Examining First- and Second-Order Factor Structures for News Credibility

Robert N. Yale, Jakob D. Jensen, Nick Carcioppolo, Ye Sun & Miao Liu

Published online: 12 Sep 2015.

Submit your article to this journal ⊄

Article views: 80

View related articles ⊄

View Crossmark data ⊄

Routledge
Taylor & Francis Group

Examining First- and Second-Order Factor Structures for News Credibility

Robert N. Yale
University of Dallas

Jakob D. Jensen
University of Utah

Nick Carcioppolo
University of Miami

Ye Sun and Miao Liu
University of Utah

The construct of news credibility has been of interest to communication scholars for decades, yet researchers have struggled to develop a measure of news credibility that demonstrates a reliable factor structure and construct validity. This study uses confirmatory factor analysis (CFA) and discriminant analysis to evaluate Abdulla and colleagues' (2004) measure of news credibility. Results indicate that although the factor structure of the measure is replicable, the measure exhibits significant concerns related to discriminant validity. A revised measurement model employing a second-order factor for the news credibility scale that eliminates the discriminant validity concerns is proposed, and the implications of second-order factors in measurement models are discussed.

Researchers have been attempting to measure news credibility for over 50 years (Kohring & Matthes, 2007; Roper, 1985). Numerous instruments have been developed by researchers and practitioners (e.g., Flanagin & Metzger, 2000; Gaziano & McGrath, 1986; Johnson & Kaye, 1998; Kiousis, 2001; Meyer, 1988; Rimmer & Weaver, 1987; Sundar, 1999; Wanta & Hu, 1994); however, questions about underlying factor structure and construct validity persist. Many scholars have argued that news credibility research has been rife with methodological concerns related to the number of factors, the specific constructs measured, and even item sequencing and polarity biases in various news credibility scales (Kiousis, 2001; Meyer, 1988; Shaw, 1973).

Aside from the methodological problems, from a theoretical standpoint a key question relates to the ways that people conceptualize news credibility. Research related to the measurement of

speaker credibility has produced multidimensional measures with strong psychometric properties that rarely struggle to replicate across studies (e.g., McCroskey & Teven, 1999), indicating that conceptions of credibility likely include a coherent understanding of subdimensions such as expertise and trustworthiness. It may be that news credibility is similar in that people are able to conceptualize different factors of news credibility and evaluate those aspects separately when making judgments. However, it might also be the case that consumers of news evaluate credibility more heuristically, frustrating attempts to generate multi-dimensional measures of news credibility. For example, Kohring and Matthes (2007) explicated several first-order constructs of news credibility, but ultimately found that news consumers best understood the variable as a single second-order factor.

Contributing to this discussion, Abdulla and colleagues (2004) developed a measure of news credibility designed to capture the presumed multidimensional nature of the construct. Their instrument—referred to here as the news credibility scale—was refined using exploratory factor analysis and found to vary by medium (newspaper, online, television). For newspaper credibility, three factors emerged: honesty, currency, and balance (Abdulla et al., 2004).

The goal of the current research is to further evaluate the news credibility scale using confirmatory factor analysis (CFA) and discriminant validity analysis. CFA allows researchers to test and refine proposed measurement models as well as larger theoretical frameworks. CFA is a logical, and necessary, follow-up step to exploratory factor analysis (like that carried out by Abdulla et al., 2004). Discriminant validity analysis examines the viability of proposed latent variables by assessing whether they outperform measurement error and other latent variables in the model (Farrell, 2010; Fornell & Larcker, 1981). To date, no study of news credibility has utilized discriminant validity analysis. Further, the results of these analyses may shed light on why the measurement of news credibility has been so difficult to achieve.

MEASURING NEWS CREDIBILITY

One of the earliest measures of news credibility was created by Roper Research Associates in their studies designed to gauge the relative credibility of various media outlets, including magazines, newspapers, radio, and television (Roper, 1971, 1985). However, the Roper measure had key weaknesses that limited its utility for academic inquiries of news credibility. Carter and Greenberg (1965) found that the wording of the Roper questions resulted in artificially inflated ratings of both newspaper and television news credibility compared with other media. Additionally, an analysis of the Roper measure by Shaw (1973) found evidence that the proximity of questions about media use and media credibility invoked a need for self-consistency, potentially biasing the results in favor of the news medium most frequently used by respondents.

Recognizing the weaknesses of previously used measures of media credibility, Gaziano and McGrath (1986) developed a 12-item measure using principal components analysis to identify a single factor of news credibility. That single-factor structure was supported by Rimmer and Weaver (1987); the measure also appeared to be internally reliable as they reported a Cronbach's alpha reliability coefficient for the 12-item measure of .90. However, Meyer (1988) argued the single-factor structure of Gaziano and McGrath's measure was actually a by-product of the semantic differential item format. Response effects related to the polarity of the semantic differential items created the unidimensional structure of the Gaziano and McGrath measure

artificially. Eliminating the polarity artifact revealed two factors of media credibility: believability and community affiliation (Meyer, 1988).

To address the conflicting results about the factor structure of news credibility, Abdulla et al. (2004) modified the Gaziano and McGrath instrument, converting it from a semantic differential scale into a five-point Likert scale with anchors from 1 = *strongly disagree* to 5 = *strongly agree*. Exploratory factor analysis indicated that newspaper credibility is a construct with three underlying factors: honesty, currency, and balance.

Abdulla and colleagues' (Abdulla et al., 2004) news credibility measure builds meaningfully upon past credibility research. Yet there are still several unanswered questions about the news credibility measure and about the underlying construct of news credibility. First, the factor structure identified by Abdulla et al. (2004) has not been confirmed using CFA. Until the factor structure is replicated, it is difficult to know if the news credibility measure is a valid measurement tool. Past studies have typically used principal components analysis. However, measurement researchers have suggested that principal components analysis is a procedure that should be used for data reduction rather than determining the most parsimonious groupings of measured variables (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Gorsuch, 1990; Morrison, 2009). As many of the published scales used to measure news credibility have been developed using principal components analysis (e.g., Abdulla et al., 2004; Gaziano & McGrath, 1986; Meyer, 1988), it is entirely possible that each of these measures suffers from methodological problems that make identifying the underlying dimensions of news credibility extremely difficult.

Second, existing research has suggested that news credibility could have one, two, three, or (most recently) four underlying factors (Abdulla et al., 2004; Gaziano & McGrath, 1986; Kohring & Matthes, 2007; Meyer, 1988). This ambiguity in the literature raises questions not only about the factor structure but also the discriminant validity of identified and proposed factors (Farrell, 2010). After all, a particular factor structure could be supported by CFA but still invalid due to discriminant validity concerns (Fornell & Larcker, 1981). Discriminant validity examines whether proposed latent variables outperform measurement error and other latent variables in the model. The first sign of discriminant validity problems is unusually high correlations among latent variables. Unfortunately, past news credibility research has not reported correlations among latent variables. Thus, the current study evaluates the proposed factor structure and discriminant validity of the news credibility scale.

