University of Amsterdam

UvA-DARE (Digital Academic Repository)

Measuring credibility perceptions in CSR communication: A scale development to test readers' perceived credibility of CSR reports

Lock, I.; Seele, P.

Article

Measuring Credibility Perceptions in CSR Communication: A Scale Development to Test Readers' Perceived Credibility of CSR Reports

Irina Lock[1] and Peter Seele[2]

Abstract
Credibility is central to communication but often jeopardized by "credibility gaps." This is especially true for communication about corporate social responsibility (CSR). To date, no tool has been available to analyze stakeholders' credibility perceptions of CSR communication. This article presents a series of studies conducted to develop a scale to assess the perceived credibility of CSR reports, one of CSR communication's most important tools. The scale provides a novel operationalization of credibility using validity claims of Habermas's ideal speech situation as subdimensions. The scale development process, carried out in five studies including a literature review, a Delphi study, and three validation studies applying confirmatory factor analysis, resulted in the 16-item Perceived Credibility (PERCRED) scale. The scale shows convergent, discriminant, concurrent, and nomological validity and is the first validated measure for analyzing credibility perceptions of CSR reports.

[1]University of Amsterdam, The Netherlands
[2]Università della Svizzera italiana, Lugano, Switzerland

Corresponding Author:
Irina Lock, Amsterdam School of Communication Research (ASCoR), University of Amsterdam, Nieuwe Achtergracht 166, 1018WV Amsterdam, The Netherlands.
Email: lock@uva.nl

A credible communicator is one whom audiences believe; a credible message is one upon which recipients rely. In general, credibility is defined as "the capacity to be believed or believed in" (Oxford English Dictionary, 2016). However, it is not only individuals who aim for credibility in communication, but credibility is also central to communicating organizations. Ironically, this is often showcased by the lack of credibility, for instance, when corporate blunders such as the Volkswagen emission scandal evoke public attention (Clemente & Gabbioneta, 2017).

Credibility in organizational communication is often challenged because of discrepancies between words and deeds that result in "credibility" or "legitimacy" gaps: expectation mismatches between what stakeholders expect organizations to do and what they perceive organizations actually do (Sethi, 1975). Credibility gaps have opened up between companies and stakeholders because companies communicate about their ethical responsibilities regarding society and the environment (i.e., corporate social responsibility [CSR]) in a noncredible fashion (Illia, Zyglidopoulos, Romenti, Rodríguez-Cánovas, & del Valle Brena, 2013), jeopardizing their "license to operate" in society (Donaldson & Dunfee, 1999). Hence, stakeholders' expectations regarding CSR and companies' relative actions diverge (Dando & Swift, 2003). This article develops a measurement scale to study stakeholders' credibility perceptions of one particular CSR communication tool: the CSR report.

CSR reports are one of the most effective tools to communicate CSR (Hooghiemstra, 2000) and are thus most affected by the issue of credibility. Although credible CSR communication is fundamental to corporate legitimacy (Coombs, 1992; Seele & Lock, 2015), to date the "credibility gap" in CSR reporting (Dando & Swift, 2003) has not evoked researchers' attention above and beyond conceptual claims.

This study systematically and empirically studied credibility perceptions by developing a scale to analyze recipients' perceptions of the credibility of CSR reports. To develop this scale, we use a novel operationalization of credibility by conceptualizing the validity claims of the ideal speech situation (truth, sincerity, appropriateness, and understandability; Habermas, 1984) as subdimensions of credibility, showcasing the potential of Habermasian theory for empirical research (Forester, 1992). This credibility concept is consistent with current CSR and CSR communication theories that include the notion of legitimacy and are built on Habermasian theory (Scherer & Palazzo, 2011; Seele & Lock, 2015).

Credibility Concepts and a New Approach for CSR Communication

Dunbar et al. (2015) argue that "[c]redibility assessment is so critical to the process of communication that the construct of credibility occupies a vaulted position in the communication discipline" (p. 650). Despite its prominence, credibility is not defined clearly, often mistaken with trust, and operationalized in various, often vague, terms.

Communication scholarship conceptualizes credibility in three primary ways, related to source, recipient, and interaction models (Jackob, 2008). Source conceptualizations refer to the recipients' perceptions of the character traits of the speaker, such as his or her trustworthiness or expertise (Hellweg & Andersen, 1989). Using this conceptualization, for example, Sias and Wyers (2001) found that source credibility predicted employee information-seeking behavior. Examining the organization as source, Fombrun (1996) found that constituents' perceptions of an organization's credibility (i.e., trustworthiness and expertise) enhanced their perceptions of the corporate brand and predicted higher purchase intentions (Lafferty, 2007). Newell and Goldsmith (2001) also used this conceptualization in their development of a scale for corporate credibility.

Recipient conceptualizations have been used by marketing scholars measuring credibility as a form of CSR attribution that, together with CSR awareness, constitutes CSR beliefs or as a necessary attribute of CSR messages and communication channels (Du & Vieira, 2012).

Credibility is also considered a subcategory of trust, defined as "a characteristic which is attributed to individuals, institutions or their communicative products (verbal or written texts, audiovisual displays) by somebody (recipient) in relation to something (events, facts, etc.)" (Bentele & Nothaft, 2011, p. 216). However, trust and credibility are often not clearly separated (Stamm & Dube, 1994). Trustful relationships are central to communication between companies and their stakeholders with respect to CSR. The notion of organization–public relationships incorporates this trust as a central component (Heath, Waymer, & Palenchar, 2013). Thus, the more trust displayed toward an organization or its messages, the more credible the communication process is perceived. Credibility, thus, can also be seen as an outcome of trustful interactions between stakeholders and companies (Jackob, 2008).

There is general consensus among scholars that credibility is a multidimensional perception construct. As Jackob (2008) notes, "Credibility is a perceptual state, i.e., the outcome of an attribution process in which recipients of messages form judgments about their sources and therefore assess them as credible or not" (p. 1045). This definition and the above credibility

conceptualizations in the CSR and organizational communication literature demonstrate that most concepts fail to include all three dimensions of credibility: source, message, and recipient (Melican & Dixon, 2008). Given the procedural character of CSR communication, credibility in this context is interaction based and should thus cover all three dimensions. Furthermore, to grasp the meaning of the concept adequately, it must incorporate a deliberative and legitimacy-based notion.

The connection between dialogic communication and legitimacy was first established in deliberative democracy research in referring to political institutions' legitimacy (Chang & Jacobson, 2010). This idea was adopted by CSR scholars, who applied the Habermasian idea of deliberative democracy and communicative action to business–society relationships (Scherer & Palazzo, 2011): In a global economy where nation states are losing power, powerful multinational corporations step into the global governance gaps and assume responsibilities that were previously with the state. With these new "political" responsibilities, corporations are called upon to help solve public issues. In this postnational constellation, corporations' license to operate is established in communication with stakeholders regarding solutions for these problems (Scherer & Palazzo, 2011). Thus, corporations obtain their right to conduct business through moral legitimacy. Contrary to pragmatic and cognitive legitimacy (Suchman, 1995), moral legitimacy is continuously negotiated in the public sphere.

In this moralized (Castello, Morsing, & Schultz, 2013) communication, CSR reports are one of the most effective tools (Hooghiemstra, 2000) and play a crucial role as facilitators by helping companies respond to the information needs and expectations of stakeholders and, thus, manage their perceived legitimacy (Hahn & Kühnen, 2013). CSR reports can, therefore, be considered organizational artifacts that "allow people to do things, and inspire people to feel or react a certain way" (Rafaeli & Vilnai-Yavetz, 2005, p. 10). They are discrete, independent corporate editorial works that provide information about corporate responsibility and citizenship (Biedermann, 2008) and represent "a prime channel for companies to communicate their response to CR [corporate responsibility] issues" (Dawkins, 2005, p. 111). Thus, companies around the world report on CSR to foster their reputation, meet the needs of various stakeholders, and demonstrate their ethical commitment (Dando & Swift, 2003). The importance of CSR reporting is highlighted by the increasing regulation of this corporate disclosure by governments. Outside Europe (e.g., South Africa) and within the European Union (on the transnational and national level, for example, France and Denmark), governments have introduced laws that require companies to report on their CSR activities (European Union [EU], 2014; Ioannou & Serafeim, 2016).

This regulatory move illustrates the importance of CSR reporting and is necessary because CSR reports often suffer from "cherry picking" behavior regarding the reported items (Milne & Gray, 2013, p. 17) and weak comparability, rendering them noncredible communication tools in the eyes of many readers (Chen & Bouvain, 2009; Lock & Seele, 2016). Although at first glance the CSR report appears to be credible, it may not be perceived as such by the audience (Seele, 2016). This credibility gap in CSR reporting (Dando & Swift, 2003) harms both sides: Stakeholders cannot satisfy their information needs with regard to CSR, while companies, because of stakeholders' lack of trust and their own low credibility, risk their license to operate.

