Message Delivery: A Revised Approach for Public Relations Measurement

David Michaelson
Teneo Strategy

Don W. Stacks
University of Miami

Jennifer Clark
Teneo Strategy

Abstract

The measurement of communication impact has become a long-standing discussion in the public relations research community. While discussions about standardized measurement date back decades (Lindenmann, 2005, pp. 3-4), the current debates about standardized measurement became prominent with the work of Michaelson and Stacks (Michaelson & Stacks, 2011). In their 2011 article, they argued that public relations measurement needed to be standardized to demonstrate communication activity impact on absolute levels of impact as well as on relative measures. The question this paper explored is "What is the root cause for the lack of adoption of standardized measurement by the public relations profession?" In so doing, this paper proposes a revised approach to standardized measurement that is based on message delivery based on a revision of an analytic approach originally proposed by Michaelson and Griffin (2005) that is steeped in message level content analysis. It expands this approach by creating a unified message delivery-scoring model that provides for the inclusion of intended messages, the omission of intended messages, and the presence of negative messages in media. The research includes rigorous testing that demonstrates the validity and statistical reliability of the proposed model and its relationship to sentiment analysis. The analysis also examines the statistical impact of negative messages in media and concludes that the presence of negative messages effectively detracts from the overall sentiment of an article just as the presence of positive messages increases overall sentiment.

Keywords: content analysis, measurement, messages, sentiment, standardization

Introduction

The measurement of communication impact has become a long-standing discussion in the public relations research community. While discussions about standardized measurement date back decades (Lindenmann, 2005, pp. 3-4), the current debates about standardized measurement became prominent with the work of Michaelson and Stacks (Michaelson & Stacks, 2011). In their 2011 article, they argued that public relations measurement needed to be standardized to

demonstrate communication activity impact on absolute levels of impact as well as on relative measures. The argument they presented was that the relative measures of impact were particularly critical as they placed communication measurement in context by providing realistic objectives and goals when compared to similar programs (p. 4)[1].

Since the publication of that article, there has been considerable discussion in the measurement community about creating generally accepted industry standards. This discussion has included the creation, in 2012, of the Coalition for Public Relations Research Standards to actively develop and test measurement standards for use by the public relations profession.[2] (Thorson, Michaelson, & et.al., 2015, p. 4) While it is currently inactive, prior to its current state, the Coalition published numerous proposed standards that appear on the Institute for Public Relations website[3]. Apart from proposed measures based on audience impact[4], virtually all the other proposed standards concentrated almost exclusively on *media* analysis, both traditional and social.

Despite the development and testing of these proposed standards, there has been scant progress in the adoption of any standard measures by the public relations community. While several large corporations[5] announced adoption of the Coalition standards, an analysis conducted in 2015 by Thorson, et. al. showed that while a minority of professionals claim to having adopted standards (25%), many of these professionals were not referring to the published standards, but were likely referring to other forms of measurement. (pp. 16-19)

The question that needs to be explored is "What is the root cause for the lack of adoption of standardized measurement by the public relations profession?" Over the years, researchers have attempted to understand why measurement has not been broadly adopted by the profession. These analyses have included the Generally Accepted Practices (GAP) studies conducted by the Annenberg School of the University of Southern California[6] as well numerous other studies conducted by other organizations.[7] These studies identified several root causes including the overall cost of and time required for measurement (Macnamara, 2011, p. 5) and a series of other concerns ranging from uncertainty about how to measure to a lack of demand for or interest in measurement (pp. 5-7). A possible explanation for the lack of demand or interest in public relations measurement is a perceived inability among public relations professionals to identify clear benefits to the practice of measuring communication activities. This concern was raised by Michaelson and Griffin in 2005 (Michaelson & Griffin, 2005, p. 4) and continues to remain true today.

The associated *cost of measurement* will continue as a concern within the public relations profession and in turn limit the adoption of these procedures. While the concern with cost is

---

[1] "Program" and "Campaign" are both used in the communication literature. In this discussion, we prefer the term, "program" as it signifies a strategic use of message theory, method, measurement, assessment, evaluation, and interpretation.

[2] The Coalition included The Council of Public Relations Firms, The Public Relations Society of America (PRSA), the Global Alliance for Public Relations and Communication Management, the Institute for Public Relations (IPR) and AMEC.

[3] A complete list of these published standards appear on the following website: http://www.instituteforpr.org/category/research/research-methods-standards/

[4] Audience impact measures are discussed in detail in Michaelson & Stacks, 2011

[5] McDonald's USA, Southwest Airlines, General Motors, and General Electric.

[6] http://ascjweb.org/gapstudy/

[7] For a detailed discussion on these prior studies refer to Jim Macnamara's paper: PR Metrics – Research for Planning and Evaluation of PR and Corporate Communications (Macnamara, 2011)

easily understood, the more troubling concern is the perception of uncertainty about the utility to be gained from actually measuring public relations. Other than a generalized belief that public relations needs to demonstrate its value to the "C-Suite" (Lofgren, 2014),[8] there is no clear statement or position by the profession about why measurement is an essential part of any communication program.[9] Without a clearly stated benefit, the core challenge for the profession is finding a clear expression of the utility that should be gained from measuring public relations activities. Is measurement a justification to secure funding and be self-congratulatory or is it a tool to improve communication performance and assure that programs are successful in contributing to meeting business objectives?

What Should Measurement Do?

In its broadest sense, public relations measurement needs to determine how communication activities directed at a target audience is advancing business objectives. The objective however is not to demonstrate the "value" of public relations to the "C-Suite" or to justify budgets. There is already ample evidence[10] – scientific as well as anecdotal – that demonstrates public relations can have an impact if it is properly executed (Michaelson & Stacks, 2009; Newlands, 2014;Stacks, 2017).