STUDY 1

Participants and Design

All participants read one of two brief news articles that reported the results of recent clinical research related to cancer prevention and treatment. The articles were represented as coming from a major national newspaper in the region of data collection and were either 286 words or 513 words in length. After reading the article, participants completed a survey that included the Abdulla et al. (2004) news credibility scale. Participants were paid $10 for completing the study.

A total of 399 participants completed the study, with more females (66.2%) than males (31.6%); 2.2% failed to report their sex. Participants ranged from 18 to 84 years of age, with a mean age of 36.68 years (*SD* = 16.33). In terms of race, 83.2% of participants were Caucasian,

11.7% African American, 3.1% Hispanic, Latino, or Spanish Origin, 1.0% Asian or Pacific Islander, 1.8% American Indian or Native American, and 2.3% described themselves as "other" (participants could select more than one category). From an educational standpoint, 3% had less than a 12th grade education, 52.4% had a high school education, 8% had an associates or technical degree, 33.6% had completed a four-year degree, and 3.3% did not answer. The mean household income was $51,769.46 ($SD = $42,954.35).

Procedures

Participants were recruited from three different shopping malls located in the Midwest. At each location, managers allowed the research team to set up a table and twelve chairs in one of the main intersections of the mall. A team of three to five researchers recruited mall shoppers from 9 A.M. to 9 P.M. over a period of seven days. Participants were recruited verbally and using six large canvas signs (with the name of the university supporting the research). When participants approached the research team they were randomly assigned to one of four different studies, each designed to take approximately 20 minutes to complete (one of which was the present protocol). Approximately 350–400 participants completed each of the four studies. The remaining three studies were unrelated to the current investigation and are reported elsewhere.

Measures

Newspaper Credibility Scale

The newspaper credibility scale (Abdulla et al., 2004) is an 11-item scale with three factors that measure the perceived credibility of news sources. Respondents evaluate news sources using 5-point scales ranging from $-2 = $ *strongly disagree* to $2 = $ *strongly agree* on the following subscales and items: balance (balanced, report the whole story, objective, fair, accurate), honesty (honest, believable, trustworthy), and currency (up-to-date, current, timely). The subscales of balance, honesty, and currency have demonstrated adequate internal consistency with Cronbach alpha coefficients of .78, .65, and .70, respectively (Abdulla et al., 2004).

Results

Item Analysis

Approximately 3% of the data were missing and replaced using hot deck imputation with location and news article used as deck variables (Myers, 2011). Calculation of scale and subscale reliability for responses to the newspaper credibility scale established a full-scale Cronbach's alpha of .88 ($M = .56$, $SD = .55$). For the subscales, balance achieved a Cronbach's alpha of .80 ($M = .40$, $SD = .62$), honesty was .78 ($M = .62$, $SD = .65$), and currency was .77 ($M = .78$, $SD = .65$).

Scores on each of the subscales were non-normally distributed. Univariate tests of normality revealed the following subscale scores: for the balance subscale, skewness $= .93$ ($SE = .12$), kurtosis $= 2.79$ ($SE = .24$); for the honesty subscale, skewness $= .42$ ($SE = .12$), kurtosis $= 2.08$ ($SE = .24$), and for the currency subscale, skewness $= .55$ ($SE = .12$), kurtosis $= 2.53$ ($SE = .24$).

As a full measure, the news credibility items exhibited significant multivariate abnormality, skewness = 15.19, z score = 18.79, $p < .001$, and kurtosis = 193.10, z score = 15.07, $p < .001$.

Micceri (1989) argues that non-normality of data is far more common than researchers are aware and must be accounted for in the analysis. Given the observed multivariate skew and kurtosis in this data, we adjusted the model when conducting the CFA by using the asymptotic covariance matrix. Thus, a Satorra-Bentler (S-B) χ^2 is reported to adjust for non-normal distributions (Hu, Bentler, & Kano, 1992; Satorra & Bentler, 2010). The S-B χ^2 is robust against violations of normality with regard to kurtosis because it specifically corrects the normal theory chi-square by a scalar value that includes the observed multivariate kurtosis (Curran, West, & Finch, 1996), thus accounting for whatever level of kurtosis is present in the data. However, given the level of skewness observed in this data, replication is necessary to evaluate the potential error introduced by non-normality.

Confirmatory Factor Analysis

Confirmatory factor analysis was conducted on participant responses to the news credibility scale items using LISREL 8.80. Model fit was evaluated using the root mean square error of approximation (RMSEA; Nevitt & Hancock, 2000), the comparative fit index (CFI; Bentler, 1990), and the non-normed fit index (NNFI; Bentler & Bonett, 1980), as these indices restrict random variation with incorrect model specification, variations in sample size, and estimation method selection (Fan, Thompson, & Wang, 1999). Hu and Bentler (1999) suggest good model fit is indicated by values greater than .95 for the CFI and NNFI. RMSEA values less than or equal to .06 indicate good fit, while excellent fit is indicated by values lower than .05 (Holbert & Stephenson, 2008; Hu & Bentler, 1999). Additionally, Model AIC was used to compare different models; lower scores indicate a better fitting model (Akaike, 1987).

Based on the previously identified factor structure of the news credibility scale by Abdulla et al. (2004), a three-factor measurement model was specified. The three-factor model exhibited good fit, S-B χ^2 (41, $N = 399$) = 98.07, $p < .001$, RMSEA = .06 (90% CI: .04—.07), NNFI = .98, CFI = .98, Model AIC = 148.07 (See Figure 1).

Given the high correlation between the latent constructs of balance and honesty, a two-factor model was specified in which the observed variables loading on balance and honesty were specified to load on the single latent factor of balance. Compared with the three-factor model, the two-factor model demonstrated somewhat worse fit, S-B χ^2 (43, $N = 399$) = 110.56, $p < .001$, RMSEA = .06 (90% CI: .05—.08), NNFI = .98, CFI = .98, Model AIC = 156.56 (see Figure 1).

As the two-factor model still demonstrated high correlation between the latent constructs of currency and balance, a third model was specified in which all of the observed variables loaded onto a single latent construct, news credibility. The one-factor model demonstrated poor fit, S-B χ^2 (44, $N = 399$) = 236.87, $p < .001$, RMSEA = .10 (90% CI: .09—.12), NNFI = .94, CFI = .95 Model AIC = 280.87 (see Figure 1). These results suggest that the initial model suggested by Abdulla et al. (2004) is the best fit for the data. However, the high correlations between latent constructs suggest that there may be questions about discriminant validity.

