12-2011

# A Content Analysis of Reliability in Advertising Content Analysis Studies.

Weize Wang
*East Tennessee State University*

# A Content Analysis of Reliability in Advertising Content Analysis Studies

---

A thesis

presented to

the Faculty of the Department of Communication

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Arts in Professional Communication

---

by

Weize Wang

December 2011

---

Stephen W. Marshall, Ph.D., Chair

John M. King, Ph.D.

Beichen Liang, Ph.D.

Keywords: Content analysis, reliability, intercoder reliability, advertising, communication, journal articles.

ABSTRACT

A Content Analysis of Reliability in Advertising Content Analysis Studies

by

Weize Wang

Content analysis is a systematic research method for examining symbolical content in communication by recording or transcribing these messages into categories. Reliability is one of the most distinctive attributes of content analysis methodology comparing to other techniques in communication. A content analysis was conducted by analyzing the method sections of published journal articles in *Communication Abstracts* from January 2006 through January 2011 by searching "advertising" and "content analysis". Results suggested that television is still the most focused medium in advertising content analysis research. Most of the content analysis studies employed 2 coders for coding reliability assessment data and final data. Moreover, content analysis researchers had improved in reporting reliability and reliability coefficients. However, there was a low percentage of studies that reported specific reliability for each variable as well as the lowest acceptable level for the reliability coefficients.

# ACKNOWLEDGMENTS

The following is the acknowledgement of my graduate thesis: At the very first, I am honored to express my deepest gratitude to my dedicated advisor, Dr. Stephen W. Marshall from Department of Communication, for his valuable support and suggestions. Dr. Marshall helped me by his profound knowledge and rich research experience. Besides, he put high priority on our thesis writing and was willing to discuss with me whenever I meet problems. I have learned a lot from him not only about thesis writing but also the professional ethics. I am also extremely grateful to Dr. John M. King and Dr. Beichen Liang who provided me patient, meticulous guidance and invaluable suggestions. I could not complete this work without their support and encouragement. In addition, I wish to extend my thanks to the faculty of Department of Communication and Department of Marketing for their support of this study. I owe special thanks to Dr. Dan Brown who is the graduate coordinator of Department of Communication for his priceless assistance during my graduate study. It was a pleasant experience and memory working on the thesis with them. Thanks are also due to my graduate friends who never failed to give me great encouragement and suggestions. Special thanks should go to Ms. Mariam Ayad, Weiwen Zhou, Beichen Wen, Yuan Long, Xiaonan Hu, and Dinah Tetteh.

At last but not least, I would like to thank my family for their support all the way from the very beginning of my graduate study. I am thankful to all my family members for their thoughtfulness and encouragement.

CONTENTS

Page

# CHAPTER 1

## INTRODUCTION

Content analysis is a fundamental method based on data generated from human observers' judgment by recording or transcribing the textual, visual, or audible messages in human communication (Krippendorff, 2004). The method has been widely used in various disciplines: communication and journalism, marketing, education, psychology, anthropology, and other social science subjects. Fowler (1986) suggested that content analysis was introduced in more than three fourths of research method courses at the master's level. In addition, there has been a growth in scholarly publications of content analysis: more than 2,000 results of published journal articles would emerge by searching "content analysis" between the year 2001 and 2011 from *Communication Abstracts*.

As a quantitative research method, the most essential advantages of this scientific research method are that it provides accurate insight of communication content, and its replicability (Berelson, 2000). However, because lacking of a standard reliability measure in content analysis, a growing concern of quality and creditability of content analysis studies has been debated in many methodology studies. In order to help with reliability in content analysis, some researchers provided guidelines and recommendations for enhancing quality of work, in terms of calculating coefficients of and reporting intercoder reliability (Neuendorf, 2002).

To some extent, this study is a replication of the study of Lombard,Snyder-Duch, and Bracken (2002) particularly in recent advertising research. We are employing content analysis method to investigate how communication researchers assessed and reported

intercoder reliability in recent published content analysis articles. This article introduces the content analysis methodology at first, and then demonstrates the significance of intercoder reliability. By reviewing the previous studies conducted to provide suggestions for enhancing reliability, we examine the method sections and procedures of reliability assessment of published content analysis articles under the recommended guidelines. The study examines recent studies to help researchers find out problems in their methodological decisions and improve the reliability of content analysis in future.

# CHAPTER 2

# LITERATURE REVIEW

## *Content Analysis*

Content analysis is a systematic research method for examining symbolical content in communication by recording or transcribing these messages into categories (Berelson, 2000; Stemler, 2001). This method is "the study of recorded human communications" (Babbie, 2000, p. 305). It is one significant technique in quantitative research and social sciences because it is an unobtrusive technique that enables researchers to examine messages "in view of the meanings, symbolic qualities, and expressive contents" (Krippendorff, 2004, p. 44) under certain contexts. Compared to other quantitative research, content analysis is more focused on validity and reliability. Researchers employ content analysis methodology to study "texts, images, and expressions that are created to be seen, read, interpreted, and acted on for their meanings" (Krippendorff, 2004, p. 13). Janis (2009) classified this research technique into three categories: pragmatical content analysis, semantical content analysis, and sign-vehicle analysis. Pramatical content analysis is the procedure in which messages are categorized by their possible causes or effects. Semantical content analysis is the procedure that classifying messages based on their meanings. Sign-vehicle analysis measures "what it purports to measure: the frequency occurrence of a given sign-vehicle" (Janis, 2009, p. 359).

