Teaching Intercoder Reliability: A Gentle Introduction to Content Analytic Methods for Graduate Students.

Patric R Spence, University of Kentucky
Kenneth Lachlan, University of Connecticut

Available at: https://works.bepress.com/patric_spence/39/
Teaching Intercoder Reliability: A Gentle Introduction to Content Analytic Methods for Graduate Students.

Patric R. Spence & Kenneth A. Lachlan

Quantitative content analysis has grown over the course of the last fifty years providing scientific evaluations of the nature of varying program elements in written, print, audio, and visual media. Despite its tremendous utility in communication research, graduate students often express apprehension at learning basic content analytic techniques. This is especially common among students who are fearful of quantitative research methods, or who are apprehensive about their own mathematical abilities. On the other hand, some students who have such apprehensions may actually move toward the method of content analysis due to an erroneous belief that it is not statistical, not mathematically rigorous, or because they believe it’s easy. This article provides a discussion of the content analytic methods with a few exercises that can be used to teach content analysis and intercoder reliability in a way that does not intimidate those students who may be apprehensive of quantitative methods, but at the same time presents content analytic techniques as sophisticated and rigorous.

Learning Objectives

Teaching the concept of intercoder reliability is sometimes difficult for the instructor of a research methods class and often the concept is frightening for the graduate student that has had little quantitative instruction. The importance of intercoder reliability is of central concern when the method of content analysis is taught, as the confidence in any conclusions about systematic analysis of media content are entirely dependent on the quality of the data and coding. Reliability evaluation therefore exists to protect against the possibility of contamination in scientific data regarding media content (Krippendorf, 1980). We provide a brief outline of the fundamental concepts of content analytic methodology, followed by a few strategies to help the

---

Patric R. Spence (M.A., Michigan State University, 2002) is a doctoral student in the Department of Communication at Wayne State University. Ken Lachlan (Ph.D., Michigan State University, 2003) is an Assistant Professor in the Communication Department at Boston College. Address all correspondence to the first author at Department of Communication, Wayne State University, 508 Manoogian Hall, Detroit, MI 48202. Email: spencepa@wayne.edu
instructor teach the concept of intercoder reliability, primarily to students that are apprehensive concerning statistical methods. Although these strategies can be used in a content analysis class, they are best for an introductory methods class that will only spend one or two periods on the topic.

Content analysis basics.

Content analysis is popular in the behavioral sciences because of its ability to be utilized for both written and oral communication as well as its ability to compare data across time and context. The method allows the researcher to identify particular words, phrases or concepts within the text(s) being examined. Though many definitions of content analysis have been offered over the years (Berelson, 1952; Weber, 1990; Berger, 1991), a complete and concise contemporary definition is offered by Neusendorf (2002), who defines it as “summarizing, quantitative analysis of messages that relies on the scientific method (including attention to objectivity-intersubjectivity, a prior design, reliability, validity, generalizability, replicability, and hypothesis testing) and is not limited to the types of variables that may be measured or the context in which the messages are created or presented” (p. 10).

Intercoder reliability is “the extent to which independent coders evaluate a characteristic of a message or artifact and reach the same conclusion” (Lombard et al., 2002, p. 589). This provides a validation to the coding scheme. Thus, intercoder reliability demonstrates that more than one person can use the coding scheme and obtain similar results. The validity of the data and any subsequent interpretations are suspect if intercoder reliability is not established or reported. Although this concept may be difficult it is important that the student understand intercoder reliability in order to produce as accurate a research study as possible.

Students become interested in the subject of content analysis or may be introduced to the topic for different reasons. Some desire to learn the method for use in future research, for others it’s a requirement. There also exists a body of students who wish to learn the method due to a belief that it is not statistical, not mathematically rigorous, or because they believe it’s easy. This means that students possessing a fear of statistics may be enrolled in the class or demonstrating interest in the method. This then gives the instructor an opportunity to gradually introduce the student to statistics and dispel common myths such as only people who have special aptitude in math can understand statistics, or that the student is “just one of those people who can’t understand statistics.” Further, students that are apprehensive concerning statistics will probably be reluctant to ask questions therefore, conducting activities as a class is an appropriate place to begin addressing the subject. However, some courses skip over the subject of intercoder reliability completely when covering content analysis, allowing the belief that the method is easy or non-statistical to continue. Although not all students enter the research methods course eagerly, any apprehension the students have can be eliminated through first emphasizing the thought processes that go into the particular method and then addressing the issue of the involved statistics (Clark, 1990).

