of questions successfully achieves the intended representation of the abstract construct. Yet the specification of a theoretical concept can result in an indefinite number of items, but it is important to remember the goal is to find the optimal sample of items and dimensions to empirically represent its abstractness. Chaffee (1991) referred to this process as meaning analysis in which the theoretician employs logical procedures to justify the conceptual and operational definitions of the construct. Careful conceptualization, item selection, and wording are necessary to ensure content validity (Clark & Watson, 1995; Cronbach & Meehl, 1955; Worthington & Whittaker, 2006). The Step 1 section provides suggestions regarding how researchers should address the validity of the concept prior to the methodological applications and statistical analyses by addressing the labeling of the construct and dimensions, creating conceptual definitions, seeking to understand the breadth of the construct, and generating and refining items for the scale (see Figure 1).

Select appropriate conceptual labels

The naming of the construct and each subscale or dimension influences future interpretations of the concept. Researchers should be thoughtful when deciding what to label each concept. During a literature review, it is common to find conceptual redundancy even though concepts vary in labels within and across disciplines. As a result, a communication scholar should share the logic associated with the selection of a label. Unfortunately, scholars rarely shared the factor naming logic of their concepts in this study (3.5%, n = 21). One suggestion to address subjectivity is to invite a panel to review the items of each factor to determine the most appropriate concept label. Often times, the first items on each subscale can provide naming guidance for each dimension.

Select conceptual definitions

One beginning step in theory building is to write and defend a formal conceptual definition of the construct. A good definition adds clarity to an ambiguous concept by telling the reader what it is supposed to mean and what it is not supposed to mean (Chaffee, 1991). For example, *quality of life* could refer to attributes such as mobility, quality of sleep, eating problems, but it also could include water quality, noise levels, and pollutants. And, thus, we have to be conceptually clear by providing a dictionary-like definition with concrete keywords or attributes that are useful for academic audiences, which then should influence the framing and wording of items included the scale.

Identify potential dimensions and items

The literature review should not only be conducted through the lens of the construct label, but authors should seek literature to identify possible dimensions of the construct and defend their association with the overarching construct. An additional function of the literature review is to gain a more concrete understanding of the construct by identifying conceptual definitions of each dimension and relevant empirical items that describe each dimension.

Conduct qualitative research to generate and validate dimensions and items

Several rounds of information gathering should be conducted following the conceptualization of the construct and prior to full data collection to reduce measurement error. The pool of scale items needs to be concise, clear, distinct, and reflect the chosen conceptual definition (DeVellis, 2012). Interviews, focus groups, and expert feedback are critical in the item generation and dimension identification process (Broom, 2006; Clark & Watson, 1995; DeVellis, 2012; Pett et al., 2003; Simms, 2008; Worthington & Whittaker, 2006). Focus group and interview research, however, does not appear often in media and communication journals in the United States (Potter & Riddle, 2007; Ye & Ki, 2012). Furthermore, previous content analyses of communication journals found that scholars very rarely combined quantitative and qualitative methods (Cooper, Potter, & Dupagne, 1994; Kamhawi, 2003; Trumbo, 2004; Ye & Ki, 2012).

This is unfortunate, because based on experience, it is very common that participants will reveal additional dimensions critical to the meaning of the construct based on results stemming from qualitative research efforts. In regard to developing a qualitative research protocol with the specific goal of scale development in mind, researchers should begin by asking participants broadly about their interpretation of the construct, followed by questions concerning the participants' interpretations of each dimension found in the literature review; including to what degree they agree with each dimension. The researcher should probe the participants for possible scale items including the wording of items representing each dimension as well.

Use feedback to refine scale

Measurement error can arise for many reasons such as complex wording or language, questions requiring estimation, vagueness in questions or response categories, double-barreled questions, and leading or biased questions. Q-sorts, pilot tests, expert feedback, cognitive interviews, and pretests are especially useful in questionnaire and item refinement (Clark & Watson, 1995; DeVellis, 2012; Pett et al., 2003; Worthington & Whittaker, 2006).

Pre-test. Pre-tests on smaller samples are useful for survey and item feedback prior to the launch of the data collection, while pilot tests, following a pre-test, can be employed to assess how the data will fall to determine whether items should be added or deleted. In this study, scholars struggled with distinguishing a pre-test from a pilot test, which resulted in the combining of both categories into one for coding purposes. The use of pre-tests in surveys dates back to the 1930s. Researchers can employ multiple pre-tests to refine their questionnaire questions and design. Pretesting can address areas such as ambiguous, leading, confusing, difficult, skipped, sensitive, and missing questions. The

goal is to reduce measurement error, response burden, and question inaccuracy. Pretest sample sizes should be small, but similar as possible to targeted respondents. Pretest sample sizes can range from 5–100 people depending upon the diversity of target subpopulations.

Pretesting can be conducted with focus groups, cognitive interviews, interviews, group debriefing, or individual debriefing. Cognitive interviews consist of probes (e.g., "What does 'some of the time" mean to you?" and think-alouds (i.e., the interviewer requests that the respondent reads each question and verbalizes what comes to mind when reading each question.). Scholars can conduct behavioral coding such as watching whether respondents hesitate or frown when reading a question. Behavioral coding consists of monitoring or reviewing taped interviews or survey participation. Following one or multiple pre-tests, statements, instructions, and the questionnaire design should be edited based on this feedback (Couper, Lessler, Martin, Martin, Rothgeb, & Singer, 2004; Drennan, 2003; Lewis, Templeton, & Bryd, 2005; Reynolds, Diamantoupoulos, & Schlegelmilch, 1993; Ruel, Wagner, & Gillespie, 2016).