Rather than focus on "self-justification," the more effective objective for a measurement program should instead be a focus on *the ability of a communication program to deliver the foundation that allows business objectives to be achieved*. In essence, public relations is a mechanism that influences behaviors that in turn create business impact. Those behaviors are affected by specific measurable areas that are the responsibility of the communication function.

In 2011, Michaelson and Stacks identified five specific measures that should be at the core of communication programs to have behavioral impact. Typically, these measures cross five areas of communication effect, each of which is central to achieving program and business success. The core measures – referred to as the B.A.S.I.C. model – examines the lifecycle of communication programs that are essential to meeting objectives. These measures are *awareness* of a company, brand, product or issue, *accurate knowledge* about a company, brand, product or issue*, relevance of messages* about a company, brand, product or issue to the target audience for a communication program, *intent of the target audience to take a desired action* based communication about a company, brand, product or issue and willingness of the target audience to *advocate* for a company, brand, product or issue based on their experience (Michaelson & Stacks, 2017, pp. 26-27).

The one element that runs throughout each of these five communication areas is the need to deliver intended messages to the targeted audience for the product, service, brand or idea. While each of the measurable areas discussed above is critical in assuring that communication positively impacts business outcomes, it diminishes when intended messages fail to reach their audience or are counteracted by negative commentary.

Linking messages to communication objectives is the foundational requirement to assure alignment between media relations initiatives and a communication program (Stacks, 2002,

---

[8] This is an example of numerous postings and discussions on this topic that appears using the search term "demonstrate value to c-suite by public relations"

[9] Academically, both Broom and Dozier (1990) and Stacks (2002) (2011) (2017) have advanced this argument. Specifically, see Stacks (2017) pp. 17-96.

[10] There is a wide-ranging number of posts on this topic yielding 69 million results using the search term "does public relations work"

2011, 2017). These messages must convey to a target audience specific and accurate information that support business objectives and in turn achieves behavioral impact (Michaelson & Griffin, 2005, p. 7). At the root of these measures are the messages that are the core of any communication program. Delivering accurate messages is the most fundamental element for communicators across each of these five measures if they are to be successful in impacting behavior among target audiences and in turn meeting business objectives. "Message accuracy," according to Michaelson and Griffin, is determined through "an analysis of four basic elements: the inclusion of basic facts, the inclusion of misstatements about these basic facts, the inclusion of incomplete, deceptive or misleading information that biases the interpretation about basic facts and the omission of basic facts." (p. 6)

Interestingly, work on public relations program excellence (Michaelson, Wright, & Stacks, 2012) proposed a model of excellence based in research and measurement. The model, at its most elemental level, argues that campaigns must set objectives and goals, link them to client/business objectives and goals, objectively measure campaign progress, analyze, interpret, and then put the findings to use in what Michaelson and Stacks (2017, p. 12) refer to as "end-to-end" measurement. Further, objective measures yield necessary information about the campaign that then can be evaluated as to whether the campaign meets intermediate and/or advanced levels of excellence. Not stated specifically, but such basic data can then be correlated with other business functions (Stacks, 2017) to provide quantitative indicators of campaign success and contribution to overall business objectives and goals.

While it is conventional wisdom that accurate delivery of messages to a target audience is an essential element of any communication program (Wetherhead, 2011), there is virtually no agreement on how to measure message accuracy or even include this measure in the recommendations for standard measures of communication performance as part of an overall measurement program that includes formative research and measures of program impact on target audiences based on exposure to intended measures. While Michaelson and Stacks included the concept of measuring message accuracy as a standard metric in their 2011 paper (pp. 18-19), they did not present specific methods for calculating or reporting on this measure.

Why Does This Challenge Exist?

An examination of the recommendations proposed by the Coalition for Public Relations Research Standards (Institute for Public Relations, n.d.) presents a common theme. Except for the outcomes[11] standards for communication impact proposed by Michaelson and Stacks (2011), the recommended standards developed by the Coalition focused exclusively on media analysis – traditional and social. The specific output[12] measures presented as possible standards include: content sourcing, transparency, mentions, validity and reliability, social media standards, items for media analysis, circulation, reach, impressions and engagement.

A further review of these recommended measures reveals another commonality that is at the root of the challenges surrounding the lack of adoption of measurement standards for public

---

[11] "Quantifiable changes in awareness, knowledge, attitude, opinion, and behavior levels that occur as a result of a public relations program or campaign." (Stacks & Bowen, Dictionary of Public Relations Measurement and Research, 2013, p. 21)

[12] "What is generated as a result of a PR program or campaign that may be received and processed by members of a target audience, and may have cognitive impact on outtakes" (Stacks & Bowen, Dictionary of Public Relations Measurement and Research, 2013, p. 21)

relations activities. While each of the proposed output standards includes an essential element that needs to be included in any detailed media analysis, overall each of these proposed standards fail to take into consideration the fundamental reasons for measuring in the first place – assessing progress toward meeting communication objectives, diagnostics to determine why meeting objectives may be impeded and prescriptive insights to redirect programs that will increase the likelihood of success, or what Michaelson, Wright, and Stacks (2012) argue are the most basic elements of communication program excellence.

