UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

The credibility of credibility measures: a meta-analysis in leading communication journals, 1951 to 2011

Hellmueller, L.; Trilling, D.

Publication date
2012
Document Version
Final published version
Published in
WAPOR Hong Kong 2012: paper presentation

Link to publication

Citation for published version (APA):
Hellmueller, L., & Trilling, D. (2012). The credibility of credibility measures: a meta-analysis in leading communication journals, 1951 to 2011. In *WAPOR Hong Kong 2012: paper presentation* World Association for Public Opinion Research/Public Opinion Programme, The University of Hong Kong.
http://wapor2012.hkpop.hk/doc/papers/ConcurrentSessionsV/VD/VD-3.pdf

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:08 Feb 2025

The Credibility of Credibility Measures: A Meta-Analysis in Leading Communication Journals, 1951 to 2011

Lea Hellmueller
Missouri School of Journalism /
University of Fribourg, Switzerland
Visiting Scholar
246 Walter Williams Hall
Columbia, MO 65211, USA
+1-8578696380
lea.hellmueller@fulbrightmail.org

Damian Trilling
Amsterdam School of Communication Research
University of Amsterdam
Kloveniersburgwal 48
1012CX Amsterdam
Netherlands
+31-205252161
d.c.trilling@uva.nl

Paper presented at the WAPOR 65th Annual Conference in Hong Kong, June, 2012

The Credibility of Credibility Measures: A Meta-Analysis in Leading Communication Journals, 1951 to 2011

Abstract

The construct of credibility has received more scholarly attention than most other communication variables. In a systematic meta-analysis of leading international communication journals (1951–2011), we examined how source, message, and media credibility are conceptualized and measured. Results suggest that various scales to measure credibility reveal inconsistency in theoretical reflection, lack operational precision, and led to insufficient replication and validation of the credibility construct. The findings provide a critical examination of 60 years of credibility research in media and communication studies.

Introduction

Over 60 years ago, Hovland and Weiss (1951) enriched communication science with a path breaking study revealing that the effectiveness of communication is largely determined by the credibility of its source. Few constructs[1] have received more scholarly attention than the issue of credibility (Kiousis, 2001). But while credibility seems to be one of the major variables in the process of communication, credibility studies have also received scholarly attention because of their lack of operational precision, the use of different items to measure the same construct, and insufficient reporting of reliability. Metzger et al. (2003) state that "disagreements about the relative importance of the dimensions of credibility led to the construction of various scales to measure this concept, each reflecting the priority of dimensions identified by particular researchers" (p. 298).

Disagreements derive from missing definitions of the credibility construct in empirical research: "The definition problem illustrated by the credibility research is a serious one, because many later decisions in factor analytic research must be based on the

[1] The term "construct" is chosen purposively to account for the level of abstraction of credibility, with construct being more abstract or general than concepts (Shoemaker, Tankard, & Lasorsa, 2004).

definition of the construct being studied" (McCroskey & Young, 1979, p. 376). McCroskey and Young argue that if the construct is not carefully defined, the product of research may yield little value because a need for isomorphism between the definition, operationalization, and measurement is a basic requirement of measurement.

Another criticism are missing details of an analysis such as how the decision for the number of factors or the type of rotation was made. Thus, methodological problems then prevent replication or modification of the results (Kohring & Matthes, 2005; Kohring & Matthes, 2007; McCroskey & Young, 1979). Despite the raised criticisms, credibility is put forward as a strong predictor of news media effects (Kohring & Matthes, 2007), as a critical factor in the selection of opinion leaders (Teven, 2008), and as the communicator's most potent means of persuasion (Höffe, 2005).

After 60 years of credibility research, it seems relevant to look back in order to move forward: The empirical richness of credibility studies has certainly come with a modest discussion on conceptualization and measurement issue. Therefore, this investigation aims to shed light on how to enhance our understanding of the reliability and validity of the credibility construct by conducting a meta-analysis of credibility research studies in leading international communication journals. The study that follows is a critical reflection and empirical investigation of credibility research—1951 to 2011. The present study examines operationalizations of credibility and methodologies of measuring source, message, or media credibility in a communication context. In addition, the present study presents an empirical overview of credibility to hopefully serve to maintain isomorphism of the construct in future research.

Literature Review

Credibility and Trust

Credibility is often used interchangeably with other constructs, particularly with the construct of trust. Some researchers use the two terms, credibility and trust, almost interchangeably. For example, Tsfati and Cappella (2003) measure trust, but rely on media credibility research. Kohring and Matthes (2007) advocate for trust over credibility to embed the construct in "theoretical concepts of society that transcend the simple identification of an information society" (p. 238). Following their argument, theories of trust refer to recipients' assessment on how societal function of selecting and conveying information about interdependencies of modern societies are performed by the news media. Whereas trust research heavily depends on media's function in society (on a rather mezzo or macro level), credibility research relies more on interpersonal factors (i.e., source credibility on a rather micro level), on characteristics of the message source, or on characteristics of the medium through which the message is delivered. Hence, the distinction is drawn between sociology based trust research that takes into consideration *societal functions* of media as basis for a trust assessment and credibility research as an dimension for *evaluating message sources* (McCroskey & Young, 1979, p. 376), Without doubt, both terms enjoy high scholarly attention, credibility more so in the United States. While credibility research in a German tradition has been influenced by sociological ideas, particularly related to system theory, such an approach never established itself in the US (Luhmann, 1990): "Germans who accept systems theory as their research program meet astonished looks if they dare to enter the United States – as if they were not quite au courant with present sociology" (p. 225). It seems not surprising then that the *American*

Behavioral Scientist has recently published an issue on credibility research in 2010 (Golan, 2010), whereas the *European Journal of Communication* put together a special issue on trust in 2012 (Golding, Sousa, & van Zoonen, 2012).