Expert feedback. Experts should consist of methodologists, intended participants, and subject-matter researchers. The goal is get their feedback on item quality and how well each item reflects the overarching construct. RESEARCHERS can provide instruction to the experts BY asking them TO provide individual feedback on items BY asking them to assess item validity through a Likert-type scale OR open-ended feedback (DeVellis, 2012; Ruel et al., 2016).

Pilot test. A pilot test is rehearsal of the actual survey in actual field conditions. The quantitative data collection part of a pilot test is especially useful in identifying how data will fall around each factor and identifying skipped questions. In order to conduct an EFA, the pilot test sample size should range from 50–100 participants. Once edits based on the pilot test are complete, the survey is set for full-scale administration (Lewis, Templeton, & Bryd, 2005; Ruel et al., 2016). At a minimum level, scholars should employ: (1) a literature review; (2) at least one type of qualitative research; (3) expert feedback; and (4) a pre-test when developing their scale dimensions and items.

STEP 2: determine sampling procedure

Decide an appropriate sample size

Following the content development stage, scholars can proceed toward factor analysis. Factor analysis is a large sample size technique. Insufficient sample sizes result in unstable factors and decreased generalizability (Kline, 2013; Tabachnick & Fidell, 2007). Methodologists vary regarding recommended sample sizes with the exception of stating that more participants result in more stable scales. Generally, most scholars recommend a sample size of at least 300 (McCrosky & Young, 1979; Henson & Roberts, 2006; Pett et al., 2003; Worthington & Whittaker, 2006). Recommendations range from a sample size of 50 (Barrett & Kline, 1981)–400 (Aleamoni, 1976). Comrey and Lee (1992) provided one guide: 50 (very poor), 100 (poor), 200 (fair), 300 (good), 500 (very good), and 1000 (excellent). If the communalities and factor loadings are low, it is suggested to increase the sample size (Mundfrom, Shaw, & Ke, 2005). A communality (h^2) is the proportion of variance accounted by each individual variable for one factor. In this study, the communalities for each item were shared in only nine studies. Communalities are considered high if they are above .80; however, the more common range in the social sciences is from .40–.70 (Costello & Osborne, 2005). Thus, lower sample sizes can be defended if a majority of communalities (<.50) and factor loadings (<.40) are high (see Worthington & Whittaker, 2006).

A scale with fewer variables, however, requires fewer participants. Every scale varies in dimensions, item numbers, communality sizes, factor correlations, and item-factor correlations (Floyd & Widaman, 1995; Osborne, 2014; Pett et al., 2003; Worthington & Whittaker, 2006). For these reasons, Guadagnoli and Velicer (1988) and other methodologists argue that item ratios are more relevant than the previously mentioned sample size defense logics. Gorsuch (1983) and others

suggested following lower minimum ratios of participants to items (5:1 or 10:1), while Osborne (2014) argued for 20 cases per variable to ensure robust, generalizable results. Costello and Osborne (2005) addressed the sample size debate by examining how varying sample sizes affected error rates regarding the factor structure of a scale. Larger solutions (20:1) produced the most correct solutions and classification of items.

Many studies with smaller sample sizes (e.g., 100 people) are published (Conway & Huffcutt, 2003; Russell, 2002). It is very common to use adequate rather than ideal sample sizes especially when dealing with difficult to access populations (Worthington & Whittaker, 2006). For example, one content analysis on a subset of public relations journals showed that a notable proportion (42.8%) of articles published relied on a sample size ranging from 101–200 people (Ki & Shin, 2006). Approximately 43% of the journal articles in this study included a sample size of less than 300 (see Table 2). Based on coder observations, studies with lower sample sizes surveyed professionals (e.g., journalists, public relations practitioners) or utilized the experimental method. Lower sample sized studies are quite common based on previous content analysis results (Conway & Huffcutt, 2003; Fabrigar et al., 1999; Henson & Roberts, 2006).

Recommendations for a sufficient sample size are challenging for these reasons. Additionally, methodologists recommend using two-to-three times as many items than one expects to be on the final scale. For example, a communication scholar would ideally write 60 items that would ultimately result in a final 20-itemed scale, which means the sample size would need to consist of 1200 participants to follow the 20:1 rule of thumb. There is no clear consensus. To be of use beyond a particular sample, the most optimal recommendation to follow is the 20:1 ratio logic to reduce the error rate, but communication scholars should abide by the minimum standard of a 5:1 item ratio of participants to number of variables.

STEP 3: examine data quality

Data cleaning is essential to ensure that findings are accurate and replicable. Researchers should check for missing data, absence of outliers, linearity, and extreme multicollinearity (Beavers et al., 2013). For example, outliers should be justifiably changed or deleted if only a few exist because of their impact on outcomes. Missing data are not ignorable, the researcher should inspect patterns of missing data. Scholars should consider deleting cases when the majority of responses (50% or more) contain missing data. It is encouraged to read books on data cleaning best practices (e.g., Hair, Black, Babin, & Anderson, 2010; Meyers, Gamst, & Guarino, 2006; Myers, 2011). One suggestion is to communicate to readers in research articles how one dealt with missing data, outliers, or other potential issues.