In summary, credibility is a basis of legitimacy (Coombs, 1992) and an outcome of reporting quality (Perera Aldama, Awad Amar, & Winicki Trostianki, 2009). The crucial role of credibility calls for a conceptualization that accounts for this legitimacy notion. Using political CSR theory (Scherer & Palazzo, 2011) grounded on Habermas's (1984, 1996) communicative action and deliberative democracy theories, the present study provides such a new approach to credibility.

Operationalizing Habermas's Validity Claims as Subdimensions of Credibility

Following Forester (1992), "Habermas's sociological analysis of communicative action . . . has a vast and yet unrealized potential for concrete social and political research" (p. 47f). According to Habermas (1984), there are two forms of social action: strategic and communicative. In the teleological model of strategic action, actors aim at achieving success. In contrast, communicative action is oriented toward agreement and mutual understanding between the participants. This notion of agreement and dialogue matches well with the Habermasian notion of political deliberation as proposed in the concept of the new political role of corporations (Scherer & Palazzo, 2011).

In communicative action, "[r]eaching an understanding functions as a mechanism for coordinating actions through the participants coming to an agreement concerning the claimed *validity* of their utterances, that is, through intersubjectively recognising the validity claims they reciprocally raise" (Habermas, 1984, p. 163; emphasis in original). These four validity claims are "communicative presuppositions" (Anderson, 1985, p. 86) in an "ideal speech situation" that can be applied to "real-world cases ranging from the general to the specific" (Dryzek, 1995, p. 104). They refer to the truth, sincerity, appropriateness, and understandability of communication as follows:

- *Truth*: objective truth of propositions or factual truth
- *Sincerity*: subjective beliefs underlying the statements
- *Appropriateness*: communication that is appropriate in its normative context
- *Understandability*: ensuring that the statement is understandable to the reader (Habermas, 1984).

In everyday communication, these claims are underlying and implicit; in argument or discourse, they become explicit (Johnson, 1993) and give power to the best argument (Chappell, 2012). Only if the speaker's statements are factually true, the speaker is sincere, the statement is appropriate in its normative context, and the actors understand what is said can we speak of a situation of communicative action that is aimed at mutual understanding and agreement, and thus addresses credibility.

Validity claims are "central" to the theory of communicative action (Habermas, 1984, p. 10) and lead to the trustfulness needed to arrive at credibility. Senders (companies issuing CSR reports) as well as recipients (stakeholders and the general public) of CSR communication must redeem these four presuppositions in the discourse. As Johnson (1993) holds, "[c]ommunicative action derives its force from the potential for rational agreement embodied in validity claims" (p. 75). Consistent with Habermas, we argue that if all four claims are fulfilled by the actors, credibility between the parties can be (re-)established.

Limitations of Existing Credibility Scales

In communication science, credibility is often context specific, for instance, in studies on Internet credibility (Melican & Dixon, 2008), credibility in interviewing (Dunbar et al., 2015), or media credibility (Stamm & Dube, 1994). The notion that credibility depends primarily on the source of communication (Hellweg & Andersen, 1989) has often been taken as a reference or starting point. Research from neighboring disciplines has engaged in similar attempts to capture this concept.

Marketing researchers developed a scale of credibility regarding attractiveness (Newell & Goldsmith, 2001) or regarding someone's trustworthiness, and thus his or her honesty and sincerity (Ohanian, 1991). Williams and Drolet (2005) developed a measure for the general credibility of an advertisement. Keller and Aaker's (1992) scale assesses the credibility of a company as a source of communication. However, none of these scales incorporate the idea of consensus and legitimacy that comprise a Habermasian operationalization, and the three dimensions of source, recipient, and message have been neglected.

Reliable and applicable scales for the validity claims of truth, sincerity, appropriateness, and understandability are also lacking.[1] Few scholars have attempted to operationalize them for quantitative empirical research, however highly context specific (Chang & Jacobson, 2010). Qualitative research using the theory of communicative action has been conducted primarily in deliberative democracy studies in the political sciences (e.g., Forester, 1992).

In sum, to the best of our knowledge, (a) no credibility perception measurement of communicational artifacts exists, (b) there is no specific instrument for the context of CSR and CSR reports, and (c) a comprehensive measurement tool for the four dimensions of the Habermasian ideal speech situation is lacking. To address these limitations, the set of studies reported below study developed and validated scales for the four constructs.

Methodological Approach of Scale Development and Validation

This project followed established scale development process methods (Hinkin, 1995; Jian et al., 2014) to develop scales that measure the four validity claims. As illustrated in Figure 1, scale development is an iterative process that in this case took five studies, reducing the initial number of scale items from 98 to 16.

Item Generation—Study 1: Literature Review

We used a deductive approach to generate items based on the relevant literature. To collect existing definitions of the four constructs of truth, sincerity, appropriateness, and understandability, a keyword search was performed for the terms "Habermas" + "communicative action" in six databases, including publications from various academic disciplines: *Business Source Premier, Communication & Mass Media Complete, EconLit, Psychology & Behavioral Sciences, PsychInfo*, and *Science Direct*. The references in the selected articles were then checked, via a snowballing technique, to find other relevant literature. We also searched for scales of (perceived) truth, sincerity, appropriateness, and understandability within the same databases. A total of 161 articles were reviewed in depth for definitions of validity claims and for scale measures. Using the mind-mapping technique (Eppler, 2006), we first defined subcategories within the constructs that emerged from the definitions. We then refined these subcategories, arriving at four subdimensions for understandability, seven for truth, 10 for sincerity, and four for appropriateness. We began to develop scale items based on these subcategories and the definitions from the literature. The written items were double-checked by a scale development expert for style and format. This process resulted in 98 items for the four
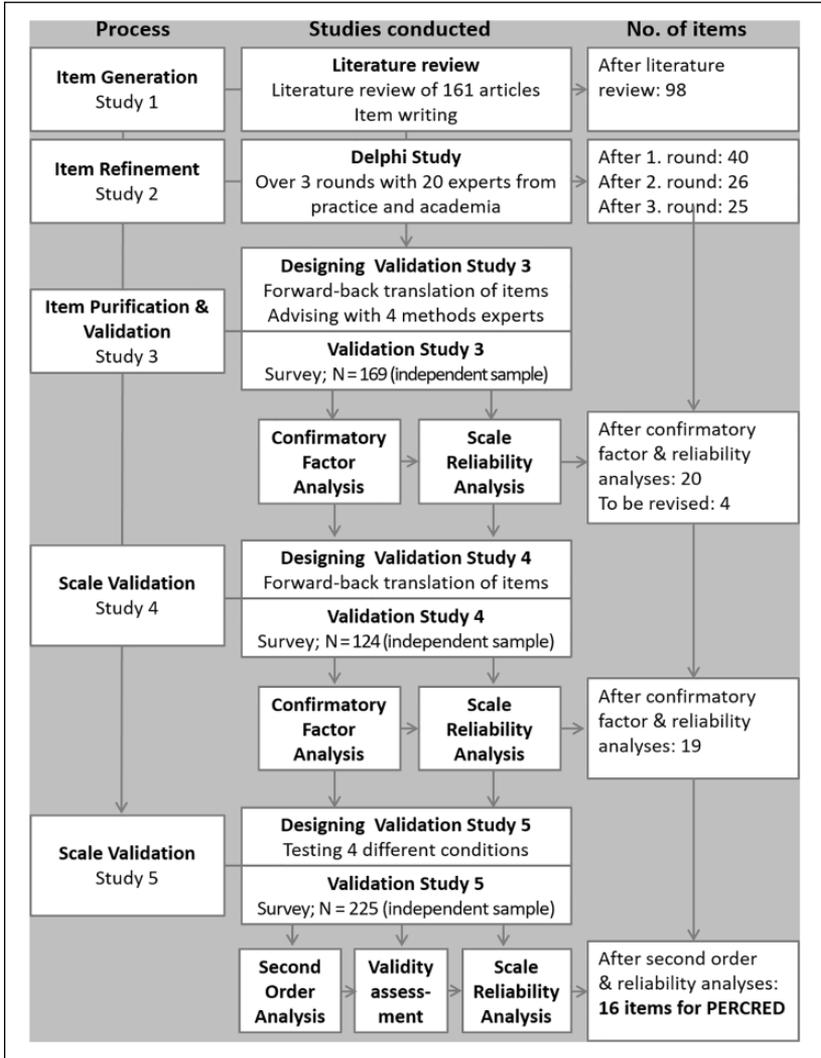
Figure 1. The scale development process to arrive at the 16-item PERCRED scale.
Note. PERCRED = Perceived Credibility.

constructs to be tested in a Delphi study (Linstone & Turoff, 1975): 18 for understandability, 28 for truth, 34 for sincerity, and 18 for appropriateness.