The most essential advantages of this scientific research method are that it provides accurate insight of communication content and its replicability (Berelson, 2000). Benefits of using content analysis method to approach communication research are (1) content

analysis can be used as an unobtrusive measure of communications while direct methods might involve bias; (2) content analysis offers potentials for examining effects of various message-content on recipients' responses (3) content analysis initiates new research on specific subjects of communication; and (4) content analysis can be employed in multimethod research (Kolbe & Burnett, 1991). Although content analysis is specifically appropriate for and has been widely used in mass communication and journalism, applications of the study method have been increasingly employed in legal, political, marketing, and commercial matters as well. Moreover, other applications of the research method lie in other empirical domains as well, including "psychiatry, psychology, history, anthropology, education, philosophy and literary analysis, and linguistics" (Krippendorff, 2004, p. 46).

Nevertheless, problems of content analysis include "the effects of researchers biases" (Kolbe & Burnett, 1991, p. 244), lack of subtleties, and weakness in providing theoretical perspectives. Moreover, this method lacks of control of confounding extraneous factors (Holsti, 1969) and is only categorically descriptive (Kolbe & Burnett, 1991). Because of these constrains of content analysis methodology, a large number of content analysis studies do not have acceptable level of quality in using the methodology (Kassarjian, 1977; Kolbe & Burnett, 1991). Kassarjian (1977) discussed the problems of using content analysis methodology in early consumer studies. He suggested that objectivity, systematization, and quantification were the most significant and distinguishing attributes of content analysis. Definite rules and procedures should be provided as directions throughout the research process. Whether one study is replicable is the criterion of its reliability (Kassarjian, 1977).

### *Reliability and Intercoder Reliability*

Reliability was defined as the extent of replicability, which is pursuing the same results through repeated measuring procedures (Neuendorf, 2002). Reliability is one of the most distinctive attributes of content analysis methodology comparing to other techniques in communication. The independence of the procedure of measuring, instrument, and judgers builds up the significance of reliability. Reliability is essential in content analysis because the goal of this research method is to achieve identifying and recording characteristics of messages objectively. However, the process of data collection in content analysis methodology is often conducted by human observers through recording or transcribing texts, pictures, or audio recordings. Therefore, reliability in content analysis studies is necessary to demonstrate to ensure the trustiness of conclusions from such data (Neuendorf, 2002).

In content analysis reliability is interpreted as intercoder reliability or the extent of agreement. Intercoder reliability is the term generally used to represent that "the extent to which independent coders evaluate a characteristic of a message or artifact and reach the same conclusion" (Lombard et al., 2002). Specifically, intercoder agreement is more appropriate for the particular characteristic of required consistency in content analysis studies. Intercoder reliability is perceived as the paramount goal and "the standard measure of research quality" (Kolbe & Burnett, 1991, p. 248). It is the indicator of whether research has weakness in its method and operational procedures. To the practical benefit of establishing intercoder reliability in content analysis research, high levels of intecoder reliability is functional for researchers to split the coding work to several coders (Lombard et al., 2002). Moreover, the assessment of reliability not only

enables coding to be efficient but also enables the entire procedure of content analysis to be supported by reviewers under scrutiny. Intercoder reliability is significant in content analysis especially when the communication content is assigned to multiple judges to code.

Although reliability is not sufficient for validity, validity cannot be established without reliability (Neuendorf, 2002). To establish intercoder reliability, more than one judge should be employed to code the same messages independently in human-coding content analysis. One important measure of reliability is category reliability which is based on researchers' ability to develop categories and present clear definitions of the categories to other coders so that coders will agree on coding decisions (Kassarjian, 1977). The category reliability is to achieve that judgers' understanding of categories is sufficiently specified for scientific usage. The degree of consistency among different coders to the same content and categories is the focus of intercoder reliability. The ratio of coding agreements to the number of all decisions made by coders is the most popular coefficient and is generally employed to determine reliability. The assessment of reliability not only enables coding to be efficient but also enables the entire procedure of content analysis to be supported by reviewers under scrutiny (Lombrad et al., 2002).

Kolbe and Burnett (1991) suggested that procedural issues most influence intercoder reliability. Issues included rules and procedures, coder training, measures pretesting, information of coders, and independence of coding. Whether these issues are reported in the text of studies affect the intercoder reliability assessment. Detailed information of operational instructions and coding procedures is required to ensure intercoder reliability. Explaining precise rules and procedures also reduces coders' personal biases and affords

reproducibility. Researchers conceive intercoder reliability as the criterion of research quality in content analysis studies. A high percentage of disagreement between coders illustrates problems in research methods signaling poor operational instructions, variables, or coder training as possible weaknesses. Calculation and reporting of reliabilities are other two significant components of intercoder reliability.

## *Intercoder Reliability Coefficients*

To measure intercoder reliability, approximately 40 different methods can be used to calculate coefficients for reliability of nominal data. However, only several indices are employed widely in communication and related research (Lombard et al., 2002). Intercoder reliability coefficients are to assess the level of agreement between coders' decisions. Calculating intercoder reliability coefficients across divergent variables is an inappropriate approach because the low reliability for certain variables that do not reach the acceptable criterion would be averaged by other variables' high reliability and be hidden. Neuendorf (2002) suggested researchers report intercoder reliability for each specific variable to ensure the trustfulness of intercoder reliability indices. We introduce these measures with mathematical facts that offer a general view of the various indices.