An example of the elements contained in these proposed output assessments that do not take these measurement objectives into consideration is the "social media measurement standards" (SMM). These proposed SMM standards include six specific measurement areas: (1) content and sourcing; (2) reach and impressions; (3) engagement and conversation; (4) influence; (5) opinion and advocacy; and (6) impact and value. (Institute for Public Relations, 2013) Each of these measures provides useful information that determine the credibility of sources, audience reach and quality of audience that can impact communication impact. Unfortunately, they are not integrated into an overall model of communication where the primary objective is to deliver intended messages to a target audience that are filtered through third party endorsers such as traditional and social media. The proposed standards for traditional media are like the social media measurement standards in their limited focus on similar metrics. (Institute for Public Relations, n.d.) These traditional media analysis measures are limited to discussions of what is a valid item for media analysis, such as an article or a broadcast segment; what is considered a mention within an article or other form of media as well as a discussion of measures such as circulation, reach, and impressions.

While each of these measures contributes to the overall measurement of public relations activity, the usability of these assessments to evaluate the performance of communication programs and in turn provide direction to improve performance through standardized interpretation that allows for comparison is extremely limited. It is commonly accepted which articles are relevant for inclusion in a content analysis and for the most part, a mention, particularly if it includes the name of a client or product is self-evident. By comparison, circulation, reach and impressions can provide an understanding if the articles reached a significant audience. However merely reaching an audience does not determine if content was observed and recalled or the impact and effectiveness on the audiences themselves.

Discussions of challenges in the usability of measurement metrics are not new. Instead, these challenges reflect a lack of understanding in the public relations industry of how research and measurement should be employed. In 2005, Michaelson and Griffin published a critical review of eight content analysis methods that were commonly used in public relations measurement at that time.[13] The central tenet of that article was that each of the analytic approaches, taken individually or in combination, was flawed since each failed to adequately address the core objective of a public relations program – delivering intended messages to targeted audiences through earned and social media. In the intervening 12 years, there appears to have been little or no progress by the profession in advancing the utility of content analysis metrics.

In their 2005 analysis, Michaelson and Griffin (pp. 6-7) identified two specific flaws that

---

[13] These eight methods include: clip counting, circulation and readership analysis, advertising value equivalency or AVE, simple content analysis, message analysis, tonality analysis, prominence analysis, quality of coverage and competitive analysis.

were common across each of the methods of content analysis in use at that time: (1) not determining message accuracy and (2) not linking messages to communication objectives. As the analysis of the media analysis standards proposed by the Coalition for Public Relations Research Standards shows, metrics for message delivery have still not entered the mainstream measurement discussion even though message delivery is fundamental in the success of a communication program.

As noted in the Michaelson and Griffin article, message accuracy is determined by an analysis of four basic elements in an article or post. These elements are "the inclusion of basic facts, the inclusion of misstatements about these facts, inclusions of incomplete, deceptive or misleading information that biases the interpretation about basic facts and the omission of basic facts." (p. 6). In that article, they provided very precise definitions of each of these elements that comprise message accuracy.

> **Basic facts** are the fundamental information that is central to any communications program... **Misstatements** are generally understood as errors or incorrect information included in an article or publication… **Incomplete information** is a statement, opinion or point-of-view that selectively includes some information, but excludes other relevant facts… **Omissions** are the absence of key information that should be included in a specific article or publication. (pp. 6-7) *(emphasis added)*

While Michaelson and Griffin has been cited numerous times in the scholarly literature,[14] there has been scant adoption of this analytic approach for content analysis by communication professionals. Michaelson and Griffin themselves speculated that adoption of this analytic method would likely be limited due to a need for in-depth knowledge by researchers on the issues surrounding the coding, advanced research skill levels required to code each article accurately and significantly increased costs.

While each of these concerns are substantive challenges, an additional factor needs to be taken into consideration that could further limit wider adoption of these metrics. That challenge is the inherent complexity of the reporting associated with this type of analysis. The approach described by Michaelson and Griffin makes it challenging to interpret findings and understand their impact on communication programs. This complexity may be a contributing factor, making it difficult for communication professionals to compare findings with other programs to understand relative performance – one of the core requirements of a standardized measurement system. Thus, the utility of the approach as a standardized metric for media analysis is severely limited.

A Proposed Revision

Based on the discussion above, it is clear something needs to be done that addresses public relations measurement problems on the professional side of the equation. This is not to denigrate the professional who must work quickly and within time budgets that their academic counterparts rarely face, but to provide an objective and applied measurement strategy to media

---

[14] See:
https://scholar.google.com/citations?view_op=view_citation&hl=en&user=s3_tpQQAAAAJ&citation_for_view=s3_tpQQAAAAJ:u-x6o8ySG0sC

analysis with the goal of moving the profession toward a standardized media analysis metric.

As discussed, a complete communication story contains all relevant messages and does not contain negative, erroneous or misleading messages. This does not mean that all distributed messages need to appear in an article to be considered complete. Rather, only those messages that support the topic of the article need to be included. With the objective of creating a single score that reflects these findings, two scoring systems focusing on message delivery were created. Once score focuses only on the presence or absence of relevant positive/intended messages. The second scoring system takes the presence of negative, erroneous or misleading messages into consideration in addition to the presence or absence of relevant intended messages.

Positive Message Scoring Calculation

The proposed calculation when considering positive or intended messages exclusively takes two factors into consideration: (1) the presence of specific intended messages appearing in an article or story and (2) the total number of messages considered relevant to an individual article or story.

The calculation for a score that focuses only on *positive* messages (Message Delivery Score A) is as follows:

$$\frac{p * 100}{p + o} = x$$

Equation 1: Message Delivery Score A

Using Message Delivery Score A, p is the number of relevant intended messages that appears in an article or story and o is the number of relevant messages that were omitted or did not appear in that same article. An example of this calculation is an article where there are eight relevant messages for an article and three of these messages appear in an article.

$$\frac{3 * 100}{3 + 5} = 37.5$$

Message Delivery Score A is 37.5 out of a possible score of 100.