Item Refinement—Study 2: Delphi Study

We used the Delphi method to ensure content validity of the generated items and to refine the scale. Twenty-three experts in the fields of CSR, CSR communication, CSR reporting, and Habermas and practitioners in the corporate and consultancy realms were contacted personally via e-mail; 20 agreed to participate in the study. Among them were seven academics with expertise in CSR communication and reporting, five experts on Habermas, four on CSR in general, and four practitioners. We conducted three rounds over a 2-month period.

In the first round, 17 participants rated the 98 items with respect to their adequacy to measure each of the four concepts on a scale from 1, *not at all adequate*, to 5, *very adequate* (to measure the construct). We used a two-step analytic process: First, all items that scored above the construct mean (e.g., for truth) were considered to be included in Round 2; second, the comments of the participants on every item were analyzed qualitatively. Based on these two steps, the number of items was reduced to 40.

Prior to the second round, the items were edited by a professional English language editor. Fifteen members of the expert panel participated in Round 2. The questionnaire format was the same as in Round 1. The remaining 26 items were refined according to the experts' comments.

In the third round, 14 participants ranked the items per construct according to their adequacy for measuring it. One more item was deleted and two items were worded negatively (inverse scores), which resulted in the final item pool of 25 items.

Item Purification and Validation—Study 3

Designing Validation Study 3. The item-generation process was conducted in English. However, the validation study was carried out at a German university. To improve the study's validity through enhanced understandability, the authors translated the items into German. As "there is no single perfect translation technique" (Maneesriwongul & Dixon, 2004, p. 183), a forward–back translation method was used with two professional translators, following Beaton, Bombardier, Guillemin, and Ferraz (2000).

Because our goal was to develop a measure for the perception of CSR reports, we needed communicational artifacts for these perceptions. CSR reports are large documents, ranging from 20 to more than 300 pages; therefore, we had participants rate one specific section. We consulted with two scale development experts to design the validation study. The population for this questionnaire consisted of undergraduate business students from a German

university. Because we assumed most of these students will be future employees of companies, the employee section of the CSR report was used to test credibility perceptions.

From an analysis of CSR reports evaluating content credibility (Lock & Seele, 2016), whose credibility evaluation is equally based on an analysis of the CSR reports along Habermasian validity claims, we selected a credible (scoring high on the four validity claim dimensions aggregated) and a noncredible report (scoring low on the four validity claim dimensions aggregated), from both the same country (Germany) and sector (consumer goods). We focused on one country and one sector only as CSR reporting is sensitive when it comes to cultural and industry bias (Chen & Bouvain, 2009). Furthermore, a consumer goods company was selected under the assumption that the business students were regular consumers of these products, which was also tested in the questionnaire. The CSR reports were anonymized to control for potential bias for or against specific brands. Comprehension questions and attention checks were built into the questionnaire, and the items were randomized to avoid response patterns. The questionnaire was pretested with a convenience sample of 12 people.

Data collection and sample. Undergraduate business students from a German university completed an online or in-class survey. The survey links (in a laptop as well as a smartphone version) were sent via e-mail to 355 students. Half (randomly assigned) received links to the survey with the credible CSR report, the other half to the survey with the noncredible report to validate the items across conditions. To trigger interest and increase the response rate, the survey used a lottery scheme, where participants could win 10 online shopping vouchers. A total of 193 students answered the survey, and 169 responses were valid.

The mean sample age was 19.91 years, with 34.91% being female participants. More than half of the sample (56.21%) had less than 6 months working experience. Regarding consumption behavior, 95.8% of the participants buy personal care products and 79.2% purchase household goods once or more than once a month, indicating that most members of the sample are regular customers of consumer goods companies.

Confirmatory factor analysis (CFA) and scale reliability. Based on Habermasian theory, a first-order model with the four validity claims (understandability, truth, sincerity, appropriateness) as latent variables and their generated items as observed variables (understandability: U1-U6; truth: T1-T7; sincerity: S1-S7; appropriateness: A1-A5) was estimated with CFA (Brown, 2015; Byrne, 2001). Because the data are nonnormally distributed according to

Mardia's test (kurtosis = 73.85, c.r. = 13.59), maximum-likelihood (ML) bootstrapping on 500 samples (confidence interval [CI] = 0.95) is used in AMOS 23. Given the relationship of the validity claims as subdimensions of perceived credibility, the factors are permitted to be correlated. In the model, all items were significantly related to their latent factors ($p < .01$) and display standardized regression weights above 0.6, except Item A1 (0.57). All four factors show significant covariances (Brown, 2015).

Multiple goodness-of-fit indices (GFIs) are used to establish model fit, although absolute thresholds as suggested by Hu and Bentler (1999) are often too conservative for applied research, and there is little consensus on the suitability of these indices to evaluate model fit (Brown, 2015). The estimated four-factor model showed the following model fit results: χ^2 (270) = 570.85, $p < .001$, root mean square error of approximation (RMSEA) = 0.08, CMIN/DF = 2.11, comparative fit index (CFI) = 0.81, Tucker–Lewis index (TLI) = 0.79, GFI = 0.78, parsimony normed comparative fit index (PCFI) = 0.75, and standardized root mean square residual (SRMR) = 0.18. RMSEA of 0.08 and the minimum discrepancy divided by degrees of freedom (CMIN/DF) below 3 suggest adequate model fit, whereas CFI and TLI should be above 0.95, GFI larger than 0.9, PCFI greater than 0.8, and SRMR close to 0.08 (Brown, 2015). The Bollen–Stine bootstrap p value was significant at $p < .01$. Thus, several parameters were inspected to increase model fit.

Model fitting in all studies of this scale development followed three steps: (a) inspection of lowest standardized parameter estimates, significance levels of standardized and unstandardized estimates and their standard errors, and significance levels of factor covariances (Brown, 2015; Byrne, 2001); (b) modification indices (MI): large error covariances beginning at values above 4 (Brown, 2015); and (c) review of standardized residual covariances exceeding the absolute value of 2.0. To ensure content validity, decisions regarding item deletions were only taken in line with the theoretical underpinnings of the study. Model fitting was stopped either when the three steps above did not suggest any changes in the model or when the model fit parameters did not show better solutions after fitting, or both.

Item A1 was deleted due to its relatively low loading of 0.57 on appropriateness, its standardized residuals, and given conceptual considerations (see section below). Second, MI above 4 suggested covarying several error terms in the model. With these modifications, the model was run again and resulted in a better fitting model with χ^2 (236) = 426.10, $p < .001$, RMSEA = 0.07, CMIN/DF = 1.81, CFI = 0.87, TLI = 0.85, GFI = 0.83, PCFI = 0.75, and SRMR = 0.18. All items loaded significantly on their factors ($p < .01$) and above 0.6, except A3, which resulted in 0.56. The latent variables showed significant covariances ($p < .01$). The appropriateness factor correlated above

.85 with all other factors, and the correlation between truth and sincerity also exceeded this threshold. This can indicate problems with discriminant validity that must be observed more closely in subsequent validation steps. Table 1 displays the first-order model with standardized and unstandardized item loadings, standard deviations and errors, means, and significance levels. Scale reliability of the four subscales was tested with Cronbach's alpha (see Table 1) and deemed appropriate for first-order models (MacKenzie, Podsakoff, & Podsakoff, 2011).

Inspection of standardized residuals indicated that U2, U3, U4, U6, and A5 consistently exceeded the absolute value of 2.0. At this point, to avoid overfitting of the data purely based on empirical results (Brown, 2015) at this early stage of scale development, we decided not to delete further items. Instead, we referred back to theory and examined the scale items more closely to subject these modified items to another validation study with an independent sample. In particular, the appropriateness items were reformulated, and the items for understandability were examined more closely.