STEP 4: verify the factorability of the data

Inspection of the correlation matrix, Bartlett's test of sphericity, and Kaiser-Meyer-Olkin (KMO) provides information as to whether factor analysis should be applied to data. The correlation matrix should include numbers at levels of .30 or higher, Bartlett's chi-square should be significant at a probability of .05 or less, and a KMO value of .60 or higher is recommended before proceeding with factor analysis (McCrosky & Young, 1979; Pett et al., 2003; Tabachnick & Fidell, 2007). Previous research, however, showed that authors do not often report that they examined these statistics prior to proceeding with a factor analysis (Henson & Roberts, 2006; Worthington & Whittaker, 2006).

STEP 5: conduct common factor analysis

The group of extraction and rotation techniques to identify latent constructs is referred to as *common factor analysis* or *exploratory factor analysis*. It is important to be aware that factor analysis is not one statistical procedure, but rather a group different statistical and methodological choices

(Beavers et al., 2013). And thus, the accuracy of the results rests on the quality of these decisions made during each step.

EFA seeks to identify the shared variance among variables. One common mistake made by researchers is the use of principal components analysis (PCA) (Goldberg & Velicer, 2006; Reise et al., 2000). PCA is conceptually and mathematically distinct from common factor analysis. Common factor analysis does not focus on explaining the amount of variance, but rather it assesses the sources of common variation. The problem with PCA in latent measurement development is that the sizes of components are inflated because they include error variance, which can lead researchers to retain too many components (Conway & Huffcutt, 2003; Costello & Osborne, 2005; Goldberg & Velicer, 2006; Preacher & MacCallum, 2003; Snook & Gorsuch, 1989). Common factor analysis (i.e., principal axis factoring or maximum likelihood) results are more generalizable when submitting hypothesized models to a confirmatory factor analysis (Haig, 2005; Worthington & Whittaker, 2006). And, thus, it is recommended to conduct common factor analysis rather than PCA because most methodologists do not support the application of it for the previously mentioned reasons.

STEP 6: select factor extraction method

The extraction method evaluates the correlation/covariation among all of the scale items and seeks to extract latent variables from the manifest variables (Osborne, 2014). Essentially, we are interested in factors, and we are transforming our data from a variable space to a factor space. Statistical package options include unweighted least squares, generalized least squares, maximum likelihood, principal axis factoring, alpha factoring, and image factoring. Osborne (2014) noted little information exists on the drawbacks and advantages of each method. Based on observations and research, scholars most often select either principal axis factoring (PAF) or maximum likelihood (ML). Both MA and PAF try to reproduce the correlation matrix. PAF, the most robust method, is recommended to use when sample sizes are small and normality is violated, and ML is supported when data is normally distributed (Fabrigar et al., 1999; Nunnally & Bernstein, 1994).

STEP 7: determine number of factors

Next, the precise number of the subscales to include is rarely clear cut. I will review the four common approaches used to determine factor number.

Eigenvalue greater than one

The most often used, but not supported approach, is the eigenvalues greater than one rule when assessing factor number (Ford et al., 1986; Henson & Roberts, 2006; Morrison, 2009; Russell, 2002). Eigenvalues measure the variance accounted for by each factor. The larger the eigenvalue, the more variance explained by a factor. In the past, Kaiser (1960) believed that eigenvalues greater than one resulted in stable dimensions. Today, several methodologists advise against using the cutoff of one eigenvalue when determining the number of factors because the number of factors above one is greatly related to the total number of variables included a model, which researchers argue leads to the extraction of too many or too few factors (Fabrigar et al., 1999; Goldberg & Velicer, 2006; Kline, 2013; Worthington & Whittaker, 2006). And, thus, it is not recommended for the development of measurement models.

Scree test

The scree test is a graphic that allows researchers to estimate the number of factors to retain. A scree is a visual plot of eigenvalues derived from the factors. The cutoff line for number of factors is determined when a line elbows off from a somewhat subjectively straight dotted line. The scree test is considered more accurate than the eigenvalue rule (McCrosky & Young, 1979; Pett et al., 2003; Preacher & MacCallum, 2003; Reise et al., 2000).

Parallel analysis (PA)

PA was developed by Horn (1965), and it compares eigenvalues from the results against a randomly ordered data set. The PA method has been recommended for determining an accurate number of factors to accept (Humphreys & Montanelli, 1975; Kline, 2013; Velicer, Eaton, & Fava, 2000; Watkins, 2006; Zwick & Velicer, 1982). Scholars should compare their data's eigenvalues with the eigenvalues produced in PA. Factors are retained when their eigenvalues are larger than the eigenvalues created by a random data set. A lack of awareness and the need to download the free program may be reasons for people not adopting PA. However, journals such as *Educational & Psychological Measurement* and the *Journal of Personality Assessment* require reporting of PA in scale development articles (Pallant, 2010). To access the stand-alone program, I recommend visiting http://edpsychassociates.com/Watkins3.html for a free download of Monte Carlo PCA for Parallel Analysis (Watkins, 2006).