### *Percentage of Agreement and Holsti's Method*

*Percentage of agreement* is the simple percentage of agreement among all coders' decisions in coding the same units of data (Neuendorf, 2002). This measure is the most popular coefficient because it is easy to understand and calculate, as well as this method also can be applied to more than two coders (Lombard et al., 2002). The indices of the

measure range from .00 (no agreement) to 1.00 (complete agreement). The conceptual formula of Percentage agreement is following:

$$PA_o = A/n$$

where $PA_o$ represents observed propotion of agreement, A is the number of coders' consensus decisions, and n is the total number of decisions the two coders have made.

Holsti's method (1969) is a variation of percentage agreement. Percentage agreement and Holsti's method (1969) would be equal when two coders code the same units of sample. Compared to percentage agreement, Holsti's method (1969) is applicable to situations in which two coders code different units of the sample. The formula is:

$$PA_o = 2A/ (N1+N2)$$

where $PA_o$ represents percentage of agreement between two coders, A is the number of two coders' consensus decisions, and N1 and N2 are numbers of decisions coders have made respectively.

Drawbacks of these two coefficients are in many dimensions. The first flaw of the indices is they lack of ability to calculate the agreement by chance. According to percentage of agreement, the probability of agreement by chance for two coders is 50%, and for three coders is 33.3%. However, in fact two coders' agreement could not be 50% automatically all the time due to chance (Riff, Lacy, & Fico, 1998). Another particular limitation of the two coefficients is that inflation of reliability may occur by increasing the number of categories when researchers know the categories will not be used frequently or the decisions are easy to be agreed on. In addition, it is difficult to judge the true reliability with percentage agreement among different variables. Moreover, the

indices only figures out agreements and disagreements, but provides no evidence of which coder's decisions are valid.

*Scott's pi and Cohen's kappa*

*Scott's pi* improves on simple percent agreement by "the agreement that is expected when the units are statistically unrelated to their descriptions" (Krippendorff, 2007, p. 80). This index takes category values into consideration and accounts for chance agreement. The number of variables and the distribution of categories, which indicates how coders use the categories in coding, are taken into account with this coefficient. The index calculates "the agreement expected by chance by looking at the proportion of times particular values of a category are used in a given test" (Riff et al., 1998, p. 129), and then chance agreement and expected agreement are calculated. Cohen's kappa was to improve pi in take discrepancy of coders' distributions into consideration by multiplicative marginals instead of additive marginals (Neuendorf, 2004). It was reported as the most widely used index for reliability (Perrault & Leigh, 1989).

However, a limitation of the methods is they tend to be conservative because the proportions of the distribution across categories are not coders' agreement but true proportion (Lombard et al., 2004). In addition, these coefficients ignore the diverse distributions of coders' values across different variables that may generate bias in data. These two coefficients are only applicable to situations of nominal data and two coders. Coders are not able to interchange each other's coding units (Krippendorff, 2007).

The conceptual formula for pi and kappa is:

$$\text{Pi or Kappa} = (PA_o - PA_E)/ (1 - PA_E)$$

where $PA_o$ stands for "observed percentage of agreement", and $PA_E$ is "proportion agreement, expected by chance".

*Krippendorff's alpha*

*Krippendorff's alpha* is a satisfied coefficient in many aspects (Krippendorff, 2004). It can accommodate to multiple coders and the situation that individuals are assigned different units during coding. It accounts for various sample sizes and missing information. It can be applied to different levels of variables (ordinal, internal, and ratio variables). When two judges code a large sample that are nominal data, the results of Krippendorff's alpha and Scott's pi will be equal. The coefficient also is applicable to data with missing values when some coders do not participate in coding all units. However, the defect of the index is it extremely difficult and complex to calculate by hand; and little support from software has been generalized to content analysis research (Lombard et al., 2002).

The formula for alpha is:

$$\alpha = 1 - D_o/D_e$$

where $D_o$ is observed disagreement and $D_e$ is expected disagreement.

*Perreault and Leigh's Method (1989)*

Perreault and Leigh (1989) developed *Ir* as a reliability index for content analysis. The approach corrects the problems of Cohen's kappa and differed from kappa in estimating chance agreement. The model provides a direct approach to compute reliability, and the index does not rely on marginal frequencies like Cohen's kappa. The formula for *Ir* is:

$$Ir = \sqrt{\{[(F_o/N) - (1/k)]\ [k/(k-1)]\}}$$

where $F_o$ stands for the number of coders' agreement, N is the total number of coders' decisions, and k is number of categories.

### *The Standard for Intercoder Reliability*

Lombard et al. (2002) provided researchers several steps of assessing intercoder reliability to follow: employing one or more indices to determine the intercoder reliability; applying software as tools to calculate the indices; setting an acceptable minimum standard of reliability; using informal reliability assessment during coder training and formal reliability assessment before the study begins and during coding the full sample; incorporating coding the reliability sample into the procedure of coding the full sample; and reporting intercoder reliability clearly and explicitly. To report the intercoder reliability clearly, researchers should explain the size, method, number of reliability coders, coding amount for each variable, intercoder reliability for each variable, the type of method to calculate coefficients, training amount, and where and how the complete information of the coding measurements, procedures, and guide could be found. We examine the intercoder reliability through the most frequently employed method of reliability coefficients, tendencies in coding procedure and rules, training amount, using computing tools, methods of study, and reporting reliability.