Scale Validation—Study 4

Designing Validation Study 4. The CFA results in Validation Study 3 indicated a poor fit of some of the understandability items. Based on the item-generation process, understandability has four theoretical subdimensions: technical comprehensibility (Items U1-U3), "talk the same language" (U4), meaning (U5), and completeness (U6). Conceptually, and based on the Delphi results, all three items for technical comprehensibility are equally comprehensible and do not exhibit underlying linguistic biases that explain the empirical results. Therefore, all three were retained for Validation Study 4. Item U4 ("The text uses too much specialist terminology") might have been subject to misunderstanding: Even if a text uses too many technical terms, it can still be understood by the reader. This item was supposed to reflect the subdimension labeled "sender and recipient talk the same language." Thus, the item was reformulated and subsequently tested in Validation Study 4 (U4: "The text uses too much specialist terminology that I do not understand"). As Items U5 and U6 represent important subdimensions of the Understandability scale, we decided to test them again in the next validation study.

All items were revised for perceived appropriateness. Emerging from the literature review and the following item generation (Stage 1), we defined four subdimensions of this validity claim: culture, norms, context, and legitimacy. During the three stages of the Delphi study, all items referring to culture were evaluated as being inadequate to measure appropriateness by the experts and thus dropped. The final scale items tested for this claim referred to the

Table 1. First-Order Model After Validation Study 3—Standardized Estimates With *p* Value, Mean, and Standard Deviation, and Unstandardized Estimates With Standard Errors, and Cronbach's α.

Factor	Item	Standardized	M	SD	Unstandardized	SE	p
Truth	T1	0.77	3.22	0.98	1.00	0.12	<.01
$\alpha = .83$	T2	0.81	3.37	0.90	0.98	0.14	<.01
	T3	0.71	3.56	0.86	0.79	0.12	<.01
	T4	0.72	3.39	0.93	0.87	0.13	<.01
	T5	0.64	2.37	0.97	0.79	0.11	<.01
	T6	0.74	3.11	1.02	—[a]	—[a]	—[a]
	T7 (recoded)	0.68	3.07	1.19	1.05	0.14	<.01
Sincerity	S1	0.80	3.38	0.79	0.91	0.20	<.01
$\alpha = .80$	S2	0.77	3.46	0.93	1.03	0.22	<.01
	S3	0.78	3.57	0.88	0.99	0.22	<.01
	S4	0.75	3.57	0.86	0.93	0.19	<.01
	S5	0.67	3.38	0.96	0.89	0.23	<.01
	S6	0.61	3.29	0.93	0.78	0.20	<.01
	S7	0.62	3.63	1.19	—[a]	—[a]	—[a]
Appropriateness	A2	0.62	3.56	0.91	0.65	0.12	<.01
$\alpha = .52$	A3	0.56	3.52	0.78	0.50	0.10	<.01
	A4	0.66	3.71	1.06	0.82	0.12	<.01
	A5	0.71	3.70	0.97	—[a]	—[a]	—[a]
Understandability	U1	0.67	4.49	0.80	0.51	0.10	<.01
$\alpha = .84$	U2	0.88	3.86	0.99	0.92	0.08	<.01
	U3	0.92	4.09	0.91	0.90	0.06	<.01
	U4 (recoded)	0.66	4.11	0.94	0.58	0.09	<.01
	U5	0.65	4.27	0.87	0.52	0.10	<.01
	U6	0.90	4.07	0.97	—[a]	—[a]	—[a]

[a]Item was constrained to 1.

subdimension norms (A1, A2), context (A3, A4), and legitimacy (A5). A1 was deleted based on its low loadings on the latent factor, consistent with conceptual considerations: Already in the Delphi study, respondents rated A1 as subjective, and two Delphi participants commented critically: "The item on expectation can elicit a cynical as well as hopeful response. I think it will make data interpretation difficult."

Second, A2 ("The text is ethically appropriate for a CSR report") was reformulated as readers might not have understood what "ethically appropriate" meant. In this case, it refers to the normative context which may be unique to the company sector. The company's industry has a strong impact on the materiality of what companies report with regard to CSR and what stakeholders might perceive as being important (Chen & Bouvain, 2009). Thus,

the authors decided to integrate the sector specificity into this item and reformulated it: "The CSR report fits to the context of the _____ industry and its social and environmental challenges." This item is now referred to as A1.

Third, regarding the context of the communication act, we generated two items, A3 ("The text is embedded in the context of CSR") and A4, an inverse item ("The text is not appropriate to the situation"). A3 might seem tautological—a CSR report is by definition embedded in the context of CSR. A reformulated item reflecting the initially intended theme is "I think that the topic presented in the text relates to CSR" (now: A2). A4 opens the question as to which situation the item refers. In the reformulated item, the situation is specified as the communication process between company and recipient when reading a report on CSR topics. A reformulated item that better describes this situation would be, "As a reader of this CSR report, I feel that the text addresses CSR issues well" (now: A3). Finally, A5 was intended to reflect the legitimacy of the sender of the communication: "The author speaks legitimately on behalf of the company." However, often, it is not clear to the reader at first glance who the author of a CSR report is. Therefore, the new item avoids the question of authorship, and shifts focus to the text and its legitimacy: "I think the text rightfully represents the company" (now: A4).

In sum, 24 items were included in Validation Study 4. The reformulated appropriateness items and U4 again underwent a forward–back translation procedure.

Data collection and sample. Data were collected using an online and in-class survey among business students from the same university, using the same study design and procedure as in Validation Study 3. The samples of Validation Studies 1 and 2 were independent. A total 133 students completed the survey, and 124 responses were valid, exceeding the threshold of $N = 100$ (MacCallum, Widaman, Preacher, & Hong, 2001). Mean age of participants was 19.92 years, with 29.84% being female; 62.90% of participants reported less than 6 months working experience. Rather than measuring consumption behavior, the students were asked about their intention to purchase a product from the company on a 7-point scale (Yin, 1990). The weighted average of the three items resulted in a mediocre intention to purchase of $M = 4.5$ ($SD = 1.4$).

CFA and scale reliability. The goal of this second validation study was to confirm the four-factor structure found in Validation Study 3 and to test the reformulated items. CFA was performed on the 24 items and four factors using ML estimation with bootstrapping on 500 samples (CI = 0.95) in AMOS 23, as Mardia's test for multivariate normality showed that the data are not normally distributed (kurtosis = 60.01, c.r. = 11.83).

To arrive at the final model, three iterations of model fitting were conducted, with covarying of errors, and with Items A2, U4, U1, and U5 deleted due to relatively low loadings <0.5 and after analysis of their standardized residual covariances. Almost all items of the final model showed significant ($p \leq .01$) loadings above 0.5 with their four latent factors (see Table 3). Only A3 loaded at 0.48 on appropriateness. However, given that this item is conceptually important as it represents the subdimension context, the authors refrained from excluding it at this point of scale development. The Bollen–Stine bootstrap of the final model was not significant ($p > .01$), indicating that the model fits the data well (Byrne, 2001).

Three of the four factors showed significant covariance ($p < .05$). Understandability, however, showed a nonsignificant covariance with truth (see Table 2). Also, one of its items, U2, loaded slightly above 1, which could be problematic (see Table 3).

The final model contained 19 items, and its chi-square test was significant ($p < .05$) with $\chi^2(137) = 172.77$. However, relying on this test alone as a measure of the goodness of fit is not recommended. The model showed very good values for other GFIs: CFI = 0.96, TLI = 0.95, CMIN/DF = 1.26, PCFI = 0.77, RMSEA = 0.05, and SRMR = 0.06 (Brown, 2015). No significant differences in credibility perceptions (based on this scale) between the two reports resulted from an independent samples t test, $t(122) = -0.60, p > .05$, which confirms the credibility gap; the CSR reports are not perceived to be as credible as they were rated in a previous study.

In general, the CFA partially confirmed the results of the CFA of Validation Study 3 and the four-factor structure with the newly developed items for appropriateness and (see Table 3). Moreover, the results indicate that understandability is the weakest of the four-factor structures. This might be due to the student sample. One can reasonably assume that students have higher reading skills than the general population. Hence, before evaluating a second-order model with these results, another step of validation with a nonstudent population was conducted to test whether understandability remained a weak construct in a more general population. Therefore, all items of the understandability factor were included again in Validation Study 5.

Scale Validation—Study 5

Designing Validation Study 5. Validation study 5 had four goals: (a) to validate the four-factor structure in a more general population, (b) to validate the scale with a broader sample of CSR reports, (c) to establish a second-order model showing how the four latent factors relate to Perceived Credibility (PERCRED), and (d)

Table 2. Covariances of Latent Factors After Validation Study 4.

Pair			Estimate	SE	p value
Understandability	↔	Truth	.10	0.08	.07
Understandability	↔	Sincerity	.15	0.09	.02
Understandability	↔	Appropriateness	.15	0.07	.03
Truth	↔	Appropriateness	.41	0.08	.00
Sincerity	↔	Truth	.45	0.08	.00
Sincerity	↔	Appropriateness	.43	0.08	.00

Table 3. First-Order Model After Validation Study 4—Standardized Estimates With p Value, Mean, and Standard Deviation, and Unstandardized Estimates With Standard Errors, and Cronbach's α.