The limitation of Message Delivery Score A is that it assumes all messages will have an equal impact on a target audience. While this limitation could be addressed through primary research using regression statistics that measure the contribution of individual messages (independent variables) on a desired behavioral outcome (dependent variable), this type of research is impractical in day-to-day practice. A possible consideration is to identify intended messages as primary, secondary, or tertiary and limiting the analysis only to primary messages and therefore focusing the analysis only on the primary communication objectives. This would minimize the need for weighting message contribution and provide a more robust, albeit basic metric that can be easily used as a relative measure in assessing communication performance.

Scoring Calculation Including Negative or Erroneous Messages

While Message Delivery Score A only considered positive or intended messages, there are other elements that can impact the delivery of messages and therefore the potential impact of a story or article on a target audience. The presence of negative, misleading or erroneous messages can effectively counter the impact of any positive mentions and therefore diminish the

potential impact of an article or story on a target audience.

As with positive or intended messages, the challenge remains in determining the contribution of negative messages to an overall message delivery score. Two basic approaches toward addressing this challenge are raised for consideration.

One approach is to assume that message delivery is assessed as the proportion of relevant positive or intended message delivery in the context of all relevant included messages. That approach (Message Delivery Score B) is calculated using the following formula:

$$\frac{p * 100}{p + o + n} = x$$

Equation 2: Message Delivery Score B

Using Message Delivery Score B, p is the number of relevant intended messages that appears in each article or story and o is the number of relevant messages that did not appear or were omitted in that same article and n are the negative messages included in an article. An example of this calculation is an article where there are eight relevant messages for an article and four of these messages appear in an article. In addition to the positive or intended messages, three negative messages also appear.

$$\frac{4 * 100}{4 + 4 + 3} = 36.4$$

Message Delivery Score B is 36.4 out of a possible score of 100.

An alternative approach to the inclusion of negative or erroneous messages in the calculation assumes that negative messages counter the impact of positive messages in each article. In this case, message delivery is assessed as the proportion of positive or intended message delivery in the context of all intended messages minus the presence of negative or erroneous messages. That approach (Message Delivery Score C) is calculated using the following formula:

$$\frac{(p - n) * 100}{p + o} = x$$

Equation 3: Message Delivery Score C

Using Message Delivery Score C, p remains the number of relevant intended messages that appears in an article or story and o is again the number of relevant messages that did not appear or were omitted in that same article. n are the negative messages included in an article. An example of this calculation is again an article where there are eight relevant messages for an article and four of these messages appear in that article. In addition to the positive or intended messages, three negative messages also appear.

$$\frac{(4 - 3) * 100}{4 + 4} = 12.5$$

Message Delivery Score C is 12.5 out of a possible score of 100.

In each of these three options, message delivery is measured using a calculated index with a maximum value of 100, where 100 is considered a complete story. A complete story is defined as having all relevant intended messages present and negative or erroneous messages do not appear in the article or story. Thus, for purposes of this analysis, the score for where only

positive messages are included in the calculation is 37.5. By comparison, the overall delivery score that includes relevant intended messages as well as negative or erroneous messages in its assessment is calculated as either 36.4 or 12.5 depending on how negative or erroneous messages contribute to the overall calculation. This raises the open question on which of the scores is most reflective of actual message delivery and possible message impact.

Assessing Validity and Reliability for Each Proposed Calculation
On an initial review, each of these calculations appears to have surface validity in assessing message delivery through an intermediary. Each measures the presence of intended messages in light of all overall relevant messages with the presence of negative or erroneous messages also taken into consideration for two of the proposed calculations. A more complete analysis that assesses the validity of each of these alternatives as well as assesses the reliability of each calculation is needed to determine which of these approaches will function best as a standardized measure of message delivery.

To assess the validity[15] and reliability[16] of each of these alternative calculations, an analysis of 95 news articles on an industrial company that appeared in major business media outlets in the United States during 2013 was conducted. Intended, negative, or erroneous messages were identified through a review of outbound communication for the company including press releases and other sources. A total of 17 intended messages and 13 negative or erroneous messages were included in the analysis.

Each article was coded for the presence of intended messages, the omission of relevant intended messages, and the presence of negative or erroneous messages. Articles were also coded for topic, publication, presence of company spokesperson and sentiment. Sentiment was coded on a five-point Likert-type scale from the perspective of a neutral reader. The scale ranged from "very positive" (5) to "very negative" (1) with a "neutral" mid-point (3). Coding was conducted by a single experienced researcher[17] who was familiar with the company and the messages (intended and negative).

The distribution of relevant included messages ranged from 0 to 4 with a mean of 1.76 intended messages per article. The distribution of relevant omitted messages ranged from 0 to 7 with a mean of 1.83 omitted relevant messages per article. The distribution of negative or erroneous messages ranged from 0 to 4 with a mean of 1.05 negative or erroneous messages per article.

Validity Assessment
To evaluate validity, message delivery scores were determined for each article included in the analysis. Separate scores were determined for each the three proposed calculations. When only positive messages were included in the calculations (Message Delivery Score A), the mean message delivery score was 49.3 with a median of 50. When negative messages were included in the base number of messages (Message Delivery Score B), the mean message delivery score as

---

[15] "Validity—m. The extent to which a research project actually measures what it is intended, or purports to measure; see also: measurement validity" (Stacks & Bowen, 2013, p. 33)

[16] "Reliability—m. The extent to which results would be consistent, or replicable, if the research were conducted a number of times." (Stacks & Bowen, 2013, p. 26)

[17] By using a single coder, measures of intercoder reliability were unnecessary in this analysis.