Preparation for the Project

Primarily, the preparation for this activity involves the development and understanding of at least some type of basic coding scheme. For the example that we offer a simple counting scheme is given addressing the number of children in television commercials. Obviously, this may be adapted for other content coding schemes, but the essential element is that a scheme has been developed before beginning the activity. Also note that the technique discussed below, using count variables, leads to the discussion of intercoder reliability in continuous terms. From a data analytic standpoint, this leads to discussion of Cohen’s Kappa and other measures of continuous agreement. Clearly, the activity could be adjusted to allow for calculation of Scott’s Pi or other measures of agreement along categorical judgments.
While the counting of children in a series of advertisements may seem simple on face value, part of the preparation for this exercise involves teasing out the operational definitions of what exactly constitutes a child and how they will be unitized in the analysis. Further, it requires the development of coding sheets for these specific analyses and the preparation of and familiarity with the formula for the reliability statistic to be computed, as well the means of interpretation for these statistics.

From an operational definition standpoint, the goal here may be to initially come up with a definition of what constitutes a child (one unit of analysis). A definition should be drafted and distributed to students, indicating the precise definition of child offered in the current coding scheme. For example, a child could be operationally defined as “any noticeable, active, human figure that is, in the judgment of the coder, to be of less than 13 years of age.” This serves not only to assist the students in coding, but as a point that may serve to illustrate the importance of developing a priori coding categories. From a procedural standpoint, this may strike home the point that even something as simple as counting the number of children in an advertisement must be carefully explicated before beginning any type of systematic analysis of reliability between coders.

Further, some discussion should be given not only to units of analysis in terms of recording units, but in terms of sampling units. Also critical to the discussion of content analysis below is an understanding and preparation of the framework of content that will be scrutinized as a sampling unit (see Krippendorf, 1980; Pool, 1959). If we are to conduct an analysis of the number of children in a series of commercials, then what is our sampling frame? For example, a series of seven commercials could be viewed and considered one sampling unit. For purposes of illustration, this sampling frame could also be broken up later for reanalysis as seven independent units of analysis, as opposed to one large one. This is also helpful in demonstrating the ways in which intercoder reliability may be effected by decision made during the conceptual stages of the study when decision are being made regarding unitizing.

In any case, as part of the preparation for the activity, a clear set of operational definitions must be developed, along with sampling scheme criteria pertaining to unitizing and sampling units. Preparation must also include familiarity with the reliability formulas themselves, along with certain classroom items that are essential for conducting the exercise. These include handouts with the sampling and operational definitions, some sort of organized coding sheet onto which judgments can be made, the availability of some sort of visual classroom media (such as a projector for discussion of the reliability formula) and of course a videotape or DVD containing the media that will be subject to analysis.

The Project

In beginning when a discussion of intercoder reliability has begun, the instructor should review what has already been covered in the course concerning what content analysis is. Take the students back to a basic concept of comparison. Begin an activity where the class is counting children (ages 13 and under) in television commercials. Two students will be coders A and B, (the remainder of the class should be coding on their own). After this is complete the instructor will walk the class through the (percent) agreement between coders A and B. For this exercise it may be best to have the number of commercials be 10 (or something else easily divisible) thus making the calculations easy and non-threatening. The instructor should then have a code sheet on the blackboard, overhead projector or PowerPoint from which he/she will demonstrate the agreement between the coders. It is possible that perfect agreement will occur due to the small number of units, this is acceptable. In this case explain to the students that a perfect agreement has occurred. After this, choose two new coders and do a similar class exercise. Again, if perfect agreement is achieved it is not a problem, just take the class through the activity. After this is done a few times go back to the first activity (children in television commercials) and change one of the codes
making a disagreement occur (or a larger disagreement than what was achieved). Ask the students for the new reliability score. Do this a few more times between the coders to making sure to reduce and increase the reliability score. At this point the instructor should move into a discussion concerning how percent agreement is a statistic.

This is a chance to introduce the students to a simple formula and some equations. Introduce the conceptual formula, The formula provided by Neuendorf (2002) may be appropriate due to is less than intimidating appearance:

\( PAo \) represents the proportion agreement, observed, were \( A \) is the number of agreements between the two coders and \( n \) is the total number if units that the two coders have coded (Neuendorf, 2002). The statistic has a range from .00 (no agreement) to 1.00 (perfect agreement). Then take the formula and relate it back to the previous activities that were completed in class. Thus using the first activity. \( A \) then is the number of times coder A and B agreed that children were present in the television commercials (in this case imagine 8) and \( n \) is the total number of commercials watched (10 commercials were watched). This provides a percent agreement of .80. This can be repeated a few times and then accompanied with an explanation of the range of the statistic (it may be a bit too early to begin a discussion about what are considered good levels of agreement). A brief explanation can begin at this point by the instructor explaining that this is a popular coefficient and further published examples can be brought into the explanation at this point.