Factor	Item	Standardized	M	SD	Unstandardized	SE	p
Truth	T1	0.53	3.42	1.08	0.76	0.16	≤.01
$\alpha = .81$	T2	0.68	3.30	0.89	0.80	0.13	≤.01
	T3	0.74	3.43	0.89	0.87	0.13	≤.01
	T4	0.72	3.21	0.98	0.94	0.15	≤.01
	T5	0.57	2.58	1.10	0.83	0.13	≤.01
	T6	0.72	3.25	1.05	—[a]	—[a]	—[a]
Sincerity	S1	0.62	3.38	0.85	0.78	0.18	≤.01
$\alpha = .82$	S2	0.65	3.11	0.90	0.87	0.24	≤.01
	S3	0.65	3.13	0.95	0.92	0.17	≤.01
	S4	0.57	3.11	0.96	0.83	0.20	≤.01
	S5	0.56	3.19	0.96	0.81	0.21	≤.01
	S6	0.66	3.27	1.02	1.00	0.20	≤.01
	S7	0.59	3.43	1.14	—[a]	—[a]	—[a]
Appropriateness	A1	0.52	3.19	0.97	0.82	0.15	≤.01
$\alpha = .65$	A3	0.48	2.99	0.95	0.73	0.15	≤.01
	A4	0.69	3.10	0.90	—[a]	—[a]	—[a]
Understandability	U2	1.00	3.78	1.06	1.88	1.37	≤.01
$\alpha = .78$	U3	0.98	3.93	1.01	1.75	1.26	≤.01
	U6	0.50	3.93	1.13	—[a]	—[a]	—[a]

[a]Item was constrained to 1.

to demonstrate convergent, discriminant, concurrent, and nomological validity of the final scale.

We continued validating the scale with a German-speaking sample, using the SoSci Panel, which contains a participant pool of more than 90,000 people (Leiner, 2017). The CSR report sample was enlarged by two more reports to test the degree of validity for the scale, not only for German consumer

goods reports and their employee sections but also for other sections, formats, and cultures.

Borrowing from the same database (Lock & Seele, 2016), two more reports were chosen: one noncredible report in a combined report format from the real-estate industry in Switzerland, and another credible report from the oil and gas sector in Austria in the form of a stand-alone publication. The excerpts were a section with general information on the report for the Swiss publication and a section on the environment for the Austrian report. Thus, the scale could be validated in three different national contexts (Austria, Germany, Switzerland), with three different topics (employees, general information, environment), in two different reporting formats (stand-alone and combined reports), and with two different levels of credibility (credible and noncredible). The names of the companies were anonymized. The study was pretested with six people in addition to undergoing the panel's review process to ensure technical functionality. The scale items were randomized to avoid response patterns.

In addition, validated scales were included in this study to describe the nomological net: trust (Kim, 2001), trustworthiness (McCroskey & Teven, 1999), corporate distrust (Adams, Highhouse, & Zickar, 2010), corporate credibility (Newell & Goldsmith, 2001), and intention to purchase (Yin, 1990).

Data collection and sample. A sample size of 300 respondents was targeted with SoSci Panel; data were collected in three rounds. The final sample of $N = 225$ (after comprehension and attention checks) had a mean age of 38.75 years, and 55.56% were female. Most participants (81.33%) have worked for more than 6 months.

Eighty-six percent of the German population have completed secondary education; in this sample, this figure is up to 97.8%. Seventy-three percent of Germans have a paid job; in this study's sample, only 62.2% are employed (Organisation for Economic Co-operation and Development [OECD], 2015). Furthermore, the sample for this study is younger than the median age of 46.1 years, and more females are represented than in the general population (ratio male/female = 1.06; Central Intelligence Agency [CIA], 2015).

Second-order analysis and scale reliability. To confirm the four-factor structure from previous studies and determine whether these four factors reflect the higher order construct PERCRED, we conducted a second-order analysis in AMOS 23 using ML estimation with bootstrapping on 500 samples (CI = 0.95), as the data did not satisfy the requirements for multivariate normality (kurtosis = 59.68, c.r. = 18.65). All items of the previous study's final model

plus all understandability items from Validation Study 3 were tested to analyze whether the factor shows different results with a nonstudent sample. A total of 22 items were tested.

A second-order model with PERCRED as the higher order construct and perceived truth (six items), sincerity (seven items), appropriateness (three items), and understandability (six items) was tested. We referred to the theory represented by the subdimensions of the four constructs to ensure content validity. The first model showed insufficient model fit statistics. The MI suggested deleting Items S5 and U4 due to their high standardized residual covariance. The second 20-item model still displayed weak model fit. The MI statistics reported consistently bad values for the standardized residual covariances for S1, S6, and T1, which led to deletion of these items. The MI's standardized residual covariances demonstrated that A3, A4, and S7 could be deleted as well. However, examining the theoretically deduced subdimensions, we observed that deleting these items would eliminate important subdimensions of sincerity and appropriateness. Appropriateness covers three subdimensions: norms, context, and legitimacy. By deleting A3 and A4, context and legitimacy would be lost; thus, these items were retained. Furthermore, sincerity is defined by trustworthiness and intentions. Deleting S7 would eliminate the important dimension of trustworthiness; hence, it was retained in the scale. The subdimension intention is represented by three items, S2, S3, and S4, which potentially leads to an overemphasis. Linguistically, the wording of S2 and S3 appears to be closer to the idea of intention, which is why S4 was deleted from the scale.

The final second-order model contained four factors that reflect PERCRED employing 16 items. Figure 2 shows the item and factor loadings. All item loadings are above 0.5 and significant ($p < .001$). Moreover, the four factors load significantly ($p < .001$) on the higher order construct PERCRED. Again, understandability had a low loading of 0.44, while truth loaded at 0.99, sincerity at 0.97, and appropriateness at 0.79. The overall model fit was good with $\chi^2(86) = 158.91$, $p < .001$, Bollen–Stine bootstrap $p > .01$, CFI = 0.96, TLI = 0.94, CMIN/DF = 1.81, PCFI = 0.70, RMSEA = 0.06, and SRMR = 0.06 (Brown, 2015). Reliability for this PERCRED scale was high ($\alpha = .88$); reliability would not be improved by deleting any items. Composite scale reliability of the PERCRED scale is CR = .94 (Raykov, 1997). In conclusion, this 16-item scale appears to be a good representation of PERCRED (see Table 4).

Convergent validity. The PERCRED scale demonstrated convergent validity because (a) all four factors are significantly related ($p < .001$) to the higher order construct PERCRED and loadings are above 0.5 except for understandability (0.44); (b) all items relate significantly ($p < .001$) at above .05 to their
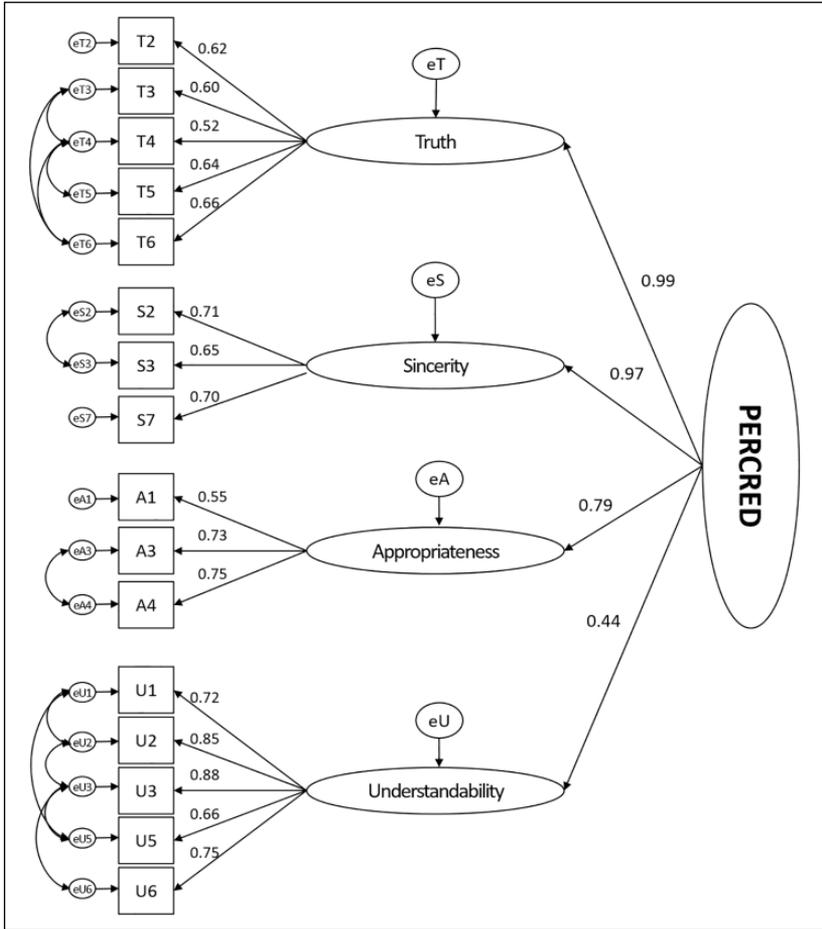
Figure 2. Final second-order model of perceived credibility after Validation Study 5 (parameter estimates are standardized).

respective first-order factors, which verifies the assumed relationships between the constructs (MacKenzie et al., 2011); (c) scale reliability is confirmed with Cronbach's α and composite reliability (Hair, Tatham, Anderson, & Black, 2006); and (4) the average variance extracted (AVE) is 0.68, above the desired threshold of 0.5 (MacKenzie et al., 2011).