We argue that credibility and trust bear conceptual overlaps; they are indeed interrelated, but for the sake of conceptual clarity and reliability of our measures, we analytically distinguish between the two and hence decided to focus on credibility solely for the here presented meta-analysis.

Source Credibility

Credibility research can be grouped into three different approaches: source credibility, media credibility, and message credibility (Metzger, et al., 2003). We briefly discuss all of them, starting with source credibility.

The path-breaking study by Hovland and his associates (1951) set the framework for credibility research, particularly for source credibility studies. The Yale studies program on attitude change was based on the former psychological studies of "prestige" (e.g. Arnett, Davidson, & Lewis, 1931) in which it was concluded that "the extent of agreement is usually higher when the statements are attributed to 'high prestige' sources" (Hovland & Weiss, 1951, p. 635). They adapted this basic idea and argued that the source (by source they meant single persons, groups, organisations or media) is a multidimensional predictor factor of communication effectiveness besides the content and the medium of a message. "Opinions were changed immediately after the communication in the direction advocated by the communicator to a significantly greater degree when the material was presented by a trustworthy source than when presented by

an untrustworthy source" (Hovland & Weiss, 1951, p. 650). The immediate reaction to the "fairness" of the presentation and the "justifiability" of the conclusion is significantly affected by the evaluation of the trustworthiness of the source. Credibility became to be considered a universal characteristic of a general communication source, based on the two dimensions of credibility, i.e., expertness ("how well informed and intelligent") and trustworthiness (Hovland, Janis, & Kelley, 1959) .

Obviously, their concept is too simplistic from today's point of view. Several variables of the message receiver were not taken into account. Cronkhite and Liska (1976) noted that the credibility of sources depends heavily upon the specific function they perform in specific topic situations for specific listeners. Moreover, King (1976) captures a perspective that posits the varying relevance of source attributes, as a function of the receiver's decision needs and message topic. Looking at source credibility, one does not only have to emphasize on the perceived attributes, but also on the communication needs of the receiver or the potential function served by sources. Since the 1970s, we can observe an increasing precision in measurement from the seminal Yale studies to the factor analytical approach. A movement followed from one dimensional to multidimensional measures and from applying items of source credibility to applying specific items of media credibility (Kohring & Matthes, 2007, p. 237). The problem of the explorative factor analysis includes that non-credibility-based dimensions were interpreted as dimensions of the construct (Kohring & Matthes, 2007, p. 235). For example, dimensions like attraction or objectivity have interdependency with credibility, but—on theoretical grounds—are not part of the construct itself. This concern resonates with the isomorphism claim explained in the introduction. McCroskey and Young (1979)

argue convincingly that the key for identifying the variables for factor analytic research is "the maintenance of isomorphism between the construct (as defined) and the measurement" (p. 376). Hence, if variables like attraction are not examined for its isomorphism with its theoretical definition then the results is very likely to be the extraction of a dimension that does not really exist and "conclusions about the originally defined construct are negated" (p. 376). For example, we then no longer predict our dependent variable with the construct (e.g., source credibility) we are assuming to measure but with interrelated variables (e.g., attraction) that are not part of the construct itself. It seems important that neglecting rigid isomorphism weakens our measurement.

On the other hand, McCroskey & Teven (1999) examined the body of literature on source credibility and argue that one dimension—goodwill—of the construct has been 'lost'. The authors argue that Aristotle's conceptualization of ethos/source credibility was based on three dimensions: competence, trustworthiness, and goodwill (p. 90). The authors refer to Aristotle's conceptualization as it serves as milestone in rhetorical communication theory.

In fact, their exhaustive empirical investigation with a variation of sources, ranging from public figures such as Madonna, political figures such as Newt Gingrich, to the partner you dated last (interpersonal source) reveal that the amount of variance attributed to the goodwill and trustworthiness factors were more predictive than competence. Interestingly, the study finds notable justification to treat credibility as three-dimensional rather than unidimensional. In fact, the three factors together accounted for approximately 61 percent, whereas a one factor unrotated solution accounted for 51 percent only for the variance in the respondents' perception of

believability. Hence, McCroskey & Teven (1999) argue that goodwill (i.e., perceived caring) is indeed a meaningful predictor of believability and likeableness and should take its place in the conceptual and operational future of communication research dealing with ethos and source credibility as a *multidimensional construct.*

Message Credibility

As outlined above, the credibility of a message is influenced by its source. Another line of research has shifted the focus from investigating the effect of different sources (communicating the same message) to the effect of different messages communicated by the same source.

Message credibility was explored by focusing on characteristics of messages that could make them more credible. The research on the structure of messages has demonstrated consistent results: unorganized messages are perceived less credible than well-organized messages. Message organization has further been shown to affect perceptions of the source (Hamilton, 1998; McCroskey & Mehrley, 1969). Credibility judgements are influenced by message content factors such as information quality, language intensity, and message discrepancy. The tendency today is to turn to the information quality literature to develop scales assessing evaluations of a message's accuracy, comprehensiveness, currency, reliability, and validity (Metzger et al., 2003).