The statistical formula provided by Fornell and Larcker (1981) was used to test the discriminant validity of the latent variables in the three-factor model. Discriminant validity is demonstrated if the average variance extracted (AVE) for each construct is greater than the square
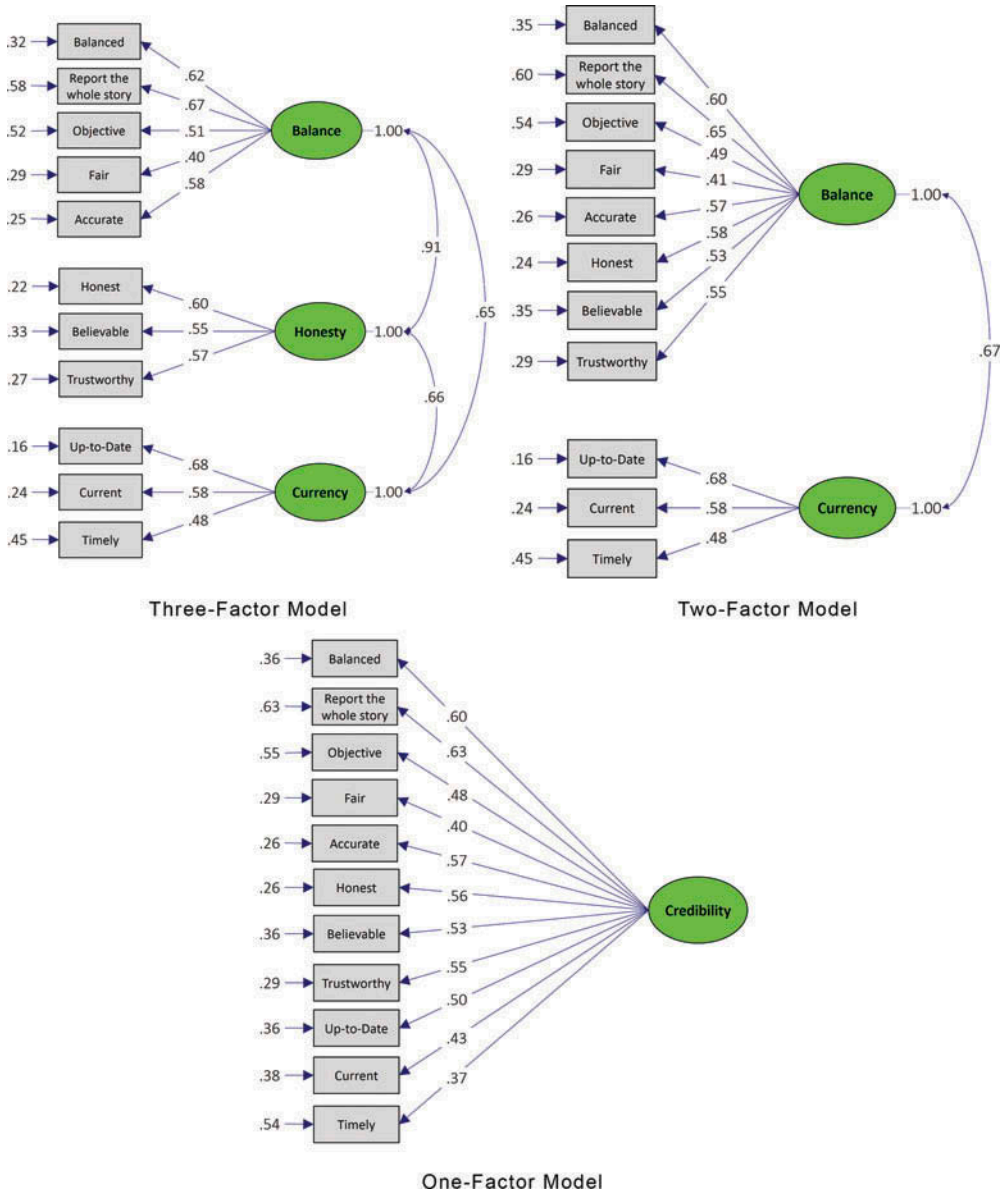
FIGURE 1 Confirmatory factor analyses testing three possible models of
news credibility (Study 1 data, $N = 399$). The three-factor model exhibited
the best fit, Model AIC $= 148.07$.

of the correlation (R^2) between two constructs. In all cases, AVE (balance $= .32$, honesty $= .33$,
currency $= .34$) was lower than the corresponding R^2 value (see Figure 1 for correlations).
These results suggest that the news credibility scale suffers from significant problems related

TABLE 1
Principal Axis Factor Analysis with Direct Oblimin Rotation for Study 1 ($N = 399$)

Item	Factor		
	1	*2*	*3*
Balanced	.756	.100	.111
Report the whole story	.771	−.073	−.017
Objective	.789	−.033	.175
Fair	.410	.290	−.093
Accurate	.632	.199	−.069
Honest	.623	.019	−.368
Believable	.628	.059	−.116
Trustworthy	.429	.222	−.300
Up-to-date	.002	.822	−.105
Current	.015	.733	−.103
Timely	.060	.711	.225

Note. Principal axis factor analysis of the news credibility measure.

to discriminant validity. Additionally, all AVE values were below the threshold of .5, indicating that, on average, the variance explained by the latent factor is less than the amount of error remaining in the items (Hair, Black, Babin, & Anderson, 2009). This finding suggests that the measure also exhibits problems related to convergent validity.

Farrell (2010) advocates using exploratory factor analysis (EFA) to determine whether the insufficient discriminant validity is the result of items cross-loading with multiple factors. The analysis was conducted using principal axis factoring with direct oblimin rotation. The number of factors to extract was fixed at three based on the expected factor structure of the measure. Only one item (report the whole story) demonstrated concerning cross-factor loadings, suggesting that removing the cross-loading item is insufficient to overcome the significant correlations between latent variables. Table 1 presents the factor loadings based on the study 1 sample.

STUDY 2

While the three-factor model was a good fit to the data, insufficient discriminant validity undermined the validity of the model and could not be corrected using Farrell's (2010) recommendations. In some cases, discriminant validity problems may be the result of sampling error, necessitating the analysis of a new sample to assess whether or not the problems persist with new data (Bollen, 1989). It is also possible that the high level of skewness present in the sample introduced random error that undermined the factor structure.

Study 2 replicates the procedure of study 1 with different news articles to determine whether or not the CFA result is consistent and whether or not the discriminant validity problems and skewness nonnormality are eliminated with a different sample.