Krippendorff (2004) provided three conditions of using an agreement coefficient to test intercoder reliability in content analysis. First, the applied data should be reliable. Such data are collected by different coders through their independent duplications of the coding, categorizing, and measuring processes to the same units of messages according to the same coding guide. Second, coders should treat the analyzed units of messages

separately. Third, researchers must accept assessing reliability rely on imperfect data. The sample of reliability assessment should permit disagreement between coders to avoid biases.

The debate of the standards of intercoder reliability for measurable variables from previous studies has suggested that .80 or greater could be the lowest acceptable level in most of the time. Riff et al. (1998) demonstrated that the lowest acceptable level of reliability coefficient should be based on the studied categories. If categories and definitions in one study have already been widely defined and studied, similar higher levels of reliability are expected. The range of the lowest acceptable level of reliability coefficients usually is between .80 and .90. The minimum requirement for reliability indices was suggested as .70. Research with reliability lower than that would be doubted with its method and value, and would be difficult to interpret. The beyond-chance statistics including the Scott's pi and Cohen's Kappa allow a looser criterion. Neuendorf (2002) suggested that when there is a lack of a uniform criterion of meaningful significance in content analysis, the expectation for researches is fully clarifying reliability coefficient separately for every measured variable. Kassarjian (1977) stated that the lowest acceptable level of coefficients of reliability should be above .80. In sum, the intercoder reliability could be satisfied in studies where coefficients are above .85.

### *Studies of Intercoder Reliability in Content Analysis*

The goal of studying research method is to "help a discipline improve" (Riffe & Freitag, 1998). Perreault and Leigh (1989) paid attention to the quality of nominal data that were coded by human judges. The researchers discussed the advantages and limitations of the existing measures for reliability assessment and then they developed a

new appropriate method for calculating intercoder reliability in marketing research. Several suggestions were offered for content analysis researchers that explicit category definitions and coding guide should be provided; multiple coders should be recruited in large sample studies; assessment of specific variable and category would help identify ambiguous; and pretest would help build high level of intercoder reliability.

Kolbe and Burnett (1991) conducted a study to investigate the reliability and objectivity of content analysis research in consumer studies. The researchers examined the methods of 128 content analysis studies from consumer behavior research according to directives for content analysis as requested by Kassarjian (1977) from dimensions of objectivity, quantification, sampling, and reliability to improve the research method of content analysis. The results suggested that percentage of agreement was the most popular reliability index and 32% of the content analysis articles employed this method to calculate intercoder reliability. At the same time, 31.3% of the studies did not report any coefficient of intercoder reliability, with an additional 19% ambiguously reported the calculating method for reliability. To reporting reliability index, 35.9% of articles reported overall average reliability; 24.2% of articles reported reliability on individual variables.

A similar investigation was conducted by Riffe and Fritag (1997) where the researchers examined content analysis articles which were published from 1971 through 1995 in *Journalism & Mass Communication Quarterly*. Researchers explored how the content analysis published in journals had changed during the 25 years in concentration of media and content, sampling, reliability reporting, and the use of other method. Results of the study suggested that the main concentration was on news content in American

media. There were few studies involved a second method besides content analysis. Half of the examined articles reported intercoder reliability, and there was a growth of reporting reliability in mass communication research.

To examine the coders' impact on reliability, Peter and Lauf (2002) investigated how the characteristics of coders influence intercoder reliability in cross-national content analysis. Language skills, political knowledge, and coding experience were examined as affecting factors to intercoder reliability. Results suggested that coders with higher level of language skills and political knowledge coded more reliably. At the same time, coding experiences did not have an impact on intercoder reliability.

In addition to different types of content in studies, Potter and Levine-Donnestein (1999) claimed that the assessment method of reliability should differ based on that the diverse types of content researchers analyzed and whether the research was based on theory. When the content analysis study is designed without a guide of theory, the researchers do not have a solid ground of developing a coding scheme, so that they must be more careful to set the coding scheme. Employing multiple coders to code the same overlapping messages would improve the convincingness of testing the consistency of coders' decisions. When calculating the intercoder reliability coefficient, Potter and Levine-Donnestein (1999) suggested that percentages of agreement should be applied to a formula that gets rid of chance agreement.

Lombard et al. (2002) conducted research to assess intercoder reliability through intercoder agreement. They examined 200 articles from years 1994 to 1998 sampled from the *Communication Abstracts* database by searching "content analysis" as the keywords. Results suggested that researchers in mass communication field usually failed to evaluate

intercoder reliability and depended on percent agreement. Lombard et al. indicated that only 69% of content analysis articles reported intercoder reliability, and only 41% of articles discussed reliability for specific variables. The specific index for reliability was not reported in most situations, and Holsti's method was the most frequently reported as the calculating method and accounted for 15%. Meanwhile, the Scott's *pi* accounted for 10% of the reported method, percent agreement accounted for 9%, Cohen's kappa accounted 7%, and Krippendorff's alpha accounted for 3%. There were only 2% of the collected articles reported the computing tools they employed to calculate intercoder reliability coefficients. The lowest standard of acceptable intercoder reliability was reported as .75 from the articles; however, the minimum reported reliability was .40 at the same time. Nine percent of the collected content analysis studies did not report the information of how many coders participated in reliability coding. Only 41% mentioned the reliability for specific variables; 14% reported the coders' training amount.

Marshall and Roberts (2008) discussed the appropriate approach to improve objectivity and reliability. They provided other researchers a framework of criteria to follow that : (1) addressing rules and procedures clearly with offering category details and clear coding guide; (2) judge training; (3) pretesting during initial coder meeting; (4) coding independently; (5) reporting the number of coders; and (6) evaluation of intercoder reliability coefficients with more than one indices.