Media Credibility

Media credibility studies, in contrast to message credibility studies, look at rather mezzo- or macro-level instead of the micro-level. The credibility of each message, the

argument goes, is directly influenced by the medium in which it appears. Accordingly, media researchers measured credibility to explain the public's use of various news media outlets. Mass communication scholars looked at media credibility in determining the relative believability of forms of communication (newspaper, radio, television). This line of research can be traced back to Roper's seminal comparative media credibility studies. Every two years, a survey asked U.S. audiences in which version they would most likely believe if they receive conflicting reports covering the same stories. Audience were asked to rate whether they would most likely believe the story in radio, on television, in magazines or in newspapers (Roper, 1985). Methodological problems include reliability of measurement and a measure on one single item (Kohring & Matthes, 2007).

The debate on which of these media receives the highest medium credibility seemed to be settled towards the end of the last century. But with the advent of the Internet, media credibility research received a renewed interest. Scholars were interested how credible online news sites, blogs, or 'the Internet' are compared to offline media (Stavenuiter, Trilling, & Bakker, 2012). These studies showed inconclusive results, and, in fact, illustrate a methodological problem of the media credibility concept: It assumes, that credibility differences within one media type are smaller than between media types. However, considering media pluralism in terms of external versus internal pluralism (i.e., pluralism that exist across media outlets with low pluralism among one medium vs. pluralism within one medium in terms of content, but less pluralism across media), it seems questionable whether the major credibility differences exist between different media types. Hence, it is premature to argue that a quality newspaper is roughly as credible as a tabloid, but differs from it's own website in terms of credibility.

There were few attempts to build on previous research to validate media credibility scales. One exception is Meyer's (1988) study whose main concern was to validate existing scales (see Table 1). He validated the Gaziano-McGrath scales in which a different factor structure emerged. In a cross-validation study of these credibility scales, the Meyer 5-item credibility index was found to be reliable (Cronbach's alpha .83) and empirically valid (West, 1994). West (1994) concluded that the Meyer modification of the Gaziano-McGrath index would serve well as a standard for future credibility research, because through the development, validation, and cross-validation. However, Kohring and Matthes (2007) point out that the reported "Goodness of Fit Indices (GFIs) of .87 and .85 indicate that the theoretical model does not fit the data" (p. 237). Furthermore, a missing theoretical foundation of the study has to be critically noted.

<< TABLE 1 ABOUT HERE >>

On other occasions, the construct of source credibility and media credibility are studied interchangeably. For example a study by Yoon (2004) that examines journalists' perceptions of stem cell and cloning organizations measured source credibility with Meyer's (1988) credibility scale. As stated above, the scale was originally conceptualized to measure the credibility of media. Hence, items like "bias" may not be valid items to measure source credibility of organization, as journalists in fact expect organizations to be biased, because companies often communicate in their interests as supporting their organizational norm of preserving the company's reputation. Here again, bias refers to a media function. Objectivity has long been the dominant norm in journalism (Schudson, 2001) and being biased seems contradictory to neutrality, factualness, or balance claims

of objectivity. On the other hand, objectivity or the absent of bias does not necessarily define an organization's function in the same way as it defines the journalism profession.

In essence, there exists a wide range of credibility research studies each contributing to our overall understanding of the construct. However, we badly need a systematic analysis to better understand how theoretical definitions are linked to their corresponding empirical operationalization.

Research Questions

As outlined above, an empirical richness of credibility studies has come with a modest discussion on conceptualization and measurement issues. We aim to fill this gap by identifying what it is that we are actually measuring when we explicate the construct of credibility. Hence, we empirically test what conceptual definitions, operationalized items, and methods have been applied to examine the three types of credibility.

We pose the following three research questions:

RQ1: How is credibility defined and operationalized?

RQ2: Which methods and items are applied to measure the credibility construct?

RQ3: Which factors are identified as part of the credibility construct (one-dimensional vs. multidimensional)? How do they relate to the theoretical conceptualization of source, message, and media credibility?

Method

A quantitative content analysis was conducted to analyze how credibility is measured in leading communication journals. Journals were selected according to three

main criteria inspired by a meta-analysis on framing studies (Matthes, 2009): (1) must have a focus on mediated communication contexts, (2) must include a variety of credibility studies such as message, source, and medium credibility, and (3) must have a quantitative approach to enhance the comparability of the articles collected. Communication journals included *American Behavioral Scientist (n=7)*[2]*, Asian Journal of Communication (n=1), China Media Research (n=1), Communication Research (n=10), Communication Research Reports (n=1), European Journal of Communication (n=2), Human Communication (n=1), International Communication Gazette (n=9), Journal of Advertising (n=1), Journal of Communication (n=12), Journal of Computer-Mediated Communication (n=7), Journalism Practice (n=1) Journalism and Mass Communication Quarterly (n=18), Mass Communication & Society (n=7).*

Within these journals, the articles to include in the analysis were identified by searching for 'credibility' in title and/or abstract. Articles that turned out to be no empirical studies on credibility were removed from the sample. Because credibility research has been around for over 60 years (i.e., Matthes, 2009) all articles were included and no time frame was set.

Applying the criteria mentioned above, we identified 96 articles based on their abstracts. After reading the full articles, 22 articles were removed as they turned out to not meet all of the requirements, mainly because they did not measure credibility quantitatively. Therefore, our final sample consists of *N*=75 articles. With 18 out of 75 articles published in Journalism and Mass Communication Quarterly, the journal seems to be the platform where credibility debates preferably take place.