Minimum average partial (MAP)

MAP involves partialling each successive factor from a correlation matrix to create a partial correlation matrix. MAP, introduced by Velicer, examines off-diagonal partial correlations after successively removing the effects of factors (Velicer, 1976). The average of the squared correlations of the off-diagonal partial correlation matrix is computed after each factor is extracted and its effect on the correlations between items is excluded. The average should decline as long as shared variance is being extracted, and the average will then increase when error variance dominates. Factors are retained when the averaged square partial correlation reaches its lowest value (Goldberg & Velicer, 2006; Watkins, 2006). MAP is not as readily available, which likely partially explains its low use. O'Conner (2000) provided guidance on HOW TO USE these procedures WITH SPSS, SAS, and MATLAB syntax commands for both MAP and PA. (https://people.ok.ubc.ca/brioconn/nfactors/nfactors.html).

Social scientists concerned about the optimal number of factors should determine them based on theory and multiple tools. Researchers should use a combination of the following: theory/literature review, visual scree test, parallel analysis, and minimum average partial when determining the appropriate number of factors in a hypothetical model (Conway & Huffcutt, 2003; DeVellis, 2012). If these previously mentioned procedures reveal that multiple factor numbers are a possibility, such as 3, 4, and 5 factor models, authors should rerun analyses examining 2–6 factor models to determine the optimal factor number for the scale.

STEP 8: rotate factors

Rotation is necessary in order to more clearly identify the scale's factors (or dimensions). Oblique and orthogonal are two types of rotation methods available to researchers. Content analyses reveal most scholars use Varimax, an orthogonal rotation, which forces factors to not correlate (Borgotta, Kercher, & Stull, 1986; Conway & Huffcutt, 2003; Ford et al., 1986; Norris & Lecavalier, 2010).

Orthogonal produces uncorrelated factors. The problem is that it is uncommon for factors to not correlate with one another in the social sciences (DeVellis, 2012; McCrosky & Young, 1979). A Varimax rotation, the most popular orthogonal rotation type, pushes high factor loadings higher and low factors lower because they are not allowed to correlate (Tabachnick & Fidell, 2001). Varimax is biased against finding a general factor when one exists and it generates more cross loadings (Fabrigar et al., 1999; Gorsuch, 1997). As a secondary analysis of data shows in Table 3, a Varimax rotation resulted in 9 significant cross loadings. It is also produced 13 lower and 12 higher loadings than a Promax rotation. If factors are substantially correlated, an oblique solution improves the ability of the researchers to approximate the structure of the model (Fabrigar et al., 1999; Ford et al., 1986; Gorsuch, 1997). If factors do not correlate, scholars should assess whether they are measuring two separate constructs rather than one. Thus, it is recommended to use oblique rotation because it more accurately represents most models in communication research because it allows factors to correlate.

Table 3. Secondary data analysis comparison of orthogonal (1st item) and oblique (2nd item) rotations (n = 551).

Items	1	2	3	4	5	6	7
1	**0.53**	−0.02	0.03	0.37	−0.16	0.06	0.08
1	−0.05	−0.02	**0.69**	−0.03	−0.06	0.12	0.05
2	**0.46**	0.27	0.06	−0.12	0.07	0.02	−0.09
2	**0.36**	0.13	−0.09	−0.15	0.21	0.10	0.08
3	0.26	−0.21	0.03	0.22	−0.13	0.07	0.12
3	−0.06	0.00	**0.46**	0.08	−0.20	0.00	0.02
4	**0.39**	**0.32**	0.07	0.12	0.19	−0.27	−0.15
4	0.04	−0.04	0.03	0.07	**0.60**	0.11	−0.01
5	**0.45**	0.04	0.14	**0.48**	−0.07	0.03	−0.01
5	−0.09	−0.20	**0.76**	0.02	0.12	0.09	0.14
6	**0.54**	0.12	0.04	0.31	0.06	−0.02	0.04
6	−0.04	0.10	**0.51**	−0.03	0.20	0.03	0.07
7	**0.51**	0.26	0.10	0.27	0.23	−0.21	0.08
7	−0.01	0.10	**0.35**	0.01	**0.48**	−0.12	−0.04
8	0.24	0.31	0.05	−0.05	0.15	−0.20	−0.15
8	0.14	−0.02	−0.18	−0.01	**0.47**	0.10	−0.03
9	**0.55**	0.08	−0.24	−0.02	0.04	0.00	0.12
9	−0.03	**0.55**	0.08	−0.11	0.02	0.07	−0.09
10	**0.45**	0.10	0.05	0.09	0.13	0.04	0.18
10	0.14	0.29	0.27	−0.11	0.05	−0.19	0.01
11	**0.62**	0.08	**−0.36**	−0.02	0.10	0.08	−0.07
11	−0.14	**0.67**	−0.02	−0.09	0.07	0.25	0.07
12	**0.60**	−0.05	**−0.44**	−0.11	0.10	0.10	0.07
12	−0.17	**0.88**	−0.08	−0.08	−0.10	0.15	−0.01
13	**0.56**	−0.08	−0.11	−0.08	−0.25	−0.18	0.25
13	0.18	0.22	0.13	0.05	−0.12	0.12	**−0.37**
14	**0.57**	0.05	−0.03	−0.06	−0.03	−0.11	0.19
14	0.23	0.28	0.09	−0.01	0.05	−0.01	−0.20
15	**0.44**	0.11	**0.33**	−0.24	0.04	0.11	0.03
15	**0.76**	−0.04	−0.08	−0.08	0.01	−0.13	0.10
16	**0.44**	−0.14	−0.07	−0.02	0.23	0.11	0.03
16	0.02	**0.49**	0.04	0.09	0.00	−0.09	0.16
17	−0.06	**0.50**	−0.05	−0.03	−0.02	0.18	0.14
17	0.13	0.12	0.04	**−0.61**	−0.06	−0.11	0.00
18	0.20	**−0.65**	0.11	0.05	0.02	−0.14	−0.06
18	−0.06	−0.07	0.08	**0.70**	−0.04	0.03	0.02
19	**0.46**	−0.08	0.07	−0.09	0.19	0.11	−0.12
19	0.25	0.25	−0.06	0.12	0.09	0.01	0.23
20	**0.45**	**−0.62**	0.11	−0.02	0.18	−0.15	0.00
20	0.05	0.21	0.01	**0.71**	0.05	−0.09	0.03
21	**0.57**	−0.15	0.03	0.03	0.05	0.28	−0.13
21	0.21	0.26	0.20	0.05	−0.12	0.12	**0.33**
22	**0.56**	−0.24	0.14	−0.15	−0.09	−0.06	0.01
22	**0.43**	0.05	0.00	0.27	−0.05	0.09	−0.05
23	**0.50**	−0.19	−0.17	−0.17	0.26	−0.02	0.11
23	0.04	**0.69**	−0.19	0.17	0.03	−0.12	−0.03
24	**0.67**	0.05	−0.24	−0.09	−0.22	0.03	−0.14
24	0.13	0.28	0.00	−0.05	−0.03	**0.49**	−0.03
25	**0.59**	−0.09	0.09	−0.17	−0.13	0.05	0.01
25	**0.49**	0.10	0.01	0.07	−0.10	0.14	−0.01
26	**0.62**	0.00	−0.02	0.10	**−0.33**	−0.03	−0.16
26	0.18	−0.14	0.29	0.06	0.00	**0.51**	−0.01
27	**0.59**	0.04	**−0.33**	0.04	−0.14	−0.13	−0.23
27	−0.17	**0.26**	0.00	0.09	0.19	**0.58**	−0.05
28	**0.54**	0.16	0.21	−0.27	−0.07	−0.03	0.00
28	**0.71**	−0.03	−0.15	−0.06	0.09	0.07	−0.06
29	**0.47**	0.08	**0.38**	−0.20	−0.07	−0.04	−0.03
29	**0.77**	−0.26	−0.05	0.06	0.10	0.00	0.00
30	**0.52**	0.15	0.19	−0.05	−0.06	−0.06	0.07
30	**0.47**	−0.04	0.12	−0.06	0.11	0.00	−0.07
31	**0.48**	0.10	0.16	−0.11	−0.04	0.17	0.03
31	**0.51**	0.08	0.11	−0.16	−0.09	0.00	0.11
32	**0.32**	−0.11	0.18	0.23	0.11	0.30	−0.08
32	0.10	0.02	0.45	0.02	−0.09	−0.06	0.40