### *Research Question*

Based on the findings concluded by Lombard et al. (2002), this study is to examine the applications of content analysis methodology through evaluating the intercoder reliability within advertising research in the previous 5 years. Our research is focusing on

the problems Lombard et al. (2002) found in their study and seeking whether there is an improvement in intercoder reliability of content analysis studies in the recent 5 years. The exploratory research question is: How adequately and consistently has reliability been assessed and reported in published advertising studies with content analysis method?

# CHAPTER 3

# METHOD

Previous studies indicated that there is a call for improving reliability in content analysis research. In demonstrating the reliability in recent content analysis studies, we are interested in how researchers reported reliability in content analysis studies. We employed content analysis to review published journal articles in advertising in which content analysis was the primary research method from *Communication Abstracts* between January 2006 and January 2011.

Objectivity and reliability is the core of content analysis methodology. Kolbe and Burnett (1991) found that there were many gaps between the requirements of content analysis methodology and the operational procedures of researchers in objectivity and reliability. In order to ensure the objectivity, Kolbe and Burnett suggested researchers provide clear coding rules and procedures, report coder training, pretest measures, code independently, and not participate in coding if the researchers were the authors. They also investigated calculating and reporting intercoder reliability by examining the selective use of intercoder reliability index and how the indices were reported. They suggested researchers report intercoder reliability for each category because the overall reliability would not be trustworthy and would obscure results.

## *Sample*

*Communication Abstracts* was selected as the archive of the study. The sample in this study was obtained by searching content analysis articles that were published between January 2006 and January 2011 by using the keywords “content analysis” and

"advertising". The initial results for our search were 163 journal articles. However, after eliminating articles that were in Spanish, did not employing content analysis as the research method, and could not be found from library resources, 91 articles from 41 journals were selected for coding. The unit of analysis was the method section of each content analysis journal article collected.

### *Rules and Procedures*

Variables of the studies were: publication year of the study; publication journal name; the method used in the research (all quantitative, some quantitative and some not, or not quantitative); what medium was analyzed (newspapers, magazines, television, internet, radio, film, data from respondents, other printed medium, or other electronic medium); whether the information of the coders were reported (reported or not reported). number of coders who participated in coding the actual sample; whether the training amount was reported (reported or not reported); whether the intercoder reliability was discussed in the study (yes or no); the name of reliability method used in the study (Krippendorff's alpha, Scott's pi, Cohen's Kappa, Holsti's method, "simple agreement only or "percentage agreement" only or "intercoder reliability" only, more than one method, Perreault and Leigh's (1989) method, or other); whether the lowest specific reliability criterion was reported (yes or no); whether the specific reliability for one or more variables was reported (reported or not reported); whether the specific reliability for each variable was reported (reported or not reported); computing tools were used to calculate reliability (yes, no, or not mentioned).

### *Judge and Judge Training*

After developing the coding scheme for the reliability study, the author and another coder, who are both graduate students majoring in communication, participated in the coding work. We coded 20 articles during the training, and then we conducted training for half an hour under the guidance of a professor who is highly experienced and knowledgeable in content analysis. Questions and disagreement were discussed before coding the real sample. The full sample was coded by the both coders independently and each of them coded 91(100% of the sample) articles. We randomly selected 50 articles for the reliability test sample. To develop the final dataset, the coding decisions were randomly selected from both of the coders' coding results.

Percent agreement, Scott's pi, Cohen's kappa, Perreault and Leigh's (1989) method, and Krippendorff's alpha were employed as coefficients to examine intercoder reliability for each variable. SPSS Macro was used to calculate Krippendorff's alpha. Percent agreement, Scott's pi, Perreault and Leigh's (1989) method, and Cohen's kappa were calculated by hand. Holsti's method was not applied in this study because the two coders coded the same amount of the reliability sample. The lowest acceptable level for for Scott's pi, Perreault and Leigh's Ir, and Cohen's kappa is .80, and for Krippendorff's alpha is .70. If this was not the case, percent agreement should be .90 or higher (Lombard et al., 2002).

# CHAPTER 4

## RESULTS

We analyzed 91 content analysis studies in advertising from 41 journals during the years 2006 to 2011. Results showed that14.3% (n= 13) of researched articles were published in 2006, 27.5% (n= 25) in 2007, 25.3% (23) in 2008, 18.7% (n=17) in 2009, and 14.3% (n= 13) were published in 2010. The results for all variables in our study are shown in Table 1.

Results suggested that 84.6% (n= 77) of the content analysis articles were all quantitative in nature, and 15.4% (n= 14) involved other methods at the same time. The analyzed medium of content analysis research in advertising has been focused on television in the last 5 years: 36.3% (n= 33) of the research sample selected television as the primary medium to conduct research. Magazines and internet are the second and third favorable medium for researchers to analyzed with the technique of content analysis, that the percentage of the studied articles are 20.9% (n= 19) and 16.5% (n= 15) respectively. In addition, 23.1% of the articles involved more than one type of media.

We found that most of the studies provided coders' information that coders usually were recruited from students in universities and the authors themselves. We found that 87.9% (n= 70) coders' information and most of them employed two coders (61.5%) to conduct coding the actual study sample, which is in accordance with Kolbe and Burnett's (1991) findings. Moreover, the data revealed that 85.7% (n= 78) of the studied sample discussed and reported intercoder reliability. There was 53.8% (n= 49) of the sample reported with specific reliability for at least one variable, and only 36.3% (n= 33) reported with specific variable for each variable in their studies. Training amount of

coding was reported as number of hours coders had spent or number of units coders had coded before coded actual sample. There were less than half (46.2%) of the articles reported training amount for coding in advance.