Discriminant validity. Discriminant validity is given if the indicators of the construct "are distinguishable from the indicators of other constructs"

Table 4. Final Scale Items in English.

	Subdimensions of validity claim		Items final model
Truth	Accuracy	T2	I think that the statements in the text are accurate.
	Truthfulness	T3	I think that the claims made in the text are correct.
	Truthfulness	T4	I am confident that the statements are true.
	Facts	T5	I think that the text uses the best evidence at hand.
	Facts	T6	The arguments are justified by the facts in the text.
Sincerity	Intentions	S2	The text reflects the genuine intentions of the company.
	Intentions	S3	I think that the company's intentions correspond with the text.
	Trustworthiness	S7	The text is not misleading.
Appropriateness	Norms	A1	The CSR report fits to the context of the _____ industry and its social and environmental challenges.
	Context	A3	As a reader of this CSR report, I feel that the text addresses CSR issues well.
	Legitimacy	A4	I think the text rightfully represents the company.
Understandability	Technical	U1	I understand the text.
	Technical	U2	The text is clearly written.
	Technical	U3	The text is written in an understandable way.
	Meaning	U5	I understand the meaning of the text.
	Completeness	U6	The text is easy to read.

Note. The descriptive statistics of the PERCRED scale (all cases aggregated) are as follows: $M = 3.22$, $SD = 0.61$, Minimum = 1.44, and Maximum = 4.88. Hence, overall, participants perceived this study's CSR reports as credible at a mediocre level. PERCRED = Perceived Credibility; CSR = corporate social responsibility.

(MacKenzie et al., 2011, p. 317) and is established if the model of the PER-CRED scale is sufficiently distinct from other models. Corporate credibility (Newell & Goldsmith, 2001) and PERCRED are distinct constructs that measure different dimensions of one concept, credibility. The former focuses on the perceived credibility of the company, whereas PERCRED measures

credibility perceptions of the company's communicational artifact, the CSR report. Both constructs are significantly positively correlated r = .63, p < .01.

To test discriminant validity of the PERCRED scale, two models were estimated in which PERCRED and corporate credibility correlate. The unconstrained CFA model, $\chi^2(235) = 660.54$, $p < .01$, and the constrained model, $\chi^2(236) = 674.74$, $p < .01$, where the correlation between PERCRED and corporate credibility was constrained to 1, differed significantly in chi-square values. According to a chi-square difference test, $\Delta\chi^2(236) = 14.2$, $p < .01$, the unconstrained model shows a significantly lower chi-square value than the constrained model, which provides evidence of discriminant validity of the PERCRED scale (Fornell & Larcker, 1981).

Concurrent validity. Concurrent validity of the PERCRED scale was established by examining the relationship between the developed construct and the outcome variable intention to purchase (Yin, 1990), a construct largely used in marketing research. Yin (1990) developed a 7-point bipolar scale consisting of three items to measure this construct. Lafferty and Goldsmith (1999) found that corporate credibility has a significant effect on purchase intentions. The perception of credible CSR reports also affects the willingness to purchase a company's product because credibility decreases distrust, which negatively affects consumers' purchase intentions (Adams et al., 2010). A linear regression with PERCRED as the dependent variable and intention to purchase as the independent variable was performed and was significant $R^2 = .11$, $\beta = .34$, $B = .13$, $SE = 0.03$), $t(1) = 4.92$, $p < .001$.

Nomological net. Cronbach and Meehl (1955) stated, "To 'make clear what something is' means to set forth the laws in which it occurs. We shall refer to the interlocking system of laws which constitute a theory as a nomological network" (p. 290). To analyze a nomological network, MacKenzie et al. (2011) suggested that researchers "test whether the focal construct is significantly related to other constructs hypothesized to be in its nomological network" (p. A4). Thus, regression analyses were performed with the four concepts of trust, distrust, trustworthiness, and corporate credibility to illustrate the relationships of PERCRED within its nomological network. *Trust* was conceptualized as one construct of the organization–public relationship scale developed by Kim (2001) and consists of four items measured on a 7-point scale. Trust and PERCRED are assumed to demonstrate a positive relationship. PERCRED might even partially overlap with the trust construct. In fact, regression analysis confirms this assumption (see Table 5). The validity claim of sincerity is sometimes also translated as *trustworthiness*; it refers to the perceived honesty of communication (Habermas, 1984). Trustworthiness is likely positively

Table 5. Evidence of Nomological Validity.

Construct (no. items)	Source	n	α	R^2	β	B	SE	T	df	p
Trust (4)	Kim (2001)	224	.83	.30	.55	0.42	0.04	9.73	1	<.001
Distrust (13)	Adams, Highhouse, and Zickar (2010)	224	.92	.08	−.29	−0.25	0.06	−4.49	1	<.001
Trustworthiness (6)	McCroskey and Teven (1999)	174	.90	.44	.66	0.41	0.04	11.56	1	<.001
Corporate credibility (8)	Newell and Goldsmith (2001)	223	.83	.42	.65	0.46	0.04	12.75	1	<.001
Intention to purchase (3)	Yin (1990)	194	.93	.11	.34	0.13	0.03	4.92	1	<.001

related to PERCRED, although sincerity covers more subdimensions than trustworthiness. McCroskey and Teven (1999) conceptualized trustworthiness as part of the Aristotelian concept of goodwill. It is measured on a single-word six-item bipolar scale with a scale length of 7, which shows a significantly positive relationship with PERCRED, as evidenced by the regression analysis (see Table 5). *Distrust* is assumed to have a negative relationship with the PERCRED scale. This notion was measured by Adams et al. (2010) with 13 items on a 5-point Likert-type scale. Indeed, regression analyses confirmed a significantly negative relationship between PERCRED and distrust (see Table 5). Newell and Goldsmith (2001) understood *corporate credibility* as the amount of trustworthiness and expertise that customers perceive with regard to the company. This was measured on an eight-item, 7-point Likert-type scale. This notion of credibility partially overlaps with PERCRED's subdimensions of perceived sincerity and, although to a lesser extent, perceived appropriateness. Thus, a positive relationship is assumed between both constructs, which was confirmed by the regression analysis (see Table 5). *Intention to purchase* is an outcome variable of credibility perceptions, as described above. As the regression analysis shows, the more CSR reports of a firm are perceived as being credible, the more likely it is that subjects intend to purchase a product of this company (see Table 5).

Discussion and Conclusion

"It is high time to study credibility gaps systematically," stated Cornfield in 1987 (p. 468). Although this statement was made with regard to discrepancies in U.S. presidents' speeches and behaviors during the Vietnam War (Turner, 2007), it holds true for companies' communication today. The credibility gap in CSR communication (Dando & Swift, 2003) has opened up between sender and recipients of CSR communication because stakeholders perceived

that companies communicate about their responsibilities to society in a non-credible fashion. However, much ado about credibility in CSR communication has neither resulted in a consistent conceptualization of credibility nor led to a systematic examination of the topic.

In response to the call for a systematic study of credibility gaps and a novel, legitimacy-based notion of credibility, this study developed a measurement scale that analyzes recipients' perceptions of CSR reports in terms of their credibility. The recipients of CSR communication are diverse stakeholders, which is why the PERCRED scale was designed to apply to different audiences.