In an oblique rotation, researchers can review two matrices, but the pattern matrix should be evaluated to assess salience and simple structure. Oblique rotation options include Direct Oblimin and Promax. Both allow factors to correlate, but Promax begins with an orthogonal solution and then transforms it to an oblique solution (Hendrickson & White, 1964). Promax has been argued to be more robust, and it is recommended (Thompson, 2004).

STEP 9: retain and delete items based on a priori criteria

Ideally, scholars include around three times as many items on a questionnaire than will be included in the final scale. Following the item writing process, researchers need to winnow down the number of items by deciding which items to retain or discard. Items can be referred to as variables and questions as well. It is critical to optimize scale length to ensure participant motivation.

Simple factor structure is often determined based on several pre-established general criteria: factor item loadings at or above the .30–.50 level, no cross-loadings (i.e., significant loadings on more than one factor), no factors with fewer than three items, reliability levels, and theoretical convergence (Clark & Watson, 1995; Costello & Osborne, 2005; DeVellis, 2012; Fabrigar et al., 1999; Gorsuch, 1997; Hair et al., 2010; Kline, 2013; Norris & Lecavalier, 2010; Tabachnick & Fidell, 2007; Tinsley & Tinsley, 1987; Worthington & Whittaker, 2006). If reported in journal articles, cross loadings and significant loading magnitude levels are the most commonly applied criteria (Worthington & Whittaker, 2006).

It is suggested that scholars evaluate items based on multiple criteria: theory, communalities, items loadings, no significant cross-loadings, minimum of three salient loadings, factor reliability levels, and parsimony. I will overview minimum item loadings and number of items per factor because these issues appeared most often in the results.

Minimum factor item loading

In the present study of communication research, authors tended to vary in their factor loading cutoff levels if they were reported: (1) .20 or lower (0.3%) and (2) .30–.39 (3.7%), .40–.49 (7.5%), .50–.59 (6.7%), .60–.69 (5.3%), .70 or higher (0.3%). A factor loading is a correlation between an item and a factor. Loadings can be positive or negative depending on their correlation with other variables. Based on coder observations, it appears that several scholars followed the 60–40 criterion developed by McCroskey and Young (1979) to determine *significance*. That rule, however, is not supported by outside scholars. Based on factor analysis recommendations, acceptable levels are notably lower: .30 (Kachigan, 1986; Pett et al., 2003; Russell, 2002; Tinsley & Tinsley, 1987), .32 (Worthington & Whittaker, 2006), .35 (Clark & Watson, 1995), .40 (Ford et al., 1986; Hair et al., 2010; Reinard, 2006), and .50 (Mertler & Vannatta, 2001). A squared correlation of .30 equals .09 and a .32 means the item accounts for 10.2% of the overlapping variance with the items within a factor. Based on a review of recommendations and due to the variations found in the present study, it is recommended that a significant cut-off level be at .32, but anywhere between .30–.40 is supported by the literature review.