Participants and Design

All participants read one of two brief news articles (different from those used in study 1) that reported the results of recent clinical research related to cancer prevention and treatment. The articles were represented as coming from a major newspaper in the region of data collection and were 280 words or 648 words in length, respectively. After reading the article, participants completed a survey that included the Abdulla et al. (2004) news credibility scale. Participants were paid $10 for completing the study.

Adult participants ($N = 478$) were recruited from a shopping mall located in the Midwest using the same procedures as study 1. More females (63.9%) participated than males (34.7%), and participants were 18–84 years of age, with a mean age of 35.44 years ($SD = 15.78$). In terms of race, 76.2% of participants were Caucasian, 9.2% African American, 5.4% Hispanic, Latino, or Spanish Origin, 5.0% Asian or Pacific Islander, 1.5% American Indian or Native American, and 1.7% described themselves as "other" (participants could select more than one category). From an educational standpoint, 5.6% had not completed high school, 30.7% had a high school education, 27.6% had some college, 25.1% had completed a four year degree, 10.2% had post-graduate education, and .8% did not answer. The mean household income was $45,545.73 ($SD = \$33,580.14$).

Measures

The 11-item news credibility scale from Abdulla et al. (2004) was used to measure perceptions of news credibility.

Results

Item Analysis

Approximately 1.6% of the data were missing and replaced using hot deck imputation, with news article and news source used as deck variables (Myers, 2011). Calculation of scale and subscale reliability for responses to the news credibility scale established a full-scale Cronbach's alpha of .90 ($M = .40$, $SD = .58$). For the subscales, balance achieved a Cronbach's alpha of .81 ($M = .28$, $SD = .63$), honesty was .80 ($M = .52$, $SD = .63$), and currency was .81 ($M = .49$, $SD = .70$).

Scores on each of the subscales were non-normally distributed. Univariate tests of normality revealed the following subscale scores: for the balance subscale, skewness = -.24 ($SE = .11$), kurtosis = 1.04 ($SE = .22$); for the honesty subscale, skewness = -.81 ($SE = .11$), kurtosis = 2.08 ($SE = .22$); and for the currency subscale, skewness = -.65 ($SE = .11$), kurtosis = 1.16 ($SE = .22$). As a scale, the news credibility items exhibited significant multivariate abnormality, skewness = 18.32, z score = 23.99, $p < .001$, and kurtosis = 243.82, z score = 19.80, $p < .001$.

Confirmatory Factor Analysis Replication

The three-factor measurement model identified as the best fit in study 1 was specified, and similar to the previous result, exhibited good fit, S-B χ^2 (41, $N = 478$) = 101.99, $p < .001$, RMSEA = .06 (90% CI: .04—.07), NNFI = .99, CFI = .99, Model AIC = 151.99 (see Figure 2).
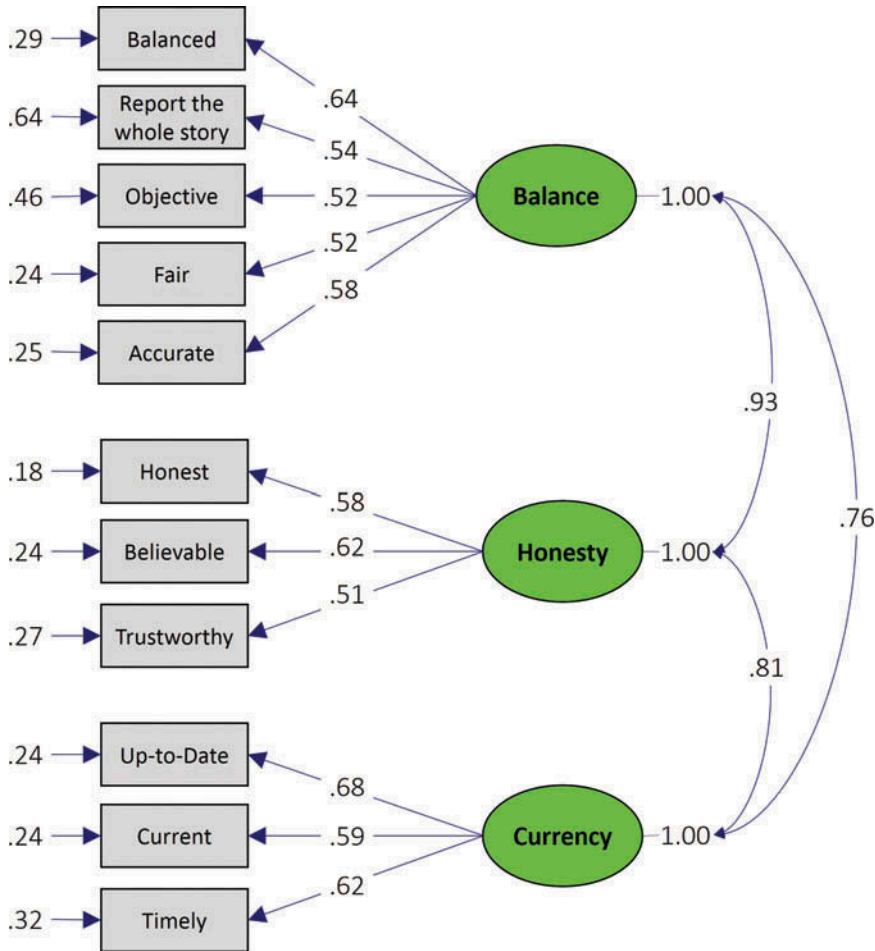
FIGURE 2 Study 2 confirmatory factor analysis replicating the three-factor model of news credibility ($N = 478$).

Correlations between latent factors were even higher with this sample than with the sample from study 1. Discriminant analysis using the Fornell and Larcker (1981) procedure revealed that in all cases, AVE (balance = .32, honesty = .33, currency = .40) was lower than the R^2 value and was below the threshold of .5 required to suggest adequate convergent validity (See Figure 2 for correlations).

Exploratory factor analysis using principal axis factoring with direct oblimin rotation with the number of factors fixed at three revealed similar results to study 1. Only one item (balanced) demonstrated concerning cross-factor loadings, but at a level insufficient to overcome the significant correlations between latent variables. Table 2 presents the factor loadings based on the study 2 sample. These results suggest that while the factor structure is replicable, the problems with discriminant validity were not the result of sampling error.