[2] Numbers of articles included in the final sample after excluding articles that did not meet all criteria (see below).

Coding and Analysis

For each article of the final sample, the following variables were coded:

- *Credibility concept.* Categories included source, message, and medium credibility. If studies explicitly named another concept (e.g., "opinion poll credibility"), this concept was coded.
- *Definition of credibility.* We aimed to code to which definition of credibility the studies referred. However, it turned out that hardly any study provided an explicit definition.
- *Method.* We distinguished longitudinal surveys, cross-sectional surveys, experiments, and content analyses.
- *Measurement*: We recorded whether a scale or some other type of measurement (like a ranking task) was used, how many points the scale had, and if semantic differentials, agreement or, some other kind of scale was used.

Only if a scale was used, the following items were coded as well:

- *Origin of Scale.* Does the study replicate an earlier used scale, and if so, which?
- *Items.* Which items are used?
- *Reliability.* Cronbach's alpha of the scale.

Based on these variables, we created two datasets: One with the article as the unit of analysis, and a second one with each scale used as unit of analysis. Most articles used one scale to measure credibility. However, n=6 articles used two different scales, n=2 articles even three. For the analysis of items used to form a scale, however, it is necessary

to keep different scales apart. Therefore, the second dataset included each scale rather than each article as a unit of analysis.

We analyzed the data by first comparing descriptive statistics and performing a Ward's linkage cluster analysis of scale items. Because a full census of articles was conducted, test for statistical significance were not necessary.

Results

As expected, results reveal three major credibility approaches: *message credibility, source credibility and media credibility*. As pointed out above, credibility has gained an increasing amount of attention in the last decade. This is reflected in our sample as well. We can distinguish one phase of fundamental research in which the construct of credibility was introduced into communication research in the 1970s. But only after the 2000s, the construct was widely used.

<< FIGURE 1 ABOUT HERE >>

Definition of Credibility

RQ1 asked: *How is credibility defined and operationalized?* By far most articles did not provide an explicit definition of credibility—which underlines the deplorable fuzziness of the construct in today's state of the research and neglects isomorphism claims, which reduces the value of empirical findings. However, most articles related to one of the three main concepts of credibility discussed above: *media credibility* (n=32, 43%), *source credibility* (n=33, 44%), and *message credibility* (n=8, 11%). Eight articles

(11%) dealt with other credibility constructs. As some articles investigated more than one concept, percentages do not add up to 100%.

Attention to different constructs of credibility changed over time (Figure 2). Due to the advent of the Internet in the 1990s, scholars addressed the question how credible "the Internet" was as opposed to newspapers or television – reflected in the high share of articles on *media credibility*. In the new millennium scholars increasingly turned to more refined questions and credibility on the *source* and *message* level gained attention.

<< FIGURE 2 ABOUT HERE >>

Methods

RQ2 asked: *Which methods and items are used to investigate credibility?*

Our findings show that the majority of studies rely on either surveys (n=38) or experiments (n=33). Rarely, content analysis data is used as well (n=3), usually as a supplement. Only one panel survey was conducted. While both surveys (M=1120, SD=1327) and experiments (M=258, SD=179) on average use rather big samples, the distribution is skewed and sample sizes differ greatly (Figures 3 and 4). This is because a considerable amount of the studies works with convenience samples, reflected in the modal survey sample size category of 250-500. In other words, especially large-scale representative surveys are still comparably scarce.

<< FIGURE 3 ABOUT HERE >>

<< FIGURE 4 ABOUT HERE >>

With very few exceptions, studies on media credibility used survey data (only 3 used experiments), while the vast majority of studies on source credibility relied on experiments. Only five source credibility studies rely on survey data, with samples between 69 and 575, M=353, SD=214). Thus, there is no single source study on source credibility with a sample big enough to be representative for a whole country, for example.

Items

Despite the increasing attention credibility research has gained among researchers, measurement of the variable differs widely. Instruments range from straightforward one-item measures to semantic differentials with as many as 55 items. While many studies refer to the scale developed by Gaziano & McGrath (1986) or other scales, items are often adapted and used selectively. Out of the 85 scales used in total in all studies, only 16 scales are replications of earlier used scales. Clearly, the field has not agreed on one instrument yet.

Three studies used ranking tasks instead of scales, and five other studies did not report the items used. These cases therefore were excluded from further analysis of the items.

It is not uncommon to directly ask respondents how credible they think something is—nine scales, out of which five media-credibility scales, applied such a one-item measure. However, most commonly used are scales of 4 or 5 items. Eight out of 23 source credibility scales use 12 items or more, media credibility is never measured with that many items.

To systematize the items used in the studies, we conducted a Ward's linkage cluster analysis of the items. Each scale used in the dataset was entered into the analysis as one case. The Duda-Hart criterion, the dendrogram, and the interpretability of competing solutions all suggested a three-cluster solution. As Table 2 shows, a wide range of different items is used in very different frequencies. The measures can broadly be grouped into Meyer-based scales (cluster 1), Johnson-Kaye-based scales (cluster 2), and own scales (cluster 3).

<< TABLE 2 ABOUT HERE>>

It is interesting to note how many studies use an enormous range of items that are used only by one or two studies. A closer look at these items shows that among them, there are many for which it is at least questionable if they conceptually can be part of credibility. For example, items like 'detailed' or 'well-written' may be used by the audience as a heuristic to evaluate credibility, but one can hardly argue that these items constitute credibility: They are rather a predictor of perceived credibility than part of it. To put it another way, items like "it provides me with useful information" are rather measuring an outcome of credibility than credibility itself.