39.7 with a median of 36.4. When negative messages were calculated that have a counteractive effect on overall message delivery (Message Delivery Score C), the mean message delivery score was 11.6 with a median of 20.
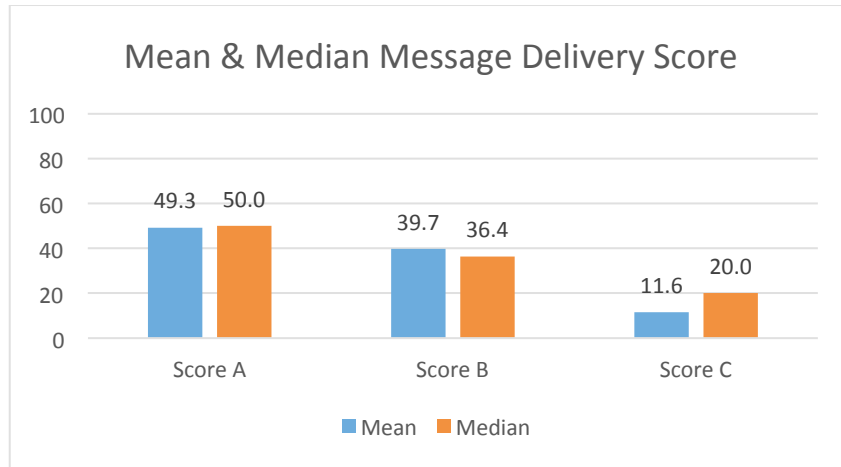

Figure 1: Mean & Median Message Delivery Scores

The distribution for Message Delivery Scores A and B ranged from zero to 100. The distribution for Message Delivery Score C ranged from negative (-) 200 to 100, reflecting a potential counteractive impact of negative messages on overall message delivery.
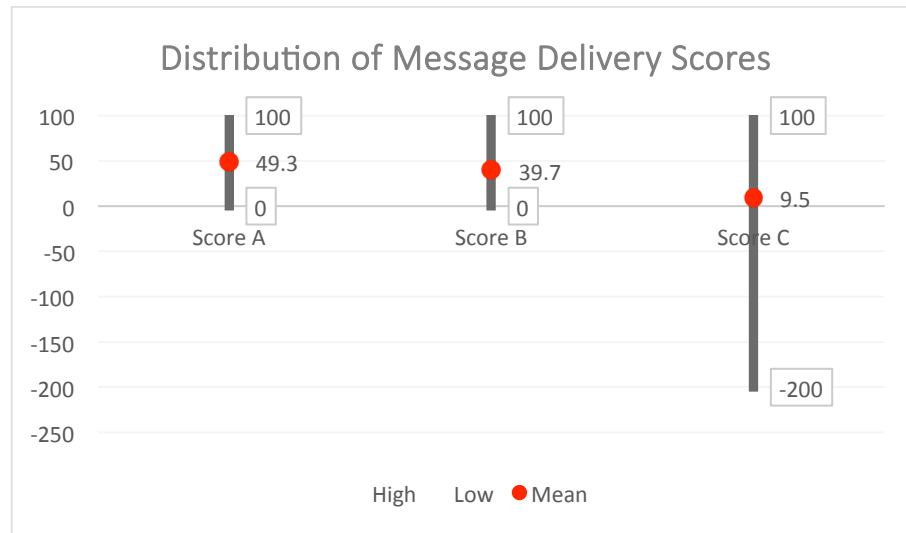

Figure 2: Distribution of Message Delivery Scores

Based on this assessment, Message Delivery Scores A and C appear to have higher validity (measuring what is intended) than Message Delivery Score B as indicators of overall message delivery. When only the presence or omission of relevant intended messages (Message Delivery Score A) is considered, it is reasonable to assume a valid score will fall between zero and 100. However, if it is reasonable to presume that the presence of negative messages in an article have the potential to counteract the impact of positive message messages (Coulter &

Sewall , 1995), then Message Delivery Score B has a lower level of validity since it does not measure the overall potential impact of messages, positive and negative, on a reader. By contrast, Message Delivery Score C demonstrates the potential impact of negative as well as positive impact on overall message delivery.

Reliability Assessment

One of the insights from the work of Michaelson and Griffin (2005) was that a different approach to content analytic units of analysis needed to be explored.

> …most content analysis tends to concentrate on tonality of an article rather than the fundamental correctness of the reporting. This concentration on tonality, consequently, fails to provide the information that is necessary to implement a corrective strategy with media, which can in turn result in an increase in accurate and appropriately positioned messages. (p. 11)

They went on to comment that in an analysis of articles conducted on behalf of MetLife, between 60% and 85% of articles on annuities and long-term care included errors or omissions (p. 11). A change in their media relations strategy based on these errors and omissions resulted in "…a significant decline in the proportion of articles with either errors or omissions as well as an overall increase in the number of articles by 45 percent" (p. 12). They went on to also point out that "While tonality was not a part of this analysis, the overall effect was that reporting on the issues was much more favorable and more in line with MetLife's media relations goals." (p. 12) Although the results from MetLife were anecdotal, the implication is that as intended message delivery increases, and negative message delivery is minimized, sentiment will also increase.