TABLE 2
Principal Axis Factor Analysis with Direct Oblimin Rotation for Study 2 ($N = 478$)

	Factor		
Item	*1*	*2*	*3*
Balanced	.450	.032	.395
Report the whole story	−.027	−.072	.641
Objective	.136	−.031	.526
Fair	.603	.032	.188
Accurate	.599	−.029	.155
Honest	.663	−.076	.047
Believable	.922	−.031	−.174
Trustworthy	.467	−.154	.122
Up-to-date	.019	−.781	.027
Current	−.010	−.154	.122
Timely	.357	−.336	.123

Note. Principal axis factor analysis of the news credibility measure.

STUDY 3

Study 2 confirmed the discriminant validity problems with the news credibility scale observed in study 1 and again exhibited high multivariate skewness. However, both data sets were collected under similar circumstances, in similar geographic areas, with articles related to the same general topic represented as originating from the same national news source, resulting in possible sampling bias and stimulus-specific effects which may have influenced the results. Study three evaluates the replicability of the CFA and discriminant analysis with a new sample obtained using different procedures and with a wider variety of stimulus article topics and news sources.

Participants and Design

A 3 (story topic) × 4 (news source: *The Wall Street Journal*, Fox News, *USA Today*, ABC News) news story randomization was embedded in an online survey about news credibility. This factorial design crossed story topics and news sources to increase the number of story topics and sources while averaging out any potential bias effects of any particular story or news source. Thus, the subsequent CFA analyses were conducted independent of this design. These specific major national newspaper and broadcast news outlets were selected to increase the potential range of source credibility effects captured by the measure as they are perceived as the two most (ABC News, *The Wall Street Journal*) and least (*USA Today*, Fox News) believable news organizations (Pew Research Pew Research Center, 2012). News stories covered the topics of the environment (Dean, 2012), business (Fox, 2014), and science (Park, 2014), and ranged in length from 396 words to 550 words. After reading the article, participants completed a survey that included the Abdulla et al. (2004) news credibility scale.

A total of 504 participants completed the online study, with representatives from 48 states and the District of Columbia. Twenty-six participants failed one or more of the screening questions and were removed from the study, resulting in a final sample size of 478. Slightly more

females (55.4%) participated than males (44.6%), and 60.88% had household incomes below the 2012 US median income of $51,371. Participants ranged from 18–75 years of age, with a mean age of 36.10 years ($SD = 12.23$). In terms of race, 83.47% of participants were Caucasian, 7.11% African American, 5.2% Hispanic, Latino, or Spanish Origin, 7.3% Asian or Pacific Islander, 1.67% American Indian or Native American, and 1.1% described themselves as "other" (participants could check more than one category). From an educational standpoint, 0.6% had not completed high school, 11.5% had a high school education, 38.3% had some college, 37.0% had completed a four year degree, and 12.5% had postgraduate education.

Procedure

Participants were recruited to complete the online survey via Amazon Mechanical Turk over a period of five days and received $0.50 and entry into a $100 raffle for completing the study. To increase the sample diversity across all available Mechanical Turk workers, we used the method of posting the survey task at different times of day and on both weekdays and weekends throughout the recruitment period, as suggested by Kapelner and Chandler (2010). Participants were prohibited from completing the study multiple times using the method explained by Pe'er, Paolacci, Chandler, and Mueller (2012). Screening questions, similar to those suggested by Downs, Holbrook, Sheng, and Cranor (2010), were embedded into the study task to identify participants who were not carefully attending to the study tasks. All participants were adults aged 18 or older living in the United States.

Measures

The 11-item news credibility scale from Abdulla et al. (2004) was used to measure perceptions of news credibility.

Results

Item Analysis

Calculation of scale and subscale reliability for responses to the news credibility scale established a full-scale Cronbach's alpha of .91 ($M = .48$, $SD = .63$). For the subscales, balance achieved a Cronbach's alpha of .88 ($M = .28$, $SD = .74$), honesty was .81 ($M = .64$, $SD = .70$), and currency was .85 ($M = .66$, $SD = .76$). Scores on each of the subscales were non-normally distributed. Univariate tests of normality revealed the following subscale scores: for the balance subscale, skewness $= -.37$ ($SE = .11$), kurtosis $= .02$ ($SE = .22$); for the honesty subscale, skewness $= -.61$ ($SE = .11$), kurtosis $= .73$ ($SE = .22$); and for the currency subscale, skewness $= -.55$ ($SE = .11$), kurtosis $= .18$ ($SE = .22$). As a scale, the news credibility items exhibited significant multivariate abnormality, skewness $= 13.06$, z score $= 19.32$, $p < .001$, and kurtosis $= 183.91$, z score $= 14.72$, $p < .001$.
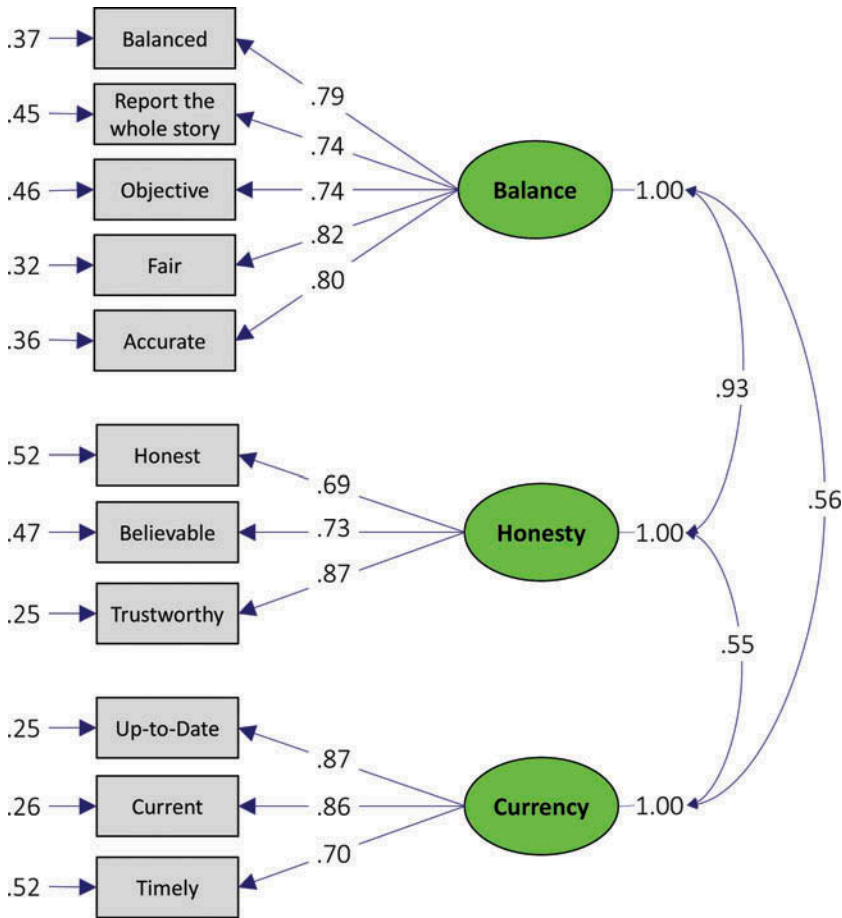
FIGURE 3 Study 3 confirmatory factor analysis replicating the three-
factor model of news.