To provide a measurement of perceived credibility, we use a novel conceptualization of credibility consistent with the Habermasian ideal speech situation and its four validity claims, contributing to the potential of this theory for empirical inquiry (Forester, 1992). This multidimensional concept of credibility was used to operationalize perceived credibility for CSR communication that leads to moral legitimacy (Seele & Lock, 2015). We argued that perceived credibility is represented by the perceived truth, sincerity, appropriateness, and understandability of CSR reports. This was confirmed by development of the 16-item PERCRED scale. In this model, perceived credibility is the higher order construct that is described by the first-order factors of perceived truth, sincerity, appropriateness, and understandability. The scale shows high reliability and construct validity through tests of convergent, discriminant, concurrent, and nomological validity. The development process to arrive at this measurement tool comprised of five studies, including a literature review, a Delphi study, and three validation studies with independent samples, student- as well as nonstudent based.

Among the four factors describing the PERCRED scale, perceived understandability showed the lowest factor loadings. In Validation Study 4, this construct showed nonsignificant covariances with truth and minimally significant relationships with sincerity and appropriateness, indicating that understandability is not the most important construct to consider in the conceptualization of perceived credibility. However, Validation Study 4 was conducted with a student sample. Given their higher education level, students may be more literate than the general population. Accordingly, understandability was tested again in Validation Study 5, which confirmed this assumption. Although understandability in the second-order model still loads weakest on perceived credibility, this relationship is significant. Perceived truth, sincerity, and appropriateness are considered more important to the concept of perceived credibility than perceptions of understandability. However, the standing of perceived understandability in the PERCRED scale depends on the sample's level of literacy.

The PERCRED scale was evaluated not only in terms of its validity and reliability but also with regard to related measures. Trust, trustworthiness, and corporate credibility were anticipated to show positive relationships with the PERCRED measure, which was confirmed by the nomological analyses. As argued in the beginning of this article, CSR reports are heavily criticized and barely trusted by stakeholders (Illia et al., 2013). This assumption is supported by the negative correlation between the PERCRED and the distrust scale. In addition, the overall mean credibility levels of perceived credibility of CSR reports landed in the middle of the scale, indicating that CSR reports are indeed perceived as being only somewhat credible.

Implications for Theory and Practice

These results have several implications for theory and practice. The novel Habermasian conceptualization of credibility and its operationalization provides a measure for the perceptions of communicational artifacts such as CSR reports. To the best of our knowledge, this is the only attempt resulting in a validated measure of the ideal speech situation in organizational communication research to date. Future studies applying the scale can help clarify the interlinkages between the validity claims and aid in establishing a model of perceived credibility of communicational artifacts such as CSR reports.

This research adds to the growing body of CSR communication studies and theory. It provides a legitimacy-based notion of credible CSR communication in the political-normative framework (Castello et al., 2013). Thus, it strengthens and expands this research stream as an alternative to approaches such as "aspirational talk," which have remained rhetoric rather than action (Ihlen, 2015). The study demonstrated theoretically and conceptually that CSR reports can serve as facilitators in the discourse between companies and stakeholders if they are credible. The link between credibility and legitimacy is therefore further confirmed (Coombs, 1992). Future studies applying this scale can provide empirical evidence on the stakeholder specificity of CSR reporting. Moreover, experimental studies modeling credible communication and moderating reputation and trust or prior attitudes can elucidate the reasons for the lack of credibility perceptions among stakeholders.

Moreover, the PERCRED scale is well suited for research in applied contexts. Complex environments challenge companies' communication strategies and behavior, especially when it comes to norm-dependent and versatile topics such as CSR. With this tool, businesses can measure their stakeholders' perceptions. They can analyze their CSR communication efforts in terms of credibility, and thus legitimacy, and test whether, although designed to be credible, they are perceived as credible. This allows corporate players to

evaluate their communication tools more effectively and may ultimately have an impact on the CSR communication strategy more generally. The credibility gap may thus eventually be bridged.

Limitations

This scale development process has several limitations: The second stage of item refinement was accomplished with a Delphi study. Experts from academia and practice were asked to participate in the study more than three consecutive rounds. Not all contacted experts were willing to participate, which could represent a self-selection bias among the participants. However, because this study aimed at experts in the fields of CSR, CSR reporting, and Habermas, some of the invited experts declined to participate because they felt their expertise was not sufficient. Moreover, the study's data were analyzed in a quantitative albeit descriptive fashion despite the small *N*. To account for this, the participants' comments were also incorporated into the results to ensure that details and linguistic nuances were considered.

To validate the reduced item pool, the items were translated from English to German. From this stage onward, the scale was validated in German. Thus, to claim validity of the scale in different languages and to ensure that no differences in meaning occurred, the forward–back translation method was applied, following Beaton et al.'s (2000) established procedure. After Validation Study 3, another round of forward–back translation was conducted, which resulted in a validated PERCRED scale in German and English.

The samples in Studies 3 and 4 were nonrepresentative student samples representing customers of the consumer goods industry, from which the CSR report sections tested in these studies were taken. Study 5's sample stems from a large pool of online respondents, who were slightly younger and better educated than the German population. Applying the PERCRED measure in other stakeholder and cultural contexts and in samples with different demographic characteristics can further validate this instrument.

The primary objective of this research was to develop a measurement scale. However, the scale cannot provide evidence of causal relationships between measured constructs, such as intentions to purchase and perceived credibility. Future applications of this scale in experimental studies should investigate causal links between credibility perceptions and intention to purchase or trust.

Acknowledgments

Declaration of Conflicting Interests

Funding

Note

1. Material on a pilot study to this scale development conducted with existing scales can be found in the file " Pilot study scale development: literature review and results," DOI: 10.13140/RG.2.2.20798.89922.

References

Adams, J. E., Highhouse, S., & Zickar, M. J. (2010). Understanding general distrust of corporations. *Corporate Reputation Review*, *13*, 38-51.

Anderson, P. (1985). Jürgen Habermas: A philosophical-political profile. *New Left Review*, *151*, 75-105.

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, *25*, 3186-3191.

Bentele, G., & Nothaft, H. (2011). Trust and credibility as the basis of corporate social responsibility. In O. Ihlen, J. L. Bartlett, & S. May (Eds.), *The handbook of communication and corporate social responsibility* (pp. 208-230). Chichester, UK: Wiley-Blackwell.

Biedermann, C. (2008). Corporate Citizenship in der Unternehmenskommunikation [Corporate citizenship in corporate communication]. In H. Backhaus-Maul, C. Biedermann, S. Nährlich, & J. Polterauer (Eds.), *Corporate citizenship in Deutschland* [Corporate citizenship in Germany] (pp. 353-370). Wiesbaden, Germany: VS Verlag.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Publications.

Byrne, B. M. (2001). *Structural equation modeling with AMOS*. Mahwah, NJ: Lawrence Erlbaum.

Castello, I., Morsing, M., & Schultz, F. (2013). Communicative dynamics and the polyphony of corporate social responsibility in the network society. *Journal of Business Ethics*, *118*, 683-694.

Central Intelligence Agency. (2015). *The world factbook: Germany*. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/geos/gm.html

Chang, L., & Jacobson, T. (2010). Measuring participation as communicative action: A case study of citizen involvement in and assessment of a city's smoking cessation policy-making process. *Journal of Communication*, *60*, 660-679.

Chappell, Z. (2012). *Deliberative democracy: A critical introduction*. Hampshire, UK: Palgrave MacMillan.

Chen, S., & Bouvain, P. (2009). Is corporate responsibility converging? A comparison of corporate responsibility reporting in the USA, UK, Australia, and Germany. *Journal of Business Ethics*, *87*(Suppl. 1), 299-317.

Clemente, M., & Gabbioneta, C. (2017). How does the media frame corporate scandals? The case of German newspapers and the Volkswagen diesel scandal. *Journal of Management Inquiry*, 1-16. Retrieved from https://doi.org/10.1177/1056492616689304.

Coombs, W. T. (1992). The failure of the task force on food assistance: A case study of the role of legitimacy in issue management. *Journal of Public Relations Research*, *4*, 101-122.

Cornfield, M. (1987). Presidential rhetoric and the credibility gap. *Communication Research*, *14*, 462-469.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.

Dando, N., & Swift, T. (2003). Transparency and assurance minding the credibility gap. *Journal of Business Ethics*, *44*, 195-200.

Dawkins, J. (2005). Corporate responsibility: The communication challenge. *Journal of Communication Management*, *9*, 108-119.

Donaldson, T., & Dunfee, T. W. (1999). *Ties that bind*. Boston, MA: Harvard Business School Press.

Dryzek, J. S. (1995). Critical theory as a research program. In S. K. White (Ed.), *The Cambridge companion to Habermas* (pp. 97-119). Cambridge, UK: Cambridge University Press.