To assess this hypothesis and the reliability of message delivery scoring using the proposed calculations, an examination of the relationship between message delivery and sentiment was conducted on the same 95 articles used for the validity assessment. The following charts show the comparative relationships of message delivery scores with sentiment as well as the linear relationships between message delivery scores and sentiment for each calculation under consideration.
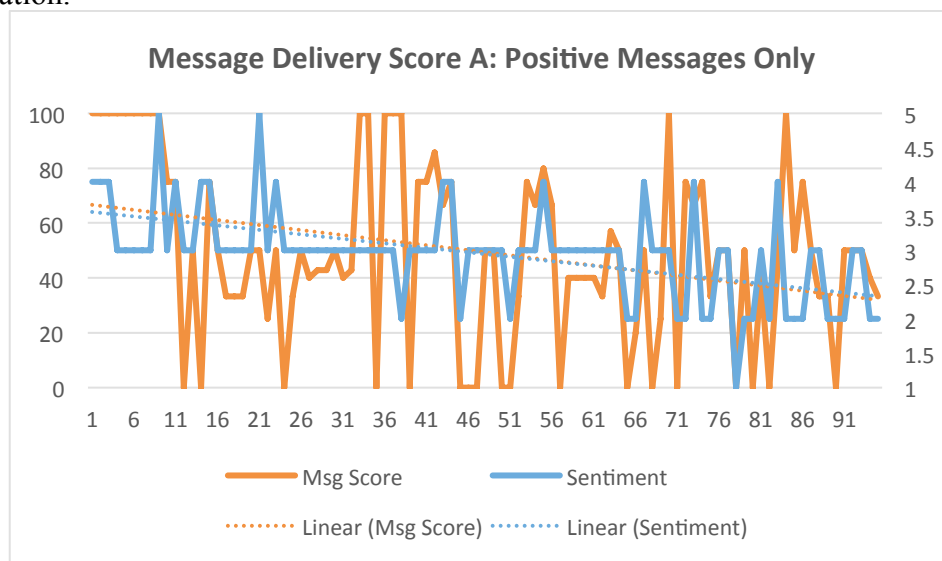

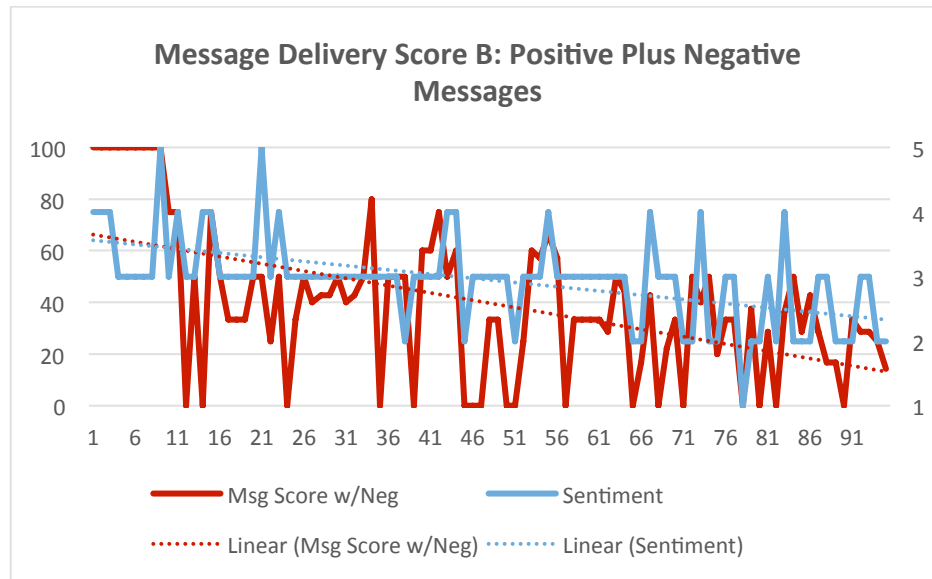*Figure 3: Message Delivery Score A Compared to Overall Sentiment*

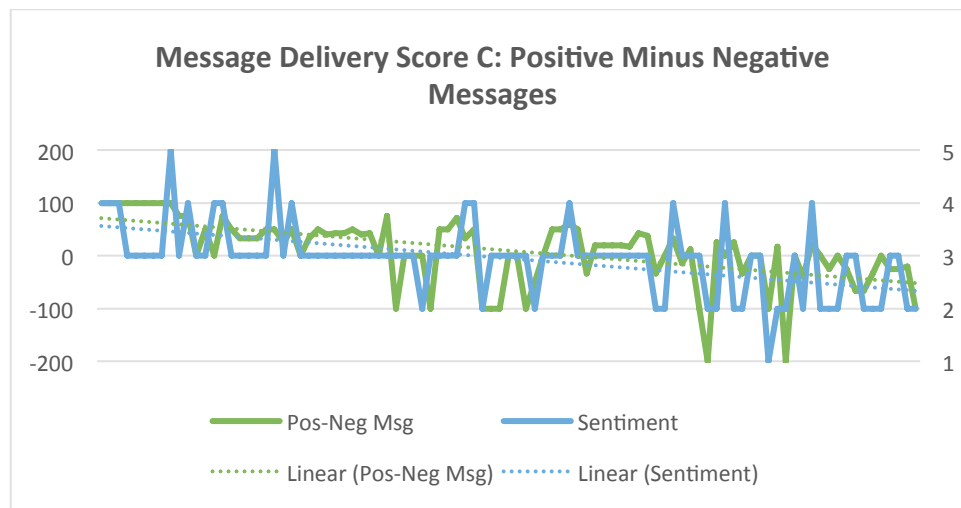*Figure 4: Message Delivery Score B Compared to Overall Sentiment*

Figure 5: Message Delivery Score C Compared to Overall Sentiment

While the findings from this analysis are based only a single case study, in this instance there is a clear linear relationship between message delivery and sentiment for Message Delivery Scores A and C. Low levels of message completeness appear to correlate with lower levels of sentiment. Conversely, high levels of message delivery show a linear relationship with higher levels of overall sentiment. This indicates there is likely a high level of reliability for measures A and C with the relationship between Message Delivery Score C and overall sentiment being particularly close. This further reinforces the validity of each of these scores (A and C). By contrast, the relationship between Message Delivery Score B and sentiment is only directional and appears much less reliable in its overall measures.