Confirmatory Factor Analysis Replication

The three-factor measurement model identified as the best fit in study 1 was specified and exhibited somewhat worse fit with this data compared to the previous two studies, S-B χ^2 (41, $N = 478$) $= 153.91$, $p < .001$, RMSEA $= .08$ (90% CI: .06—.09), NNFI $= .98$, CFI $= .98$, Model AIC $= 203.91$ (see Figure 3). The extremely high correlation between the latent factors of balance and honesty continued to create insufficient discriminant validity. Discriminant analysis using the Fornell and Larcker (1981) procedure revealed that for two of the three latent constructs, AVE (balance $= .61$, honesty $= .59$, currency $= .66$) was lower than the R^2 value (See Figure 3 for correlations). However, in this sample, all three latent constructs achieved AVE higher than .50, indicating adequate convergent validity.

TABLE 3
Principal Axis Factor Analysis with Direct Oblimin Rotation for Study 3 ($N = 478$)

Item	Factor		
	1	2	3
Balanced	.783	.033	.401
Report the whole story	.687	.076	.133
Objective	.751	−.035	.182
Fair	.789	.037	.116
Accurate	.762	.094	−.218
Honest	.668	.011	−.226
Believable	.624	.153	−.262
Trustworthy	.854	−.024	−.141
Up-to-date	.059	.816	.046
Current	−.096	.948	.018
Timely	.050	.660	−.024

Note. Principal axis factor analysis of the news credibility measure.

With this sample, the extremely high correlation between the latent factors of Balance and Honesty persisted. However, the correlations between Balance and Currency and between Honesty and Currency were greatly reduced, suggesting that a two-factor model might be a better fit to the data. A two-factor measurement model was specified which collapsed the latent factors of Balance and Honesty into a single factor: Balance. The two-factor measurement model exhibited worse fit compared to the three-factor model, S-B χ^2 (43, $N = 478$) = 184.24, $p < .001$, RMSEA = .08 (90% CI: .07—.10), NNFI = .97, CFI = .98, Model AIC = 230.24 (see Figure 3).

With this particular data set, examination of the results of exploratory factor analysis to identify problematic cross loadings per Farrell's (2010) recommendations was not helpful, as the five indicators for Balance and the four indicators for Honesty seem to group into a single-factor structure (see Table 3). However, examination of the item correlation matrix revealed some possible items to drop from the model in order to reduce the problematic correlation between the latent factors of Balance and Honesty: fair and trustworthy. Each of these indicator items demonstrated extremely high correlations with all of the items in both the Balance and the Honesty factors, suggesting that they might be inflating the correlation between the latent variables. Unfortunately, dropping trustworthy from the Honesty factor would result in a construct with fewer than three indicators, a situation that should be avoided in confirmatory factor analysis (Hair et al., 2009).

Fortunately, further examination of the item correlation matrix revealed a potential solution: the accurate indicator from the Balance factor correlated more strongly with the items in the Honesty factor. On its face, this seems to make sense, as an evaluation of a news story's accuracy seems more conceptually related with Honesty than with Balance. Moving the accurate indicator from Balance to Honesty provides the required three items per indicator, per the recommendation of Hair et al. (2009).

A three-factor measurement model was specified with the following subscales and items: balance (balanced, report the whole story, objective), honesty (accurate, honest, believable), and currency (up-to-date, current, timely). The model proved to be an excellent fit, S-B χ^2 (24, $N = 478$) = 47.45, $p < .001$, RMSEA = .05 (90% CI: .03—.06), NNFI = .99, CFI = .99,
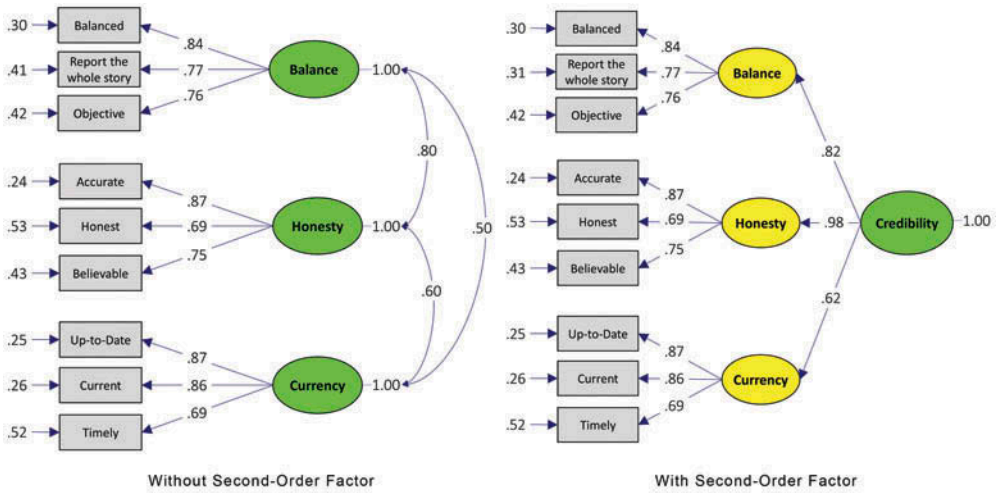
FIGURE 4 Model specification of revised news credibility measurement
model without and with a second-order factor.

Model AIC $= 89.45$ (see Figure 3). Although the removal of two items and movement of one
did reduce the correlation between latent factors (see Figure 4), it was not sufficient to eliminate
the discriminant validity concerns. Discriminant analysis using the Fornell and Larcker (1981)
procedure revealed that for two of the three latent constructs, AVE (balance $= .63$, honesty $= .60$,
currency $= .66$) was lower than the R^2 value (See Figure 4 for correlations).

Unfortunately, although a better fit was achieved with the new nine-item model specification,
the problems with discriminant validity were unable to be eliminated and combining factors in
the measurement model reduced fit to unacceptable levels. The only viable solution seems to
be to model news credibility as a second-order factor with three first-order factors. Kohring and
Matthes (2007) utilized this approach when evaluating a model of trust in news media. A second-
order model was tested with three first-order factors (balance, honesty, currency) and one second-
order factor (news credibility) (see Figure 4). The second-order model exhibited identical fit
statistics as the prior model.