Number of items per factor

It is recommended that each subscale include at least three items in order to capture the true central of each dimension. Methodologists recommend being over-inclusive with the number of items (Clark & Watson, 1995; Kline, 2013; Loevinger, 1957). One flaw with many scales is that they use one- or two-item measures to tap into an abstract concept. Two-itemed scales are only recommended if items are highly correlated (i.e., $r < .70$) (Worthington & Whittaker, 2006). An additional analysis of the articles in this study found that 15.8% of the journal articles included two-itemed factors in their new scales. Most methodologists endorse that each factor should include a minimum of three variables, however, at least four or five variables per dimension are recommended (Costello & Osborne, 2005; Fabrigar et al., 1999; Gorsuch, 1983; Kline, 2013; Reise, Thurstone, 1947; Waller & Comrey, 2000).

It is suspected, however, that many scales contain highly redundant items to achieve this ideal of three items per factor. During the development stage, scholars can select a few empirical items that are redundant in meaning in order to identify the best statements that represent the construct. The final scale, however, should not contain insufficiently distinct items that inflate reliability levels and have a negative impact on the goal of parsimony. In fact, high coefficient alpha levels may suggest an over-inclusion of certain items. Qualitative research is key to preventing redundancy issues. Scale items should be related, but also be distinct aspects of the latent factor. This goal is very difficult to accomplish, but it is a hallmark of the best scales.

STEP 10: present results

An important point that needs to be addressed is the continued practice of not reporting information regarding the logic and choices made at major decision points in the scale development process. Scholars should report the following information: construct and subscale naming logic and conceptual definitions, sample size logic, methods for determining factor numbers, Bartlett's test of sphericity and Kaiser-Meyer-Olkin test of sampling adequacy results, factor extraction method, rotational method, strategies for selecting items, eigenvalues for all factors, pattern matrix, computer program package, communalities for each variable, descriptive statistics, subscale reliabilities, and percentage of variance accounted for by each factor.

The discussion of results stemming from a scale development manuscript should reflect a discussion explaining each factor including the naming logic and conceptual definitions associated with any new factors. Scale development research explores factors, and it does not present hypotheses and research questions (Pett et al., 2003).

Conclusion

My intent is to empower the reader by dissecting the fundamentals of the scale development methodology in hopes of advancing communication and media research because scale development is a complex, multi-step process. Scale development is truly a craft that requires practice. Good theorists should explicate their concepts by precisely defining them. Thoughtful attention to measurement may result in a forced clarification of our concepts that represent our fields, which will then enable us to stand stronger on our body of scientific knowledge—what we claim we know.

The misuse associated with exploratory factor analysis likely stems from a lack of awareness in the field of communication. As noted, the reliance on statistical package defaults most often goes against best practices in scale development. The accuracy and soundness of our predictions rests heavily on not only our statistical and methodological decisions, but our theoretical logic as well. This article makes a unique contribution by studying and reviewing the concept explication process involved with scale development. It is important to note that the measurement model building process is exploratory, and a resulting measure may continue to evolve over time. A scientist should not hesitate to move back and forth between meaning and empirical analysis until the precision and structure of a measure adequately represents the latent construct because it is not a linear process (Chaffee, 1991; Osborne, 2014).

Educational leaders including professional associations can address these issues. The hiring of methodologists at educational institutions, the support of methodological and theoretical divisions, and the teaching of scale development can influence graduate students and faculty members to enact a more critical eye toward measurement quality. To engage students, it is important to conceptually explain the statistical and methodological decision logics in order to encourage the adoption of best practices. In the meantime, many reputable and dedicated methodologists have written guidelines on the ideal statistical and methodological choices associated with scale development, and I encourage readers to read the literature listed in the references section for additional guidance or read some scale development papers for manuscript writing guidance (Carpenter, Grant, & Hoag, 2016; Carpenter, Makhadmeh, & Thornton, 2015).

Future research needs to be conducted to encourage measurement literacy in other areas. For example, a content analysis could focus on index development practices in communication. Methodologists refer to indices and summated scales as two different types of measures. An index is a formative measure, whereas a summated scale is a reflective measure. Some example differences between the two include that the deletion of one item from an index negatively affects the theoretical meaning of a construct, and indices, made up of manifest items such as what is often found in content analysis work, are not expected to correlate with one another (Bollen & Lennox, 1991; Morrison, 2009). In the present study, it was found that several scholars referred to a summated scale as an index (11.7%). As a result, it will be challenging to content analyze index development efforts until the communication field addresses the differences between the two types of measures. A future content analysis could also assess how scholars applied confirmatory factor analysis techniques to validate scales. Several content analyses have been done outside the field of communication evaluating such practices (Jackson, Gillaspy, & Pure-Stephenson, 2009; Schreiber et al., 2006; Worthington & Whittaker, 2006). This article does not cover all measurement approaches. And thus, it is hoped that this article will simply offer readers a foundation on measurement theory and procedures.