Linking constructs and measurement

RQ3 asked*: How are conceptualizations of credibility reflected in the decision of which items are used to measure credibility?* In line with the literature review, we proceed to present the findings according to their relevance for source, message, or media credibility. Studies on source credibility never used a Johnson/Kaye-cluster2-scale.

Instead, three studies used a Meyer-cluster1-scale and 15 used own scales. Such an avoidance of cluster2-scale is a sign of consistency in measurement. However, the large amount of other scales reflects inconsistency about a theoretical definition of source credibility.

Looking at the items (Table 3), we see that 'trustworthiness' and 'expertness' are characteristic items to measure source credibility. They are much less common to measure other types of credibility. This makes sense, because if source credibility is defined as 'evaluating a message source' such as a public or political figure then items such as 'in-depth' or 'accurate' might not be valid measures of how credible such a person is because such items are referring to evaluating a function of media as an institution. In light with this argument, it seems crucial to not only define source credibility but also explicitly define what we refer to when we conceptualize 'source'. The Yale Group Communication Research Program focused on the influence of individual persons, institutions as well as magazines, which they defined as "sources". On the other hand, McCroskey and his associates defined "sources" as communicators, individuals and tested the perceived credibility of those speakers (e.g., Arnold & McCroskey, 1967; McCroskey & Teven, 1999). It remains clear that the same measurement cannot serve for different operationalization of sources. Evaluating a magazine is not an interpersonal communication situation.

<< TABLE 3 ABOUT HERE>>

In contrast to measures of source credibility, the items to measure media credibility are more diverse. In fact, studies on media-cred use all types of scales: We found six scales belonging to Cluster 1, nine scales belonging to Cluster2, and thirteen scales belonging to Cluster 3. This diversity can be seen as a sign of inconsistency between conceptualization and measurement.

On the item level (Table 3), the main difference between the measurement of source and media credibility is that the latter is remarkably often measured with an item tapping into *accuracy* and *fairness*, which is not the case for source credibility, as explained above.

Hence, even if most studies lack an explicit definition, and even if differences are far from clear-cut, some agreement seems to exist on their operational definition of the used constructs.

Message credibility was investigated with two Cluster 1 scales and three Cluster 3-scales. The overwhelming majority of studies on other credibility concepts used on scales (12), only three scales belonged to cluster 1 and one to cluster 2.

Reliability and Internal Consistency of the Scales

Although, as we showed, no universally applied scale seems to exist, most studies report good reliability scores – except two studies with reliabilities below .70 and fourteen studies that failed to report reliabilities. However, in a substantive number of cases, reliabilities are as high as about .95. This is not as positive as it may seem: As Streiner (2003) points out, this is a sign that some items are redundant.

Clark and Watson (1995) argue similarly and point out that, Cronbach's alpha is a measure of internal consistency rather than homogeneity. In other words, its measure is

of limited utility in establishing the unidimensionality of a scale (Clark & Watson, 1995). In accordance with Streiner (2003) the authors state that the problem that occurs is that few items correlating most strongly with the assessed or latent construct may be highly redundant with one another and thus are measuring the same thing and may create an overly narrow scale that likely will not assess the construct optimally. Clark & Watson (1995) propose to report the average inter-item correlation, which should fall in the range of r=.15-.5. In light of this paradox, it becomes clear that the goal of scale construction is to maximize validity rather than reliability.

Discussion and Conclusion

With an increasing amount of information available, selection criteria such as credibility become immensely important. How credible a person or a media organization is has frequently been discussed among scholar and practitioners alike. In stark contrast to the increasing relevance of credibility, our meta-analysis reveals that discussion on theoretical and operational definitions of the credibility construct have been modest at most. Our findings point to inconsistencies between theoretical decisions and measurements. For example, source, media, and message credibility are identified as three different constructs. In other words, in 44 percent of our analyzed cases, scholars report to measure source credibility, in 43 percent media credibility is measured, whereas in 11 percent of all studies message credibility is used to indicate what is measured. We interpret this as an assumed conceptual difference between media, source, and message credibility. However, operational definitions do not resonate with the three distinctions

made in the conceptualizations. The cluster analysis illustrates disagreement on the corresponding operational definition and measurement.

We follow from this discussion that implicit theoretical definitions such as defining credibility through its operationalized definition (i.e., the measured items) may vanish the validity of our measures and limit an in-depth comparison of our findings because the results are not comparable as they are based on diverging operational definitions. The validity of the measure especially is confounded by the inclusion of items that are outcomes, predictors or other correlates of credibility, but do not constitute the construct of credibility. As our analysis shows, this happens on a regular basis.

Furthermore, the process of choosing items should follow an attempt to maintain isomorphism between construct as it is defined and its corresponding measurement. If for example "source credibility" is defined as an "evaluation of a message source", then only evaluative scales should be chosen for use otherwise conclusions are negated (McCroskey & Young, 1979). Without doubt, the theoretical linkage to the operationalization should be an examination made prior to its application. However, this rule was not always followed consequently, as items like "it provides me with information I can use" clearly is no evaluation of the source itself. The huge amount of items that are used only once also is an indicator of the lack of theoretical reflection.