In order to confirm this new measurement model, the nine-item model with second-order factor
was tested with the data from Study 1. The revised model specification provided a better fit
compared to the original Abdulla et al. (2004) model, S-B χ^2 (24, $N = 399) = 47.84$, $p < .001$,
RMSEA $= .05$ (90% CI: .03—.07), NNFI $= .99$, CFI $= .99$, Model AIC $= 89.84$.

DISCUSSION

This research sought to confirm the factor structure and discriminant validity of the news credi-
bility scale. The results of the present endeavor uncovered larger measurement questions related
to discriminant validity and higher-order factor models, since the three-factor model proposed by
Abdulla et al. (2004) received mixed support.

While it is possible that the high level of uncorrected skewness observed in the data introduced error, there are several reasons why this is unlikely. First, investigations of the robustness of various statistical tests to violations of normality have demonstrated that tests relying on variance and covariance (e.g., CFA) are more robust to skewness nonnormality than kurtosis nonnormality (Jobson, 1999; Mardia, Kent, & Bibby, 1979) suggesting that the skewness present in these data are unlikely to undermine the results. Further, fairly consistent results were obtained with the news credibility measure in three different studies, each with different news article stimuli. Finally, each of the CFA models achieves reasonable fit, also suggesting that the skewness in the data is not introducing error that is significant enough to nullify these results.

Significant discriminant validity concerns raise questions about the utility of the latent constructs Honesty, Currency, and Balance. Specifically, the problems with discriminant validity in the Abdulla et al. (2004) measure suggest that the use of the separate factors in research would introduce significant conceptual ambiguity and potentially undermine analytical approaches. Concerning the latter, the high correlations among the latent variables would yield multicollinearity concerns. If, as was the case in this investigation, second-order factors are required to eliminate discriminant validity concerns, then for all practical purposes, the scale must be treated as a single-factor measure when used as a variable in research studies.

The dubious utility of first-order factors is not confined to the Abdulla news credibility scale. Kohring and Matthes (2007) developed a measure of trust in news media with four dimensions. Hierarchical confirmatory factor analysis supported a second-order factor model which was replicated in a second dataset. Their measurement model didn't postulate correlations among first-order factors—consistent with the logic of a hierarchical model—but the insufficient discriminant validity identified here suggest that it would be fruitful to consider those relationships. Kohring and Matthes graciously provided the correlations among the four first-order factors in the model in response to an email inquiry. They are: Selectivity of Topics with Selectivity of Facts, .82, Selectivity of Topics with Accuracy of Depictions, .62, Selectivity of Topics with Journalistic Assessment, .73, Selectivity of Facts with Accuracy of Depictions, .85, Selectivity of Facts with Journalistic Assessment, .90, and Accuracy of Depictions with Journalistic Assessment, .83.

Discriminant analysis of the Kohring and Matthes model using the Fornell and Larcker (1981) procedure revealed that, with the exception of the Selectivity of Topics factor, AVE scores (Selectivity of Topics = .55, Selectivity of Facts = .44, Accuracy of Depictions = .53, Journalistic Assessment = .44) were lower than the corresponding R^2 values. Given this confirmed lack of discriminant validity between first-order factors, it would be entirely inappropriate to use the subdimensions of the measure as variables in research, even though the scale was specifically designed as a "multiple factor model" (Kohring & Matthes, 2007, p. 238).

These results, in concert with the findings of this study concerning the factor structure and discriminant validity problems in the Abdulla et al. (2004) news credibility measure, underscore the difficulty for researchers who seek to develop a valid multidimensional measure of news credibility. Persistent appearances of discriminant validity problems suggest that it would be meaningful to further explicate the underlying construct of news credibility. Factor analytic methods have been used to study news credibility for several decades, and the majority of the work appears to have been largely data-driven. That is, researchers have ascribed names and logic to factors that manifest in particular datasets. Attempts to replicate those factors have typically proven unsuccessful, and the lack of theoretically derived, fully explicated dimensions may be an underlying cause for this reoccurring problem.

From a theoretical standpoint, these results provide support for the idea that although philosophical distinctions may be made between concepts like news balance, honesty, and currency, people seem to evaluate news credibility more heuristically. The high correlations between first-order latent factors observed in multiple samples and even with different measures of news credibility suggest that conceptions of news credibility lack coherent differentiation of subfactors. This may be one reason why developing a valid measure of news credibility has proven difficult.

Traditional credibility measurement research focused on people rather than news (e.g., McCroskey & Teven, 1999) demonstrates that when individuals evaluate the credibility of another person, they are able to consistently differentiate between various dimensions of credibility, including expertise, trustworthiness, and goodwill. Evaluating news credibility may be an entirely different kind of judgment where such nuanced distinctions are unlikely because news is more of an end product and the distance between the sender (i.e., journalist or editor) and receiver is often significant. This distance suggests that evaluations of news credibility may be more heuristic compared to the systematic evaluation of credibility that is possible with less distant subjects. Further research should investigate whether reducing this distance improves the ability of individuals to differentiate between sub-dimensions of news credibility like those present in the Abdulla et al. (2004) and Kohring and Matthes (2007) scales. For example, evaluation of the credibility of televised broadcast news may decrease the distance between sender and receiver, reducing the tendency for heuristic evaluation of credibility and improving the discriminant validity of multidimensional measures of news credibility.

Researchers should also be mindful of the significant shifts that are reshaping journalism, and how those shifts may alter how the public conceptualizes news credibility. Past research has examined shifts in public perception of news credibility over time (e.g., Roper, 1971, 1985), yet those studies assumed that the underlying structure of news credibility remained unchanged. Given the ever-changing landscape of the information media, it is possible that notions about what makes a news source credible are also shifting.

The results of this research indicate that the nine-item revision of the Abdulla et al. (2004) scale represents an improved measure of news credibility. However, despite its multidimensional structure, the inclusion of a higher-order factor to eliminate discriminant validity problems requires that the measure be used as a single nine-item scale rather than as three subscales. Researchers who wish to use the original Adbulla et al. (2004) scale to allow comparisons with existing studies should use the full 11-item scale rather than the three subscales.

The current study was limited in several ways. First, all studies used sampling methods that produced participant samples which may not be representative of the public as a whole or even adults within the United States. Second, participants responded to questions after reading one of seven news articles about various topics. The text-based news articles were either printed on paper (studies 1 and 2) or displayed on a screen (study 3) and presented as though from a major news source. Past research has suggested that news credibility could vary by medium (Abdulla et al., 2004), thus it is possible that the results observed here may not replicate for perceptions of other media channels.

Media professionals work very hard to cultivate and maintain credibility with their audiences. Communication researchers have been just as diligent when it comes to measuring news credibility. Accordingly, continued construct explication, combined with rigorous psychometric methods, will serve to rectify many of the problems identified in the current investigation.