Researchers could conduct a content analysis of manuscript reviews to determine how often factor analysis is mentioned in reviews and examine what type of guidance is provided by editors and reviewers. Additionally, a comparative analysis of journals with lower or higher word limits could be carried out to determine the extent in which manuscript word limits play a part in the communication of measurement development practices.

Journal editors and reviewers play a critical role in raising the field to these standards. I would encourage journal editors to welcome manuscripts that focus on measurement development. Not all science consists of hypothesis testing; science can also consist of the theoretical development of measures. Most of the articles in this study concentrated on both scale development and hypothesis testing (92%) rather than scale development (8%). The targeted attentiveness on measurement development by journal editors may help clarify concepts and teach journal readers about best practices.

Acknowledgments

I would like to thank Marley Watkins who provided me with a strong foundation in scale development and inspired me to continue down this path in the field of communication. I would also like to acknowledge the editor Jörg Matthes and the reviewers for encouraging my work to be much better.

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

ORCID

Serena Carpenter http://orcid.org/0000-0002-8814-4185

References

Aleamoni, L. M. (1976). The relation of sample size to the number of variables in using factor analysis techniques. *Educational and Psychological Measurement*, *36*, 879–883.

Barrett, P. T., & Kline, P. (1981). The observation to variable ratio in factor analysis. *Personality Study and Group Behavior*, *1*, 23–33.

Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in education research. *Practical Assessment, Research & Evaluation*, *18*(6), 1–13.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement – a structural equation perspective. *MIS Quarterly*, *28*(2), 305–314.

Borgotta, E. F., Kercher, K., & Stull, D. E. (1986). A cautionary note on use of principal components analysis. *Sociological Methods and Research*, *15*, 160–168.
Broom, G. M. (2006). An open-system approach to building theory in public relations. *Journal of Public Relations Research*, *18*(2), 141–150.
Carpenter, S., Grant, A. E., & Hoag, A. (2016). Journalism Degree Motivations (JDM): The development of a scale. *Journalism & Mass Communication Educator*, *71*(1), 5–27.
Carpenter, S., Makhadmeh, N., & Thornton, L. J. (2015). Mentorship on the doctoral l level: An examination of communication mentors' traits and functions. *Communication Education*, *64*(3), 366–384.
Chaffee, S. H. (1991). *Explication* (pp. 1–42). Beverly Hills, CA: Sage.
Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309–319.
Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Conway, J. M., & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods*, *6*(2), 147–168.
Cooper, R., Potter, W. J., & Dupagne, M. (1994). A status report on methods used in mass communication research. *Journalism Educator*, *48*(4), 54–61.
Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, *10*(7), 1–9.
Couper, Mick P., J. Rothgeb, J. Lessler, E.A. Martin, J. Martin, and Eleanor Singer. (2004). *Methods for Testing and Evaluating Survey Questionnaires*. New York: Wiley.
Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological test. *Psychological Bulletin*, *52*, 281–302.
DeVellis, R. F. (2012). *Scale development. Theory and applications* (3rd ed.). Thousand Oaks, CA: Sage.
Drennan, J. (2003). Cognitive interviewing: Verbal data in the design and pretesting of questionnaires. *Journal of Advancing Nursing*, *42*(1), 57–63.
Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272–299.
Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*(1), 286–299.
Ford, J. K., MacCullum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, *39*, 291–314.
Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, *103*, 265–275.
Goldberg, L. R., & Velicer, W. F. (2006). Principles of exploratory factor analysis. In S. S. Strack (Ed.), *Differentiating normal and abnormal personality* (pp. 209–237). New York, NY, USA: Springer.
Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment*, *68*, 532–560.
Haig, B. D. (2005). Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behavioral Research*, *40*(3), 303–329.
Hair, J., Jr., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
Hendrickson, A. E., & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *British Journal of Mathematical Psychology*, *17*, 65–70.
Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research. Common errors and some comment on improved practice. *Education and Psychological Measurement*, *66*(3), 393–416.
Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185.
Humphreys, L. G., & Montanelli, R. G., Jr. (1975). An investigation of the parallel criterion for determining the number of common factors. *Multivariate Behavioral Research*, *10*, 193–205.
Jackson, D. L., Gillaspy, J. A., Jr., & Pure-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, *14*(1), 6–23.
Kachigan, S. K. (1986). *Statistical analysis: An interdisciplinary introduction to univariate and multivariate methods*. New York: Radius Press.
Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141–151.
Kaiser, H. F. (1970). A second generation Little Jiffy. *Psychometrika*, *35*(4), 401–415.
Kamhawi, R. (2003). Mass communication research trends from 1980 to 1999. *Journalism & Mass Communication Quarterly*, *80*(1), 7–27.
Ki, E., & Shin, J. (2006). Status of organization-public relationship research from an analysis of published articles, 1985-2004. *Public Relations Review*, *32*, 194–195.
Kline, R. B. (2013). Exploratory and confirmatory factor analysis. In Y. Petscher, C. Schatschneider, & D. L. Compton (Eds.), *Applied quantitative analysis education and the social sciences* (pp. 171–207). New York, NY, USA: Routledge.