The same is true for the distinction between source credibility, media credibility, and message credibility: Many studies did not explicitly link their credibility concepts to one of these, but stuck to narrower, own concepts that impede linking their results to other studies. And although theoretically distinct, some studies used media credibility scales to measure source credibility, and vice versa.

A majority of studies have pointed to the multidimensionality of the credibility construct (Matthes & Kohring, 2009; McCroskey & Teven, 1999). Hence, when researcher are faced in making a decision on how many factors to extract, the multidimensionality leads to a decision of more than one factor.

Our concern was to set out the empirical differences between different forms of credibility. It revealed that in spite of the large amount of studies our understanding of different types of credibility is limited: For example, while source credibility turned out to be studied with experiments, media credibility research mostly relied on large-scale surveys. But as both methods serve different ends, the one allowing for causal inferences, the other one allowing quantifying media effects in a representative sample, it would be fruitful to also study media credibility in an experiment and source credibility in a large-scale survey. Our study serves practical claims to some extend: The Project for Exellence in Journalism has reported repeatedly that the credibility ratings of traditional mass media organization has gradually declined since the mid-1980s. For example, in a Pew Research Center survey (2009) 63% of Americans agreed that news stories are inaccurate and not credible whereas 60% perceived news organizations as politically biased, which marks the worst ratings since 1985. Valid and reliable measures will hopefully crystallize dimensions that may provoke more in-depth discussions on how to reinvest in the credibility of the news media.

On a number of media variables, accepted scales exist. We suggest to develop three news scales – for media, source, and message credibility – based on here pointed out substantive arguments that reflect the theoretical differences between these constructs. These scales should be universally applicable—or extended to various cultural

settings— to avoid the proliferation of questionable items and scales that currently dominate credibility research.

References

Arnett, C. E., Davidson, H. H., & Lewis, H. N. (1931). Prestige as a factor in attitude change. *Sociology and Social Research, 16*, 49-55.

Clark, L. A., & Watson, D. (1995). Constructing validity: basic issues in objective scale development. *Psychological Assesment, 7*, 309-319.

Cronkhite, G., & Liska, J. (1976). A critique of factor analytic approaches to the study of credibility. *Communication Monographs, 43*(2), 91-107.

Golan, G. J. (2010). New Perspectives on Media Credibility Research (Vol. 54, pp. 3-7).

Golding, P., Sousa, H., & van Zoonen, L. (2012). Trust and the media. [Proceeding]. *European Journal of Communication, 27*(3), 3-6.

Hamilton, M. A. (1998). Messagge variables that mediate and moderate the effect of equivocal language on source credibility. *Journal of Language and Social Psychology, 17*, 109-143.

Höffe, O. (2005). *Aristoteles* ((3rd ed.) ed.). München: Verlag C.H. Beck oHG.

Hovland, C. I., Janis, I. L., & Kelley, H. (1959). *Communication and persuasion. Psychological studies of opinion change* (3 ed.). New Haven, CT: Yale University Press.

Hovland, C. I., & Weiss, W. (1951). The Influence of Source Credibility on Communication Effectiveness. *Public Opin Q, 15*(4), 635-650.

Kiousis, S. (2001). Public Trust or Mistrust? Perceptions of Media Credibility in the Information Age. *Mass Communication & Society, 4*(4), 381-403.

Kohring, M., & Matthes, J. (2005). *Trust in News Media. Development and Validation of a Multidimensional Scale*. Paper presented at the International Communication Association Conference. Retrieved from http://mcc1.library.csulb.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=ufh&AN=18655970&site=ehost-live

Kohring, M., & Matthes, J. (2007). Trust in News Media. *Communication Research, 34*(2), 231-252.

Luhmann, N. (1990). General theory and american sociology. In H. J. Gans (Ed.), *Sociology in America* (pp. 253-264). Newbury Park, CA, London & New Delhi: Sage.

Matthes, J. (2009). What's in frame? a content analysis of media framing studies in the world's leading communication journals, 1990-2005 *Journalism & Mass Communication Quarterly, 86*(2), 349-367.

McCroskey, J. C., & Mehrley, R. S. (1969). The effects of disorganization and nonfluency on attitude change and source credibility. *Speech Monographs, 36*, 13-21.

McCroskey, J. C., & Young, T. J. (1979). The use and abuse of factor analysis in communication research. *Human Communication Research, 5*(4), 375-382.

Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D. R., & McCann, R. M. (2003). Credibility for the 21st century: integrating perspectives on source, message, and

media credibility in the contemporary media environment. In P. J. Kalbfleisch (Ed.), *Communication Yearbook 27*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Meyer, P. (1988). Defining and measuring credibility of newspapers: Developing an index. *Journalism Quarterly, 65*, 567–574.

Roper, B. W. (1985). *Public attitudes toward television and other media in a time of change*. New York: Television Information Office.

Schudson, M. (2001). The objectivity norm in American journalism. *Journalism, 2*(2), 149-170.

Shoemaker, P. J., Tankard, J. W., & Lasorsa, D. L. (2004). *How to build social science theories*. Thousand Oaks, CA: Sage.

Stavenuiter, R., Trilling, D., & Bakker, T. P. (2012). *The same old song? Reviewing factors that predict credibility of offline and online media*. Paper presented at the Paper presented at the Etmaal van de Communicatiewetenschap (February, 2012).

Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*, 99-103.

Teven, J. J. (2008). An Examination of Perceived Credibility of the 2008 Presidential Candidates: Relationships with Believability, Likeability, and Deceptiveness. *Human Communication, 11*(4), 391-407.