REFERENCES

Abdulla, R. A., Garrison, B., Salwen, M. B., Driscoll, P. D., & Casey, D. (2004). Online news credibility. In M. B. Salwen, B. Garrison, & P. D. Driscoll (Eds.), *Online news and the public* (pp. 147–163). New York, NY: Routledge.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*(3), 317–332. doi:10.1007/BF02294359

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. doi:10.1037/0033-2909.107.2.238

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588–606. doi:10.1037/0033-2909.88.3.588

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: John Wiley & Sons.

Carter, R. F., & Greenberg, B. S. (1965). Newspapers or television: Which do you believe? *Journalism & Mass Communication Quarterly*, *42*(1), 29–34. doi:10.1177/107769906504200104

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to non-normality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*(1), 16–29. doi:10.1037/1082-989X.1.1.16

Dean, C. (2012, January 26). With prevalence of nanomaterials rising, panel urges review of risks. *New York Times*, p. A16. Retrieved from http://www.nytimes.com/2012/01/26/science/nanomaterials-effects-on-health-and-environment-unclear-panel-says.html

Downs, J. S., Holbrook, M. B., Sheng, S., & Cranor, L. F. (2010, April). *Are your participants gaming the system?: Screening Mechanical Turk workers*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, GA.

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*(3), 272–299. doi:10.1037/1082-989X.4.3.272

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 56–83. doi:10.1080/10705519909540119

Farrell, A. M. (2010). Insufficient discriminant validity: A comment on Bove, Pervan, Beatty, and Shiu (2009). *Journal of Business Research*, *63*(3), 324–327. doi:10.1016/j.jbusres.2009.05.003

Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of Internet information credibility. *Journalism & Mass Communication Quarterly*, *77*(3), 515–540. doi:10.1177/107769900007700304

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research (JMR)*, *18*(1), 39–50.

Fox, E. J. (2014, March 31). *260,000 graduates in minimum wage jobs*. Retrieved from http://money.cnn.com/2014/03/31/news/economy/minimum-wage-college-graduates/index.html

Gaziano, C., & McGrath, K. (1986). Measuring the concept of credibility. *Journalism & Mass Communication Quarterly*, *63*(3), 451–462. doi:10.1177/107769908606300301

Gorsuch, R. L. (1990). Common factor analysis versus component analysis: Some well and little known facts. *Multivariate Behavioral Research*, *25*(1), 33–39. doi:10.1207/s15327906mbr2501_3

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.

Holbert, R. L., & Stephenson, M. T. (2008). Commentary on the uses and misuses of structural equation modeling in communication research. In A. F. Hayes, M. D. Slater, & L. B. Snyder (Eds.), *The Sage sourcebook of advanced data analysis methods for communication research* (pp. 185–218). Thousand Oaks, CA: Sage.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. doi:10.1080/10705519909540118

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112*(2), 351–362. doi:10.1037/0033-2909.112.2.351

Jobson, J. (1999). *Applied multivariate data analysis: Regression and experimental design* (Corrected ed.). New York, NY: Springer.

Johnson, T. J., & Kaye, B. K. (1998). Cruising is believing?: Comparing internet and traditional sources on media credibility measures. *Journalism & Mass Communication Quarterly*, *75*(2), 325–340. doi:10.1177/107769909807500208

Kapelner, A., & Chandler, D. (2010, Oct.). *Preventing satisficing in online surveys*. Paper presented at the CrowdConf 2010 Meeting, San Francisco, CA.

Kiousis, S. (2001). Public trust or mistrust? Perceptions of media credibility in the information age. *Mass Communication and Society*, *4*(4), 381–403. doi:10.1207/S15327825MCS0404_4

Kohring, M., & Matthes, J. (2007). Trust in news media: Development and validation of a multidimensional scale. *Communication Research*, *34*(2), 231–252. doi:10.1177/0093650206298071

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. London, United Kingdom: Academic Press.

McCroskey, J. C., & Teven, J. J. (1999). Goodwill: A reexamination of the construct and its measurement. *Communication Monographs*, *66*(1), 90–103. doi:10.1080/03637759909376464

Meyer, P. (1988). Defining and measuring credibility of newspapers: Developing an index. *Journalism & Mass Communication Quarterly*, *65*(3), 567–574. doi:10.1177/107769908806500301

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*(1), 156–166. doi:10.1037/0033-2909.105.1.156

Morrison, J. T. (2009). Evaluating factor analysis decisions for scale design in communication research. *Communication Methods and Measures*, *3*(4), 195–215. doi:10.1080/19312450903378917

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. doi:10.1080/19312458.2011.624490

Nevitt, J., & Hancock, G. R. (2000). Improving the root mean square error of approximation for non-normal conditions in structural equation modeling. *The Journal of Experimental Education*, *68*(3), 251–268. doi:10.1080/00220970009600095

Park, A. (2014, April 1). Stem cell scientist guilty of falsifying data. *CNN*. Retrieved from http://www.cnn.com/2014/04/01/health/time-stem-cell/

Pe'er, E., Paolacci, G., Chandler, J., & Mueller, P. (2012). *Selectively recruiting participants from Amazon Mechanical Turk using Qualtrics*. Retrieved from http://ssrn.com/abstract=2100631 or http://dx.doi.org/10.2139/ssrn.2100631

Pew Research Center. (2012). Further decline in credibility ratings for most news organizations. Retrieved from http://www.people-press.org/files/2012/08/8-16-2012-Media-Believability1.pdf

Rimmer, T., & Weaver, D. (1987). Different questions, different answers? Media use and media credibility. *Journalism & Mass Communication Quarterly*, *64*(1), 28–44. doi:10.1177/107769908706400104

Roper, B. W. (1971). *An extended view of public attitudes toward television and other mass media, 1959-1971: A report by The Roper Organization, Inc*. New York, NY: Television Information Office.

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

Satorra, A., & Bentler, P. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*(2), 243–248. doi:10.1007/s11336-009-9135-y

Shaw, E. F. (1973). Media credibility: Taking the measure of a measure. *Journalism & Mass Communication Quarterly*, *50*(2), 306–311. doi:10.1177/107769907305000213

Sundar, S. S. (1999). Exploring receivers' criteria for perception of print and online news. *Journalism & Mass Communication Quarterly*, *76*(2), 373–386. doi:10.1177/107769909907600213

Wanta, W., & Hu, Y.-W. (1994). The effects of credibility, reliance, and exposure on media agenda-setting: A path analysis model. *Journalism & Mass Communication Quarterly*, *71*(1), 90–98. doi:10.1177/107769909407100109