Further Statistical Assessment

While the linear analysis comparing message sentiment with message delivery scores is illuminating about the potential value of this proposed metric, further analysis was needed to evaluate the reliability of each of these three measures. A linear regression comparing the message completeness scores with sentiment was conducted for each of the 95 articles included in the analysis for this purpose. Separate analyses were conducted for each of the three proposed message delivery scores.

Using a Multiple R-squared analysis[18] where the dependent or *predicted* variable is overall sentiment of the articles and the independent or *predictor* variable is the Message Delivery score, a positive correlation[19] between Message Delivery and overall article sentiment exists for all three Message Delivery scores ($p < .05$).

	Message Score A	Message Score B	Message Score C
Multiple R-Squared	0.33	0.46	0.54

Table 1: Multiple R-Squared for Message Delivery Scores (scores without superscripts are significantly different, $p < .05$)

However, correlation levels[20] vary considerably among each of the three scores. When only positive or intended messages are included in the score, correlation levels are at best medium. When negative messages are included in the score, correlation levels significantly increase, with Message Delivery Score C achieving a high level of correlation.

An analysis of t-statistics generated for each Message Delivery Score yielded similar results. While the t-statistic for each Message Delivery score is significant ($p < .05$), the level increases significantly with the *inclusion of negative messages* and even further *when negative messages are calculated as having a counter effect* to the positive or intended messages included in an article (Message Delivery Score C).

	Message Score A	Message Score B	Message Score C
t-Statistic	3.34	5.02	6.20

Table 2: t-Statistic for Message Delivery Scores (scores without superscripts are significantly different, $p < .05$).

While the linear relationship between the Message Delivery score and sentiment is lower for score B than for scores A and C, the introduction of negative messages in the score

---

[18] Multiple R-squared is used for evaluating how well a model fits the data. The statistic presents the variance in the dependent variable (the predicted variable) that can be explained by the independent variables (the predictor variables).

[19] Correlation results are between -1 and 1. A result of -1 means that there is a perfect negative correlation between the two values, while a result of 1 means that there is a perfect positive correlation between the two variables. A result of 0 means that there is no linear relationship between the two variables. A correlation of 0, -1 or 1 is rare, with likely results somewhere in between.

[20] *Very high correlation* 0.7 to 1.0 or -.07 to -1.0
High correlation: 0.5 to 0.7 or -0.5 to -0.7
Medium correlation: 0.3 to 0.5 or -0.3 to -0.5
Low correlation: 0.1 to 0.3 or -0.1 to -0.3

significantly increases the Multiple R-squared and the t-statistic over Message Delivery Score A and these metrics continue to increase in significance with Message Delivery Score C. The most basic conclusion is that *the presence of negative messages in a story effectively detracts from overall sentiment in an article just as the presence of positive messages increases overall sentiment*.

Interpretation of Message Delivery Scores
        Each article analyzed using any of these calculations will have a message delivery score that reflects the presence of intended messages, the omission or absence of these intended messages and depending on the calculation selected, the presence of negative or erroneous messages. *This will provide a standardized score (either Z or T) for comparing the delivery of messages in individual articles.* However, this approach is also flexible in that message delivery scores from a campaign, a specific time or other unit of measure can be averaged to assess overall performance for a specific analytic period.
        While message delivery scoring shows significant promise in evaluating the impact of media relations activities, the challenge is determining how to interpret these scores due to a lack of normative data. While it is obvious that negative scores generated using Message Delivery Score C indicate problematic or ineffective media relations, it is less clear on how to interpret a positive score.

Application of Message Delivery Scores
        As noted earlier (page 9), message delivery is at the core of measurement in determining the success of *any* communication program and needs to be considered at each measureable stage of the communication process. The objectives of this approach to measurement are to (1) identify those messages which are most likely to impact a desired behavioral change, (2) determine if these messages are delivered to intermediaries and intended audiences for the messages, (3) evaluate target audience reaction to these messages, and (4) assess progress toward establishing the conditions that are necessary in order effect behavior.
        Without the ability to determine if the delivery of intended messages to a target audience occurs, an organization will not be able to effectively determine if their communication activities contribute to the business goals the communication program is intended to support. For this determination to happen, an effective measurement program needs to examine the *entire* communication process starting from *origination* of the message through *delivery* and *impact* against its target audience (i.e., "end-to-end" research and measurement). It therefore becomes necessary to understand not only *if* the message appears, but also at the key points *where* the measurement of communication performance assures intended messages are reaching their target audience, and, in turn, these messages are facilitating an intended impact that results in behavioral change.
        There are five distinct stages where the communication of intended messages needs to be measured. These stages are during the developing of the messages where these messages are also tested for efficacy, if the messages are distributed by the sender, the delivery of the messages to the target audience through media, organic reaction to the messages by the target audience as a result of exposure to the message through media and the impact of the messages on the target audience.

Figure 6: Stages of Message Measurement

The message delivery score presented in this paper fulfills one of these key measures. However, by itself is not sufficient to function as a comprehensive assessment of communication performance. A full measurement program must also include tangible impact measures that include the core assessments of awareness, accurate level of knowledge, relevance of messages, intent to take action and willingness to be an advocate.[21]

Conclusions and Next Steps

To date, the value of message delivery as a diagnostic tool has demonstrated that it has the potential to be a valuable tool in providing reliable assessments of media relations performance, as well as providing direction in improving performance of communication programs. It also has the potential to assess the relative performance of communication programs to provide context for interpreting the performance of a program. Most important, message delivery scoring has also provided support for the presumed belief that negative messaging has the potential to limit the impact of public relations programs even when positive or intended messages are delivered.