Levine, T. R., Hullett, C. R., Turner, M. M., & Lapinski, M. K. (2006). The deirability of using confirmatory factor analysis on published scales. *Communication Research Reports*, *23*, 309–314.
Lewis, B. R., Templeton, G. F., & Byrd, T. A. (2005). A methodology for construct development in MIS research. *European Journal of Information Systems*, *14*(4), 388–400.
Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*(7), 635–694.
McCrosky, J. C., & Young, T. J. (1979). The use and abuse of factor analysis in communication research. *Human Communication Research*, *5*, 375–82.
Mertler, C. A., & Vannatta, R. A. (2001). *Advanced and multivariate statistical methods: Practical applications and interpretation*. Los Angeles, CA: Pyrczak Publishing.
Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied multivariate research: Design and interpretation*. Thousand Oaks, CA: Sage.
Morrison, J. T. (2009). Evaluating factor analysis decisions for scale design in communication research. *Communication Methods and Measures*, *3*(4), 195–215.
Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, *5*, 159–168.
Myers, T. A. (2011). Goodbye, listwise deletion: Presenting hot deck imputation as an easy and effective tool for handling missing data. *Communication Methods and Measures*, *5*(4), 297–310.
Norris, M., & Lecavalier, L. (2010). Evaluating the use of exploratory factor analysis in developmental disability psychological research. *Journal of Autism Development and Disorders*, *40*, 8–20.
Nunnally, J. C., & Bernstein, I. (1994). *Psychometric theory* (3rd ed.). New York, NY, USA: McGraw Hill.
O'Conner, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Veliver's MAP test. *Behavior Research Methods, Instruments, & Computers*, *32*(3), 396–402.
Osborne, J. W. (2014). *Best practices in exploratory factor analysis*. Scotts Valley, CA: CreateSpace Independent Publishing.
Pallant, J. (2010). *SPSS: Survival manual*. New York, NY, USA: McGraw Hill.
Park, H. S., Dailey, R., & Lemus, D. (2002). The use of exploratory factor analysis and principal components analysis in communication research. *Human Communication Research*, *28*(4), 562–577.
Pett, M. A., Lackey, N. R., & Sullivan, J. L. (2003). *Making sense of factor analysis. The use of factor analysis for instrument development in health care research*. Thousand Oaks, CA: Sage Publications, Inc.
Potter, W. J., & Riddle, K. (2007). A content analysis of the media effects literature. *Journalism & Mass Communication Quarterly*, *84*(1), 90–104.
Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, *2*(1), 13–43.
Reinard, J. C. (2006). Communication research statistics. Sage Publications, Thousand Oaks, CA.
Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment*, *12*(3), 287–297.
Reynolds, N., Diamantopoulos, A., & Schlegelmilch, B. (1993). Pretesting in questionnaire design: A review of the literature and suggestions for further research. *Journal of Market Research Society*, *35*(2), 171–182.
Riffe, D., Lacy, S., & Fico, F. G. (2005). *Analyzing media messages. Using quantitative content analysis in research*. Mahwah, NJ: Lawrence Erlbaum Associates.
Riffe, D., Lacy, S., & Fico, F. G. (2014). *Analyzing media messages. Using quantitative content analysis in research*. Mahwah, NJ. : Lawrence Erlbaum Associates.
Ruel, E. E., Wagner, W. E., III, & Gillespie, B. J. (2016). *The practice of survey research*. Thousand Oaks, CA: Sage.
Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor Factor analysis and scale revision. *Psychological Assessment*, *12*(3), 1629–1646.
Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory analysis results: A review. *The Journal of Educational Research*, *99*(6), 323–337.
Simms, L. J. (2008). Classical and modern of psychological scale construction. *Social and Psychology Compass*, *2*(1), 414–433.
Snook, S. C., & Gorsuch, R. L. (1989). Common factor analysis vs. component analysis. *Psychological Bulletin*, *106*, 148–154.
Streiner, D. L. (1994). Figuring out factors: The use and misuse of factor analysis. *Canadian Journal of Psychiatry*, *39*, 135–140.
Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics* (5th ed.). New York: Allyn and Bacon.
Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn and Bacon.
Thompson, B. (2004). *Exploratory and confirmatory analysis: Understanding concepts and applications*. Washington, DC, USA: American Psychological Association.
Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago, IL, USA: University of Chicago Press.
Tinsley, H. E. A., & Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. *Journal of Counseling Psychology*, *34*(4), 414–424.

Trumbo, C. W. (2004). Research methods in mass communication research: A census of eight journals 1990-2000. *Journalism & Mass Communication Quarterly*, *81*(2), 417–436.
Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321–327.
Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin, & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas Jackson at seventy* (pp. 41–71). Boston, MA: Kluwer.
Viswanathan, M. (2010). Understanding the intangibles of measurement in the social sciences. In G. Walford, E. Tucker, & M. Viswanathan (Eds.), *The SAGE handbook of measurement* (pp. 285–313). London, UK: Sage.
Watkins, M. W. (2006). Determining parallel analysis criteria. *Journal of Modern Applied Statistical Methods*, *5*(2), 344–346.
Wimmer, R. D., & Haynes, R. B. (1978). Statistical analyses in the *Journal of Broadcasting*, 1970-76. *Journal of Broadcasting*, *22*(2), 241–248.
Worthington, R. L., & Whittaker, T. A. (2006). Scale development research. A content analysis for recommendations for best practices. *The Counseling Psychologist*, *34*(6), 806–838.
Ye, L., & Ki, E. (2012). The status of online public relations research: An analysis of published articles in 1992-2009. *Journal of Public Relations Research*, *24*(5), 409–434.
Zwick, W. R., & Velicer, W. F. (1982). Factors influencing four rules determining the number of components to retain. *Multivariate Behavioral Research*, *17*, 253–269.