Du, S., & Vieira, E. T. (2012). Striving for legitimacy through corporate social responsibility: Insights from oil companies. *Journal of Business Ethics*, *110*, 413-427.

Dunbar, N. E., Jensen, M. L., Burgoon, J. K., Kelley, K. M., Harrison, K. J., Adame, B. J., & Bernard, D. R. (2015). Effects of veracity, modality, and sanctioning on credibility assessment during mediated and unmediated interviews. *Communication Research*, *42*, 649-674.

Eppler, M. J. (2006). A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing. *Information Visualization*, *5*, 202-210.

European Union. (2014). *Directive 2014/95/EU of the European Parliament and of the Council of 22 October 2014 Amending Directive 2013/34/EU as regards disclosure of non-financial and diversity information by certain large undertakings and groups*. Strasbourg. Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32014L0095

Fombrun, C. (1996). *Reputation*. Hoboken, NJ: John Wiley & Sons.

Forester, J. (1992). Critical ethnography: On fieldwork in a Habermasian way. In M. Alvesson & H. Willmott (Eds.), *Critical management studies* (pp. 46-65). London, England: SAGE.

Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research*, *18*(3), 39-50.

Habermas, J. (1984). *The theory of communicative action: Volume 1: Reason and the rationalization of society.* Boston, MA: Beacon Press.

Habermas, J. (1996). *Between facts and norms: Contributions to a discourse theory of law and democracy*. Cambridge, MA: MIT Press.

Hahn, R., & Kühnen, M. (2013). Determinants of sustainability reporting: A review of results, trends, theory, and opportunities in an expanding field of research. *Journal of Cleaner Production*, *59*, 5-21.

Hair, J. F., Tatham, R. L., Anderson, R. E., & Black, W. (2006). *Multivariate data analysis* (Vol. 6). Upper Saddle River, NJ: Pearson Prentice Hall.

Heath, R. L., Waymer, D., & Palenchar, M. J. (2013). Is the universe of democracy, rhetoric, and public relations whole cloth or three separate galaxies? *Public Relations Review*, *39*, 271-279.

Hellweg, S. A., & Andersen, P. A. (1989). An analysis of source valence instrumentation in the organizational communication literature. *Management Communication Quarterly*, *3*, 132-159.

Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, *21*, 967-988.

Hooghiemstra, R. (2000). Corporate communication and impression management—New perspectives why companies engage in corporate social reporting. *Journal of Business Ethics*, *27*, 55-68.

Hu, L. T., & 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.

Ihlen, O. (2015). "It is five minutes to midnight and all is quiet": Corporate rhetoric and sustainability. *Management Communication Quarterly*, *29*, 145-152.

Illia, L., Zyglidopoulos, S. C., Romenti, S., Rodríguez-Cánovas, B., & del Valle Brena, A. G. (2013). Communicating corporate social responsibility to a cynical public. *MIT Sloan Management Review*, *54*(3), 15-18.

Ioannou, I., & Serafeim, G. (2016). *The consequences of mandatory corporate sustainability reporting: Evidence from four countries* (Harvard Business School Research Working Paper No. 11-100, August). Boston, MS: Harvard Business School.

Jackob, N. (2008). Credibility effects. In W. Donsbach (Ed.), *International encyclopedia of communication* (pp. 1044-1047). Malden, MA: Blackwell.

Jian, G., Shi, X., & Dalisay, F. (2014). Leader–member conversational quality scale development and validation through three studies. *Management Communication Quarterly*, *28*, 375-403.

Johnson, J. (1993). Is talk really cheap? Prompting conversation between critical theory and rational choice. *American Political Science Review*, *87*, 74-86.

Keller, K. L., & Aaker, D. A. (1992). The effects of sequential introduction of brand extensions. *Journal of Marketing Research*, *29*, 35-50.

Kim, Y. (2001). Searching for the organization-public relationship: A valid and reliable instrument. *Journalism & Mass Communication Quarterly*, *78*, 799-815.

Lafferty, B. A. (2007). The relevance of fit in a cause–brand alliance when consumers evaluate corporate credibility. *Journal of Business Research*, *60*, 447-453.

Lafferty, B. A., & Goldsmith, R. E. (1999). Corporate credibility's role in consumers' attitudes and purchase intentions when a high versus a low credibility endorser is used in the ad. *Journal of Business Research*, *44*, 109-116.

Leiner, D. J. (2017). Our research's breadth lives on convenience samples. A case study of the online respondent pool "SoSci Panel." *Studies in Communication Media*, *5*, 367-396.

Linstone, H. A., & Turoff, M. (Eds.). (1975). *The Delphi method: Techniques and applications*. London, England: Addison-Wesley.

Lock, I., & Seele, P. (2016). The credibility of CSR (corporate social responsibility) reports in Europe. Evidence from a quantitative content analysis in 11 countries. *Journal of Cleaner Production*, *122*, 186-200.

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, *36*(4), 611-637.

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, *35*, 293-334.

Maneesriwongul, W., & Dixon, J. K. (2004). Instrument translation process: A methods review. *Journal of Advanced Nursing*, *48*, 175-186.

McCroskey, J. C., & Teven, J. J. (1999). Goodwill: A reexamination of the construct and its measurement. *Communications Monographs*, *66*, 90-103.

Melican, D. B., & Dixon, T. L. (2008). News on the net credibility, selective exposure, and racial prejudice. *Communication Research*, *35*, 151-168.

Milne, M. J., & Gray, R. (2013). W(h)ither ecology? The triple bottom line, the global reporting initiative, and corporate sustainability reporting. *Journal of Business Ethics*, *118*, 13-29.

Newell, S. J., & Goldsmith, R. E. (2001). The development of a scale to measure perceived corporate credibility. *Journal of Business Research*, *52*, 235-247.

Ohanian, R. (1991). The impact of celebrity spokespersons' perceived image on consumers' intention to purchase. *Journal of Advertising Research*, *31*, 46-54.

Organisation for Economic Co-operation and Development. (2015). *OECD better life index*. Retrieved from http://www.oecdbetterlifeindex.org/countries/germany/

Oxford English Dictionary. (2016). *Credibility*. Retrieved from http://www.oed.com/

Perera Aldama, L. R., Awad Amar, P., & Winicki Trostianki, D. (2009). Embedding corporate responsibility through effective organizational structures. *Corporate Governance: The International Journal of Business in Society*, *9*(4), 506-516.

Rafaeli, A., & Vilnai-Yavetz, I. (2005). Managing organizational artifacts to avoid artifact myopia. In A. Rafaeli & M. G. Pratt (Eds.), *Artifacts and organizations* (pp. 9-22). Mahwah, NJ: Lawrence Erlbaum.

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, *21*, 173-184.

Scherer, A. G., & Palazzo, G. (2011). The new political role of business in a globalized world: A review of a new perspective on CSR and its implications for the firm, governance, and democracy. *Journal of Management Studies*, *48*, 899-931.

Seele, P. (2016). Digitally unified reporting: How XBRL-based real-time transparency helps in combining integrated sustainability reporting and performance control. *Journal of Cleaner Production*, *136*, 65-77.

Seele, P., & Lock, I. (2015). Instrumental and/or deliberative? A typology of CSR communication tools. *Journal of Business Ethics*, *131*, 401-414.

Sethi, S. P. (1975). Dimensions of corporate social performance: An analytical framework. *California Management Review*, *17*(3), 58-64.

Sias, P. M., & Wyers, T. D. (2001). Employee uncertainty and information-seeking in newly formed expansion organizations. *Management Communication Quarterly*, *14*, 549-573.

Stamm, K., & Dube, R. (1994). The relationship of attitudinal components to trust in media. *Communication Research*, *21*, 105-123.

Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, *20*, 571-610.

Turner, K. J. (2007). Credibility gap. In S. L. Vaughn (Ed.), *Encyclopedia of American journalism* (pp. 123-124). New York, NY: Routledge.

Williams, P., & Drolet, A. (2005). Age-related differences in response to emotional advertisements. *Journal of Consumer Research*, *32*, 343-354.

Yin, Y. (1990). Cognitive and affective priming effects of the context for print advertisements. *Journal of Advertising*, *19*(2), 40-48.

Author Biographies

Irina Lock is assistant professor of corporate communication at the University of Amsterdam, the Netherlands. Her research interests cover corporate social responsibility (CSR) communication, digital corporate communication, and public affairs.

Peter Seele is associate professor of CSR and business ethics at the Università della Svizzera italiana, Lugano, Switzerland. His research focuses on CSR reporting, CSR and digitalization, and ethics and the law.