The next step in this process is to conduct extended experiments that continue to determine the validity and reliability of a message completeness scoring and its relationship to traditional content analysis metrics. It should be noted that this approach is based on an examination of message delivery in traditional media formats and has not examined an application of this method to social media analysis or other forms of media. The differences in format (e.g., 140 characters) and the management of these forms of media may require significantly different metrics (and Twitter's playing with an additional 140 characters for message lengths up to 280 characters may change even those employed at this time).

In addition, these experiments should be extended to include the relationship between message delivery scores and impact on message recipients. Such experiments would look at the relationship between the ability to include or place messages in intermediary sources and changes in the core metrics of communication impact including awareness, knowledge, relevance, intent, and advocacy.[22] This process can serve as the foundation for interpreting message completeness and the success of media relations programs.

---

[21] For a full discussion of these measures, refer to Michaelson & Stacks, 2017 page 28

[22] For full description of these measures, refer to Michaelson & Stacks, 2011 and Michaelson & Stacks, 2017 chapter 2.

References

Broom, G. M., & Dozier, D. M. (1990). *Using research in public relations: Applications to program management.* New York: Prentice Hall.

Coulter , K. S., & Sewall , M. A. (1995). The Effects of Editorial Context and Cognitive and Affective Moderators on Responses to Embedded Ads. *Advances in Consumer Research, 22*, 177-183. Retrieved June 23, 2017, from http://acrwebsite.org/volumes/7693/volumes/v22/NA-22

Institute for Public Relations. (2013, June 14). *The Conclave: Complete Social Media Measurement Standards June 2013*. Retrieved April 24, 2017, from Institute for Public Relations: http://www.instituteforpr.org/11596/

Institute for Public Relations. (n.d.). *Research categorized as: Research Methods / Standards*. Retrieved April 21, 2017, from Institute for Public Relations: http://www.instituteforpr.org/category/research/research-methods-standards/

Institute for Public Relations. (n.d.). *Research categorized as: Traditional Media Measurement*. Retrieved May 2, 2007, from Institute for Public Relations: http://www.instituteforpr.org/category/research/research-methods-standards/traditional-media-measurement/

Lindenmann, W. (2005, February). *Putting PR Measurement and Evaluation Into Historical Perspective.* Retrieved April 17, 2017, from Institute for Public Relations: http://www.instituteforpr.org/wp-content/uploads/PR_History2005.pdf

Lofgren, D. G. (2014, September 9). *Demonstrating PR's Value to C-Suite Executives*. Retrieved April 18, 2017, from PR News Online: http://www.prnewsonline.com/demonstrating-prs-value-to-c-suite-executives/

Macnamara, J. (2011). *PR Metrics: How to Measure Public Relations and Corporate Communication.* Retrieved April 17, 2017, from AMEC: http://amecorg.com/wp-content/uploads/2011/10/PR-Metrics-Paper.pdf

Michaelson, D., & Griffin, T. (2005). *The Media Reality Check: A New Model for Media Content Analysis.* Retrieved from www.instituteforpr.org/topics/new-model-for-media-content-analysis/

Michaelson, D., & Stacks, D. (2009, August). Exploring the Comparative Communications Effectiveness of Advertising and Public Relations: A Replication and Extension of Prior Experiments. (D. K. Wright, Ed.) *Public Relations Journal*.

Michaelson, D., & Stacks, D. W. (2011). Standardization in Public Relations Measurement and Evaluation. *Public Relations Journal, 5*(2).

Michaelson, D., & Stacks, D. W. (2017). *A Professional and Practitioner's Guide to Public Relations Research, Measurement, and Evaluation* (Third Edition ed.). New York: Business Expert Press.

Michaelson, D., & Stacks, D. W. (2017a). Objective Based Measurement for Public Relations. *In Preparation*.

Michaelson, D., Wright, D. K., & Stacks, D. W. (2012). Evaluating Efficacy in Public Relations/Corporate Communication Programming: Towards Establishing Standards of Campaign Performance. *Public Relations Journal*.

Newlands, M. (2014, December 3). *Why Public Relations is Important for your Business*. Retrieved April 18, 2017, from Inc.com: https://www.inc.com/murray-newlands/why-public-relations-is-important-for-your-business.html

Stacks, D. W. (2002). *Primer of Public Relations Research* (1st ed.). New York: Guilford Press.

Stacks, D. W. (2011). *Primer of Public Relations Research* (2nd ed.). New York: Guildford Press.

Stacks, D. W. (2017). *Primer of Public Relations Research* (3rd ed.). New York: Guilford Press.

Stacks, D. W., & Bowen, S. A. (2013). *Dictionary of Public Relations Measurement and Research* (3rd ed.). (D. W. Stacks, & S. Bowen, Eds.) Gainesville, FL: Institute for Public Relations. Retrieved April 24, 2017, from http://www.instituteforpr.org/wp-content/uploads/Dictionary-of-Public-Relations-Measurement-and-Research-3rd-Edition.pdf

Thorson, K., Michaelson, D., & et.al. (2015). Joining the Movement?: Investigating Standardization of Measurement and Evaluation Within Public Relations. *Research Journal of the Institute for Public Relations, 2*(1), 1-25. Retrieved April 17, 2017, from http://www.instituteforpr.org/wp-content/uploads/MichaelsonEtcArticle.pdf

Wetherhead, D. (2011, December 2). *Key Message Development: Building a Foundation for Effective Communications*. Retrieved April 24, 2017, from PR Say: https://prsay.prsa.org/2011/12/02/key-message-development-building-a-foundation-for-effective-communications/