The coefficients used to calculate intercoder reliability were reported in most of the articles (90%, n=82); in the subsample of articles that reported the specific index of intercoder reliability, percentage agreement (22.0%) was reported as the most frequent coefficient used by researchers; Perreault and Leigh's (1989) method (19.5%) was the reported as the second common method to develop intercoder reliability, and succeeding with Cohen's kappa (17.1%), Scott's pi (14.6%), and Holsti's method (6.1%). In addition, 14.6% (n=12) articles employed more than one method to calculate the intercoder coefficients that percentage agreement was contained as one of the indices.

Nevertheless, the lowest acceptable levels of intercoder reliability coefficients were not reported in most of the studies. Only 28.6% (n= 26) mentioned the criterion for the employed indices of intercoder reliability. Computing tools for calculating intercoder reliability coefficients, such as calculating by hands and by software, were little demonstrated in the articles: only 8.8% (n= 8) articles mentioned the tools researchers employed to assess intercoder reliability.

Table 1

*Intercoder Reliability and Percentages and Means for All Variables*

| Variable | Percentage agreement | Scott's *pi* | Cohen's *kappa* | Krippendoff's *alpha* | Perreault and Leigh's (1989) *Ir* | % (n) or mean |
|---|---|---|---|---|---|---|
| *What is method of study in nature?* | 0.9400 | 0.8667 | 0.7664 | 0.7671 | 0.9539 | |
| All quantitative | | | | | | 84.6% (77) |
| Some Quantitative, some not | | | | | | 15.4% (14) |
| Not quantitative | | | | | | 0 |
| *What is the main medium analyzed?* | 0.9800 | 0.9725 | 0.9725 | 0.9728 | 0.9887 | |
| Newspapars | | | | | | 11.0% (10) |
| Magazines | | | | | | 20.9% (19) |
| Television | | | | | | 36.3% (33) |
| Internet | | | | | | 16.5% (15) |
| Radio | | | | | | 1.1% (1) |
| Film | | | | | | 1.1% (1) |
| Data from respondents | | | | | | 6.6% (6) |
| other print medium | | | | | | 6.6% (6) |
| other electronic medium | | | | | | 0 |
| *Second medium analyzed* | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | |
| *Third medium analyzed* | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | |
| *Was information of coders reported?* | 0.9600 | 0.6454 | 0.6479 | 0.6489 | 0.9592 | |
| Reported | | | | | | 87.9% (80) |
| Not reported | | | | | | 12.1% (11) |
| *Numbers of codes who participated in coding the actual sample* | | | | | | |
| 1 Coder | | | | | | 1.1% (1) |
| 2 Coders | | | | | | 61.5% (56) |
| 3 Coders | | | | | | 8.8% (8) |
| 4 Coders | | | | | | 4.4% (4) |
| 5 Coders | | | | | | 4.4% (4) |
| 6 Coders | | | | | | 2.2% (2) |
| 7 Coders | | | | | | 1.1% (1) |
| 16 Coders | | | | | | 2.2% (2) |
| 25 Coders | | | | | | 2.2% (2) |
| *Was the training amount reported ?* | 0.9400 | 0.8776 | 0.8780 | 0.8788 | 0.9381 | |
| Reported | | | | | | 46.2% (42) |
| Not reported | | | | | | 53.8% (49) |
| *Was the intercoder reliability discussed?* | 0.9200 | 0.5914 | 0.8980 | 0.8989 | 0.9165 | |
| Reported | | | | | | 85.7% (78) |
| Not reported | | | | | | 14.3% (13) |

Table 1 Continued

*Intercoder Reliability and Percentages and Means for All Variables*

| Variable | Percentage agreement | Scott's *pi* | Cohen's *kappa* | Krippendoff's *alpha* | Perreault and Leigh's (1989) *Ir* | % (n) or mean |
|---|---|---|---|---|---|---|
| *What is the method to assess reliability?* | 0.9200 | 0.9073 | 0.9028 | 0.8978 | 0.9532 | |
| Krippendorff's alpha | | | | | | 4.4% (4) |
| Scott's pi | | | | | | 13.2% (12) |
| Cohen's kappa | | | | | | 15.4% (14) |
| Holsti's method | | | | | | 5.5% (5) |
| "simple agreement" only, or "percentage agreement" only, or "intercoder reliability" only | | | | | | 19.8% (18) |
| more than one method | | | | | | 13.2% (12) |
| Perreault and Leigh's(1989) method | | | | | | 17.6% (16) |
| other | | | | | | 1.1% (1) |
| *Was the lowest specific reliability criterion reported?* | 0.9400 | 0.8597 | 0.8600 | 0.8612 | 0.9381 | |
| Reported | | | | | | 28.6% (26) |
| Not reported | | | | | | 71.4% (65) |
| *Was the specific reliability for one or more variables reported?* | 0.9600 | 0.9192 | 0.8790 | 0.8800 | 0.9592 | |
| Reported | | | | | | 53.8% (49) |
| Not reported | | | | | | 46.2% (42) |
| *Was the specific reliability for each variable reported?* | 0.9400 | 0.8681 | 0.8690 | 0.8695 | 0.9381 | |
| Reported | | | | | | 36.3% (33) |
| Not reported | | | | | | 63.7% (58) |
| *Was any computing tools used to calculate reliability?* | 0.9800 | 0.8464 | 0.8470 | 0.8479 | 0.9798 | |
| Yes | | | | | | 8.8% (8) |
| No or not mentioned | | | | | | 91.2% (83) |