Tsfati, Y., & Cappella, J. N. (2003). Do people watch what they do not trust? Exloring the association between news media skepticism and exposure. *Communication Research, 30*, 504-529.

West, M. D. (1994). Validating a scale for the measurement of credibility: a covariance structure modeling approach. *Journalism Quarterly, 71*(1), 159-168.

Yoon, Y. (2005). Examining journalists' perceptions and news coverage of stem cell and cloning organizations. *Journalism & Mass Communication Quarterly, 82*(2), 281-300.

Tables

Table 1
Channel Credibility Index (Meyer, 1988; West, 1994; Yoon, 2005)

Item	**Definition**
Fair – unfair	*Includes different points of views vs. does not include different points of views*
Unbiased – biased	*Opinions are well-founded vs. opinions are not well founded*
Tells the whole story – does not tell the whole story	*The essential points are included vs. the essential points are not included*
Accurate – inaccurate	*The information given usually would be verifiable and true if examined vs.* *The information given usually would not be verifiable and true if examined*
Can be trusted – cannot be trusted	*Statements are usually true/correct – statements are usually not true/correct*

Table 2
Cluster analysis of items

Item	overall (N=68)	Cluster 1 (n=14)	Cluster 2 (n=10)	Cluster 3 (n=33)
credible	24%	0%	0%	37%
accurate	50%	100%	90%	26%
authority	1%	0%	0%	2%
balanced	3%	7%	0%	2%
biased	25%	100%	0%	7%
believable	35%	36%	90%	23%
clear	3%	7%	0%	2%
competent	1%	0%	0%	2%
complete	22%	86%	0%	5%
critical	1%	0%	0%	2%
detailed	1%	0%	0%	2%
ethical	1%	0%	0%	2%
experienced	3%	0%	0%	5%
expert	10%	0%	10%	14%
factual	7%	7%	0%	9%
fair	35%	79%	100%	7%
false	1%	0%	0%	2%
honest	0%	0%	0%	0%
indepth	16%	7%	100%	0%
informed	7%	0%	10%	9%
intelligent	3%	0%	0%	5%
knowledge	3%	0%	0%	5%
neutral	1%	0%	0%	2%
objective	4%	0%	0%	7%
professional	3%	0%	0%	5%
qualified	7%	0%	0%	12%
rational	1%	0%	0%	2%
reliable	3%	0%	0%	5%
reputable	4%	0%	0%	7%
serious	1%	0%	0%	2%
successful	3%	0%	0%	5%
Thoroughly researched	1%	0%	0%	2%
trustworthy	60%	100%	10%	49%
truthful	3%	0%	0%	5%
useinfo	1%	0%	0%	2%
wellwritten	1%	0%	0%	2%
integrity	2%	0%	0%	5%
other	31%	21%	0%	42%

Table 3
Items used by concept

Item	source cred (n=21)	media cred (n=29)	message cred (n=8)	other * (n=10)
credible	29%	17%	25%	30%
accurate	29%	66%	75%	30%
authority	0%	3%	0%	0%
balanced	0%	7%	0%	0%
biased	33%	21%	25%	20%
believable	24%	41%	62%	20%
clear	0%	7%	0%	0%
competent	0%	3%	0%	0%
complete	14%	24%	25%	20%
critical	0%	3%	0%	0%
detailed	0%	3%	0%	0%
ethical	5%	0%	0%	0%
experienced	10%	0%	0%	0%
expert	19%	7%	0%	10%
factual	0%	10%	25%	0%
fair	19%	55%	38%	10%
false	5%	0%	0%	0%
honest	0%	0%	0%	0%
indepth	0%	34%	12%	0%
informed	14%	4%	12%	0%
intelligent	10%	0%	0%	0%
knowledge	5%	3%	0%	0%
neutral	0%	3%	0%	0%
objective	5%	7%	0%	0%
professional	5%	3%	0%	0%
qualified	19%	0%	0%	10%
rational	5%	0%	0%	0%
reliable	10%	0%	0%	0%
reputable	5%	0%	0%	20%
serious	0%	3%	0%	0%
successful	0%	0%	0%	20%
thoroughlyresearched	0%	3%	0%	0%
trustworthy	90%	45%	38%	60%
truthful	5%	3%	0%	0%
useful information	5%	0%	0%	0%
well written	0%	0%	12%	0%
integrity	0%	0%	0%	10%
other	40%	17%	25%	60%

Note: * These were: opinion poll credibility (2), owner credibility (2), perceived credibility (1), platform credibility (1), sponsor credibility (2), website credibility (1), and one scale that was used to measure source, message, and channel credibility.