If the instructor believes that the class is ready he/she then should begin a discussion of what chance is and the role of chance in agreement between coders. Ask a few of the students to explain what chance is, when chance occurred in their life and to provide examples. Although the role of chance is more associated with probability than statistics it is still one of the best means to introduce students that have a phobia, or believe myths of math/statistics, that the process is harmless. At the same time a foundation is being built to teach the apprehensive students about inferential statistics at a later point due to the fact that one will be making inferences regarding that sample from which the objects being coded were drawn from. An often used example when first discussing chance is that of a jar of marbles (Sawilowski & Fahome, 2003). One has a jar of marbles with a label indicating that there are 1,000 marbles in the jar. The make up of the marbles are 50% green, 25% blue, 15% orange and 10% white. The jar was shaken and then a sample of 100 marbles was taken from the jar. The expectation (or it may be better to frame it as a hypothesis or research question due to the fact that the concept should be familiar to the class) half of the sample (50%) would be green. However, 60 green marbles were obtained. Now the instructor should ask if this outcome is likely to occur if the label was accurate?

Next explain that it is more likely that a random sample of 100 marbles is more likely to produce 50 green marbles than any other sample. What about if 49 or 51 green marbles were obtained in the sample, is that an unlikely occurrence? Should the hypothesis be rejected? Is the label incorrect? Or was this an issue of chance (if possible this could also be used to discuss a very crewed version of sampling error)? Just as any number of green marbles can be obtained by chance, agreement between coders can manifest by change.

In the coding process the students at this point must understand that chance agreement can occur between coders and this can inflate the reliability score. If the coders were determining whether or not participants in the first commercial example were male or female, chance alone dictates a 50% agreement. From this point the instructor should talk about the role of chance with the previous example of percent agreement. Several agreements between the two coders could exist on the basis of chance. Ask the students if it was possible that two coders perceived that children were present in the commercials by chance, and if so how? Such chance agreements could include different beliefs on what the operational definition is, coders seeing two different
participants in the commercial and identifying them as children, or coder error. With such errors pointed out, a discussion on why percent agreement can’t account for chance should begin.

Return to the previously used code sheet on the blackboard, overhead projector or PowerPoint and again demonstrate the agreement as was done with the percent procedure. Then present the equation for Cohen's kappa:

(2) again taken from Neuendorf, PAo concerns the proportion agreement, observed, and refers to the proportion agreement that is expected by chance. Now walk through the coding procedure and work out the lesson in the same procedures that was done with percent agreement. Remind the students that just as percent agreement is a statistic, the kappa is also a statistic. The objective of using the kappa, to access reliability, is the same as using percent agreement but in a manner to ensure more confidence in the findings. Cohen's kappa (1968) is a good statistic to teach due to it being the most popular reliability assessment used (Zwick, 1988), particularly because of its accessibility in SPSS. The Kappa has a range is from .00 (agreement at chance level) to 1.00 (perfect agreement). Do several more computations of the statistic with the students. Changing the coded responses each time and working to produce the new reliability score. When the students have an understand of how this Kappa is obtained then the instructor can begin explaining how to evaluate the score.

Debriefing

Any classroom topic that involves subjects which create high levels of apprehension must have measures in place to provide feedback to the instructor concerning the students’ grasp of the subject. Although the in-class exercise outlined above is one such measure the instructor should utilize other measures of classroom assessment. For the topic of intercoder reliability three specific classroom assessment techniques (Angelo & Cross, 1993) stand out as being appropriate. The first is the end of class quiz. This does not have to be graded. The quiz would consist of two columns of coder responses and the students would be required to calculate the reliability scores with both percent agreement and a kappa.

The second technique is the minute paper. At the end of the class the instructor should ask students to answer on a half-sheet of paper the following: What is the most important point you learned today?; and, What point remains least clear to you? Or What question do you have that remains unanswered. The purpose of this exercise is to obtain data about students’ comprehension of the particular class session. This exercise is also helpful due to the students’ reluctance to ask questions about statistical topics. The instructor should review the students’ responses and note any useful comments. In the following class periods the instructor should emphasize and then discuss the issues that remain problem some. A third appropriate technique is the memory matrix, in this exercise the students are instructed to fill in the cells of a diagram for which instructor has provided labels. For example, some cells may have the statistical formula and students are to enter the label or identify whether certain types of codes would be continuous or categorical. Several applications of the memory matrix can be used.

Conclusion

Regardless of the number of class periods used to address the method of content analysis, the goal should be to introduce the students to the method in as much depth as possible. This requires a strong discussion and application of the role of intercoder reliability. Emphasizing the importance of reliable data, while keeping the explanation simple and concise, will allow students who may be apprehensive of quantitative methods to become more comfortable with these analyses. There are two goals of this teaching tip, first the authors hope to encourage methods instructors (specifically at the graduate level but also those who teach undergraduate methods) to
include a complete discussion of intercoder reliability when teaching content analysis. Second, to provide the instructor with tools and procedures to aid in teaching the topic.

**Resources**


**References**