# CHAPTER 5

# CONCLUSION

## *Improvement in Intercoder Reliability*

This content analysis has investigated the significant problems Lombard et al. (2002) demonstrated in assessing and reporting intercoder reliability which generated uncertainty in the validity of recent content analysis studies in mass communication. Kolbe and Burnett (1991) revealed only 35.9% of consumer behavior studies that were conducted with content analysis methodology from 1978 to 1989 reported reliability indices. Riffe and Feitag (1997) examined 25 years content analysis from 1971 to 1995 and found that 56% of the articles reported reliability coefficients and most of the studies neglected reporting intercoder reliability by variables. Lombard et al. discovered 69% (n=137) of content analysis studies published in *Communication Abstracts* mentioned intercoder reliability.

As we expected, there was an improvement in assessing and reporting interncoder reliability in mass communication research. Comparing to the findings of Lombard et al. (2002), reporting coders' information in content analysis has been improved so that 87.9% of the studies during the last 5 years reported coders' information while the figure was only 67% from 1994 through 1998. Training amount of coding usually is reported as specific numbers of hours or units of data judgers used to practice coding before code the actual sample. It has been increasingly reported (46.2%) comparing to Lombard et al.'s findings: only 9% of content analysis mentioned training amount in their study. The percentage of studies that discussed intercoder reliability (85.7%) also increased by

contrast with 69% in the early study. According to the results, researchers in advertising research have increasingly reported specific reliability for each variable as well as computing tool in calculating intercoder reliability coefficients.

Although our findings suggested that many dimensions of intercoder reliability have been improved in recent years, there is still a call to enhance reliability for content analysis. Small portion (about one third) of the recent studies reported reliability for each variable, less than half of the studies reported training amount, the lowest acceptable level of reliability was not reported frequently.

### *Intercoder Reliability Measures and Computing Tools*

Percentage agreement has kept being the most widely used method of calculating intercoder reliability coefficient which is in accordance with the results of Lombard et al. (2002). Using percentage agreement is the easiest way to calculate intercoder reliability indices, but it is not a sufficient support to achieve accurate intercoder reliability. In order to ensure the accuracy of intercoder reliability, researchers were recommended to employ more than one method to calculate intercoder reliability coefficients, and the current results showed that about 13% of the studies conformed to this suggestion. Interestingly, different from Lombard et al.'s findings, Perreault and Leigh's (1989) method has become the second favorable index for intercoder reliability in recent years. The index *Ir* is easy to calculate as well as it is an approach which provided the possibility of employing "an explicit model of level of the agreement" (p. 140) that to achieve the true level of reliability.

Most of the studies failed to provide information about computing tools used to calculate intercoder reliability coefficients. We calculated all the indices both by hand

and using software that Percentage Agreement, Scott's pi, Cohen's kappa, and Perreault and Leigh's (1989) *Ir* were calculated by hand, while Krippendorff's alpha was calculated by SPSS Macro. During the process of calculating the intercoder reliability for each variable, we found the problems of reliability software were: the number of available software for intercoder reliability was limited; some software could only applicable to one specific index; some helpful software was difficult to obtain online. Lombard et al. (2004) introduced several specialized software and statistical software for intercoder reliability: AGREE, which is to calculate Cohen's kappa; Krippendorff's alpha 3.12a, which is to calculate Krippendorff's alpha; PRAM; Simstat; and SPSS. However, AGREE costs users more than 400 dollars to access, meanwhile Krippendorff's alpha is a beta version that is not distributed widely. PRAM and Simstat are designed to calculate a variety of reliability indices, but the online resource of PRAM is hard to access, while Simstat is also paid software. SPSS is general used as statistical software that could calculate Cohen's kappa and Krippendorff's alpha with installing a Macro package. Some Macro packages for SAS to calculate Krippendorff's alpha are available online, but there is no widely defined syntax of SAS code that applied to reliability and it is sophisticated statistical analysis software that requires a high level of understanding of statistics and SAS code.

### *Method and Studied Media in Advertising Content Analysis*

Budd, Thorp, and Donohew (1967) suggested that researchers should use multiple research methods rather than employ content analysis as the single method. Kolbe and Bernett (1991) also suggested that content analysis offered researchers the potential of conducting multimethod research along with the method. Employing different methods

with content analysis research would improve the validity of results and reduce biases of using one single method. Our findings conclude that most of the content analysis studies in advertising were conducted with a single method. Researchers employed other method along with content analysis involved survey or qualitative research method (interview and focus group) particular to their studies.

Riffe and Fritag (1997) found the concentration of content analysis research from 1971 through 1995 was on American news media: newspapers and television. Our results conclude that television advertising was the focus of advertising research of content analysis. Magazines and internet advertising ranked as the second and third focus of advertising content analysis research.

### *Limitations*

The first limitation is the archive of *Communication Abstracts* itself that using a single database in study may create problems in reliability and validity because that each vendor has its strengths as well as its selective exclusive control of collections (Neuendorf, 2002). Communication journals that are not available in the database were neglected in our study. This influenced our findings. Moreover, the discrepancy between electronic journals and print versions of publications is a drawback to conducting content analysis with database. Kaufman, Dykers, and Caldwell (1994) compared the results of data collected by hand from print versions and from databases online. Results showed that data collected from online databases were different from the print versions so that collecting data from electronic databases would affect the results and reliability of content analysis.