Figures

Figure 1
Year of publication of the articles in the sample

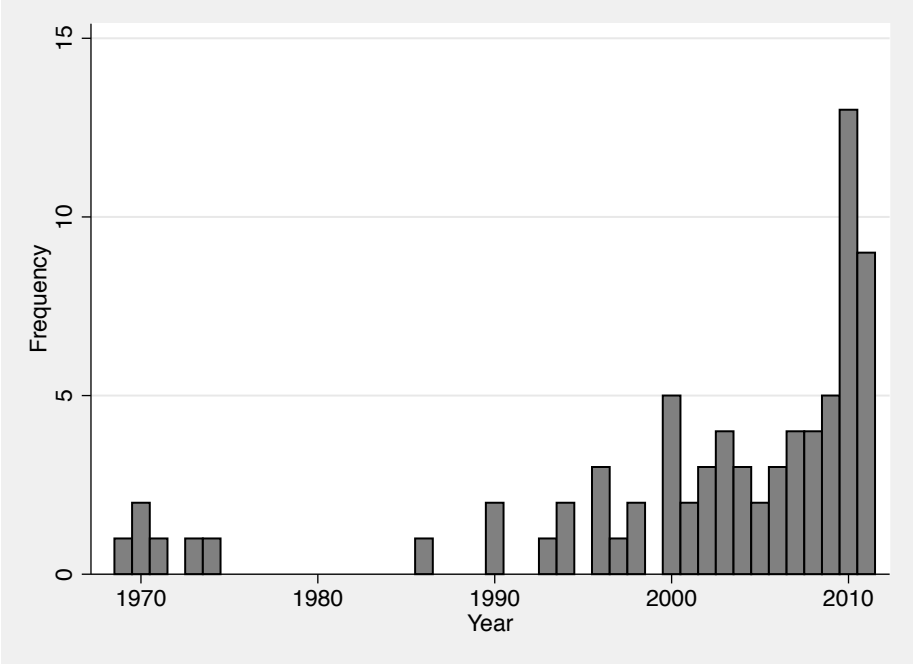

Figure 2
Usage of different credibility concepts over time

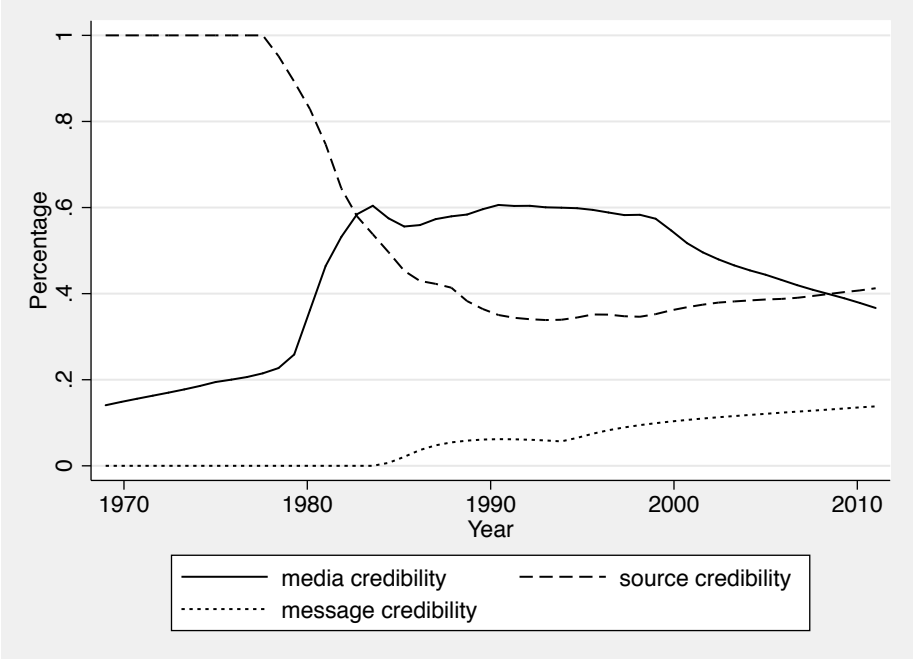

Figure 3

Survey sample sizes. [bar width = 250]

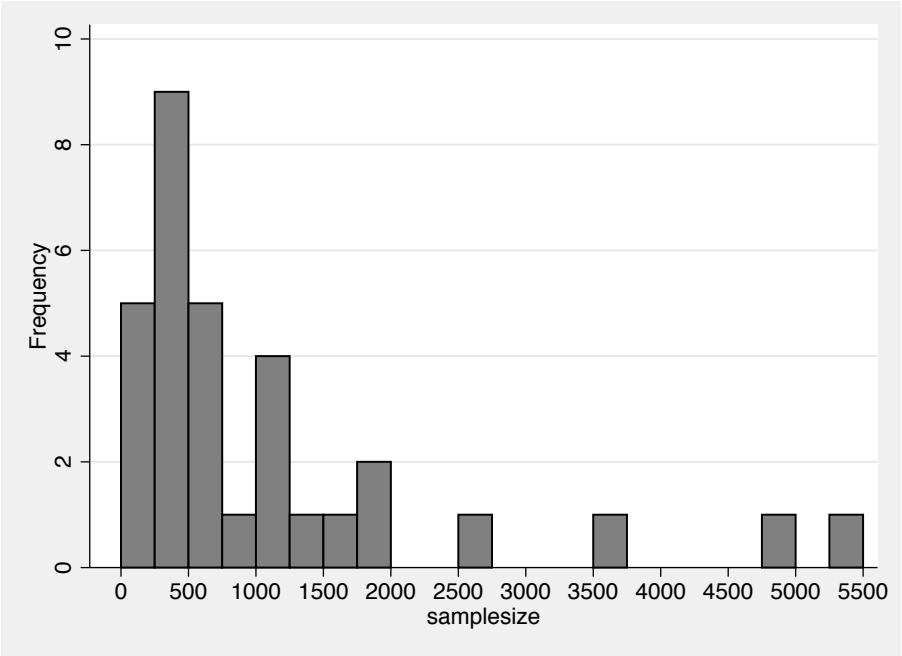

Figure 4

Experiment sample sizes. [bar width = 50]