



Standardization in Assessment and Reporting of Intercoder Reliability in Content Analyses

Matthew Lombard, Temple University

December 5, 2008

University of Michigan



Overview

- History of interest in topic
- Content analysis of content analyses
- Challenges in improving reliability
- Reliability in open ended coding
- A broader definition of reliability



History of Interest in Topic

- (Too) ambitious c.a. project: Structural Features of Television
- Variables at all levels of measurement, many ratio level
- 7 coders (programs); 4 coders (timepoints, 10 sec intervals)
- Old VHS linear editing machines
- Want to do reliability right, but...



History of Interest in Topic

- Many indices, few of them flexible
- Few guidelines (re indices, procedures, reporting)
- Krippendorff's Alpha seemed best but complex by hand, no software to calculate
- Problem variables w/o variation



Content Analysis of Content Analyses

- Literature review regarding topic
- RQ: How well do researchers in communication deal with reliability?



Content Analysis of Content Analyses

■ Literature review:

- Everyone says it's important
- Previous studies show problems in use
- Many indices (39+): Cohen's Kappa, Holsti's, Krippendorff's Alpha, percent agreement, Scott's Pi, Perreault-Leigh's Ir, Tinsley-Weiss's T, Bennett-Alpert-Goldstein's S, Lin's concordance coefficient, Hughes-Garrett's Generalizability Theory approach, Rust-Cooil's PRL approach, ...



Content Analysis of Content Analyses

- Literature review:

- Experts only agree about

- not using percent agreement, Cronbach's alpha, r, chi-square

- using indices that consider chance agreement

- Software not widely available, all 'beta,' require special/separate data setups



Content Analysis of Content Analyses

- Content analyzed journal articles with keyword phrase 'content analysis' in Communication Abstracts from 1994 to 1998



Content Analysis of Content Analyses

- 200 articles in sample
- Two coders
- Several reliability methods, to show differences
- 80% all quantitative c.a., 8% qualitative, 12% both



Content Analysis of Content Analyses

- Content that was analyzed in studies:
 - 42% newspapers
 - 30% television
 - 18% magazines
 - 3% respondent reports
 - 2% film
 - 2% radio
 - 16% other



Content Analysis of Content Analyses

- Only 69% discussed intercoder reliability

OF THAT 69% (n=137)...

- Only 41% reported reliability for specific variables
- 42% didn't report method/index used
- 9% didn't report number of reliability coders



Content Analysis of Content Analyses

- Mean # of sentences on reliability in text and footnotes was 4.5 (SD=4)
- Only 6% included reliability in a table
- Only 45% cited an article or book related to reliability



Content Analysis of Content Analyses

- 15% Holsti's
- 10% Scott's Pi
- 7% Cohen's Kappa
- 3% Krippendorff's Alpha
- 25% Percent agreement (16% "intercoder reliability," 7% "percentage agreement," 2% "simple agreement")
- 36% other



Content Analysis of Content Analyses

- 23% didn't report size of reliability sample
- 35% didn't report whether reliability sample was part of overall sample
- 91% didn't report amount of coder training
- 71% didn't report how/if discrepancies among coders were resolved



Content Analysis of Content Analyses

- Conclusion: Intercoder reliability is inconsistently assessed and reported in communication research reports



Content Analysis of Content Analyses

- Example: *“All 168 articles were coded by four trained researchers. The Holsti coefficient of inter-coder reliability was .93.”*



Content Analysis of Content Analyses

- Example: *“Correlations were computed between the second author’s ratings and the two undergraduates’ ratings for all the variables (mean $r=.88$ and mean $r=.82$, respectively). Those desiring a complete summary of the reliability coefficients should contact the authors.”*



Challenges in improving reliability

- Publicize and argue for its importance, especially in Communication
- Don't make 'perfect' enemy of 'good'
- Propose guidelines, describe available tools, and publicize
 - HCR article
 - Online supplement:
<http://www.temple.edu/mmc/reliability>
 - This talk (even if preaching to converted)



Challenges in improving reliability

Abridged guidelines:

- *Calculate and report intercoder reliability!*
- *Report:*
 - *Size of, method used to create the reliability sample, along with justification of that method*
 - *Relationship of reliability sample to full sample (i.e., same, subset, or separate)*



Challenges in improving reliability

- *Number of reliability coders, whether they include the researcher(s)*
- *Amount of coding conducted by each reliability, and non-reliability, coder*
- *Index or indices selected to calculate reliability, and justification of this/these*
- *Intercoder reliability level for each variable, for each index*



Challenges in improving reliability

- *Approximate amount of training (in hours) required to reach reliability levels reported*
- *How disagreements in reliability coding were resolved in full sample*
- *Where and how the reader can obtain detailed information regarding the coding instrument, procedures, and instructions (e.g., from the authors)*



Challenges in improving reliability

- Advocate, push, cajole, encourage, etc.:
 - More expert agreement on guidelines (e.g., re: Cohen's Kappa)
 - Development of better software
 - Incorporated into standard packages



Reliability in open ended coding

- Content analysis = quantitative (Berelson, et al.)
- But... *qualitative content analysis, textual analysis, semiotic analysis, case study, ...*
- Simplistic, impractical bias: if qualitative variable, don't lose richness by coding to quantitative; if quantitative, use closed-ended item(s)



Reliability in open ended coding

- When data are open ended, verbatim, essay, etc.
 - Human coding – reliability is critical
 - designing coding scheme (esp. based on data)
 - coding according to scheme
 - Automated (computer) coding – reliability still an issue



A broader definition of reliability

- Computer is reliable, but is the coding scheme?
- For common open ended items, aside from internal validity, we need to increase consistency of coding schemes across researchers, studies to be able to compare results



A broader definition of reliability

■ How?

- Collaboration among researchers in topic areas
- Wikis, listservs, shared databases
- Advocate use of theoretical rationales
- Advocate requiring publication of details of schemes and what worked and didn't
- Community building for shared obligation

■ Easier said than done but worth doing



More information...

- Matthew Lombard

`lombard@temple.edu`

`http://matthewlombard.com`