Another limitation of the study is the sample collected from content analysis in advertising research may be insufficient to ensure the representability of all content analysis studies during the same time period.

Coders who are graduate students in the communication program in this study were trained under the direction of an experienced professor with content analysis. All the variables were coded by each coder independently. The intercoder reliability index for the variable of "intercoder reliability discussion" scored as low as .59 for Scott's pi, meanwhile the indices for "coder information" were also as low as .65 for Scott's pi, Cohen's kappa, and Krippendorff's alpha. Because all of the data were from manifest content, coding discrepancies could be caused by obscure definitions. The measures in our study were established by replicating Lombard et al.'s (2002) instrument, which was tested and revised several times by the researchers. Our coding instrument was modified in some variables from the original instrument to adapt to our current advertising research. However, the categories of the two variables should be defined more clearly in the future research.

### *Conclusion*

Researchers have been calling for improving intercoder reliability for content analysis studies during the last 2 decades. This study is a replication of Lombard et al.'s (2002) research in the field of advertising. By comparing our findings with the earlier studies, we conclude that although improvement has been made in assessing and reporting reliability in content analysis, researchers still need to pay more attention to reporting training amount, specific reliability for each variable, the lowest acceptable level of reliability index, and computing tools. It is appropriate to employ more than one

intercoder reliability coefficients in research. We hope our findings will shed a light on researchers' understandings in reliability of the recent content analysis studies.

REFERENCES

Babbie, E. (2001). *The practice of social research* (9th ed). Belmont, CA: Wadsworth.

Berelson, B. (2000). Content analysis in communication research. In P. Marris & S. Thornham (Eds.), *Media studies: A reader* (pp. 200-209). New York, NY: New York University Press.

Budd, R. W., Thorp, R. K., & Donohew, L. (1967). *Content analysis of communication*, New York: Macmillan.

Fowler, G. L. (1986). Content and teacher characteristics for master's level research course. *Journalism Quarterly*, *63*, 594-599.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methonds and Measures*, *1* (1), 77-89.

Hughes, M. A., & Garrett, D. E. (1990). Intercoder reliability estimation approaches in marketing: A generalizability theory framework for quantitative data. *Journal of Marketing Research*, *27*(2), 185-195.

Holsti, O. R. (1969). *Content analysis for the social sciences and humanities,* Reading, MA: Addison-Wesley.

Janis, I. (2009). The problem of validating content analysis. In K. Krippendorff, & M. A. Bock (Eds.), *The content analysis reader* (pp. 358-366). Thousand Oaks, CA: Sage.

Kaufman, P. A., Dykers, C. R., & Caldwell, C. (1994). Why going online for content analysis can reduce research reliability. *Journalism Quarterly*, *70*, 824-832.

Kassarjian, H. H. (1977). Content analysis in consumer research. *The Journal of Consumer Research*, *4* (1), 8-18.

Kolbe, R. H. & Burnett, M. S. (1991). Content-analysis research: An examination of applications with directives for improving research reliability and objectivity. *Journal of Consumer Research, 18*, 243-250.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed). Thousand Oaks, CA: Sage.

Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research, 30*, 411-433.

Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*, 77-89.

Lombard, M., Snyder-Duch, J., & Bracken, C.C. (2004). Practical resources for assessing and reporting intercoder reliability in content analysis research projects. *Retrieved on September 4, 2011, from* http://astro.temple.edu/~lombard/reliability/.

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2004). A call for standardization in content analysis reliability. *Human Communication Research, 30*, 434-437.

Marshall, S. W., & Roberts, M. S. (2008). *Television advertising that works: An analysis of commercials from effective campaigns*. New York: Cambria Press.

Nueuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.

Perreault, W. D., & Leigh, L. E. (1989). Reliability of nominal data based on qualitative judgements. *Journal of Marketing Research, 26*, 135-148.

Peter, J., & Lauf, E. (2002). Reliability in cross-national content analysis. *Journalism and Mass Communication Quarterly, 79*, 815-832.

Potter, W. J., & Levine-Donnerstein, D. (1999). Rethinking validity and reliability in Content Analysis. *Journal of Applied Communication Research, 27,* 258-284.

Riffe, D., & Freitag, A. (1997). A content analysis of content analyses: Twenty-five years of journalism quarterly. *J & MC Quarterly, 74,* 873-882.

Riffe, D., Lacy, S., & Fico, F. G. (1998). *Analyzing media messages.* Mahwah, NJ: Lawrence Erlbaum Associates.

Stemler, S. (2001). An overview of content analysis. *Practical Assessment, Research & Evaluation, 7* (17). Retrieved on October 7, 2011 from http://pareonline.net/getvn.asp?v=7&n=17/

# VITA

## WEIZE WANG

| | |
|---|---|
| Personal Data: | Date of Birth: June 4, 1986 |
| | Place of Birth: Hulubeier City, Inner Mongolia, China |
| Education: | B.S. Advertising, Nanchang University |
| | Nanchang, Jiangxi, China, 2008 |
| | M.S. Professional Communication, East Tennessee |
| | State University, Johnson City, Tennessee, 2011 |
| Professional | |
| Experience: | Research Assistant, East Tennessee State University |
| | Johnon City, Tennessee, 2010-2011 |