Human Coding & Interrater Reliability In Content Analysis

Jonathan Bundy
University of Georgia

2012 SMA Annual Meeting

The University of Georgia
Terry College of Business

Human Coding

We have computers, why on Earth use human coding?

1. When the meaning of the variable resides within source/receiver interpretation
   - Attributions (Lee et al., 2004)
2. When context matters
   - Socio-cognitive variables (King et al., 2011)
3. When concepts are not easily identified by particular words or phrases
   - Strategic actions (Zavyalova et al., 2012; Lamin & Zaheer, 2012)
4. To identify grounded or emergent variables

Measurement Issues

M= t + e

(measure = true score + error)

- Sources of error = Humans
  - Coder misinterpretation, poor scheme, inadequate training, inattention/fatigue, recording error, rogue coder!
- Thus, we need ***reliability*** – the extent to which a measuring procedure yields the same results on repeated trials
  - More specifically, ***interrater reliability*** – the amount of agreement or correspondence among two or more coders

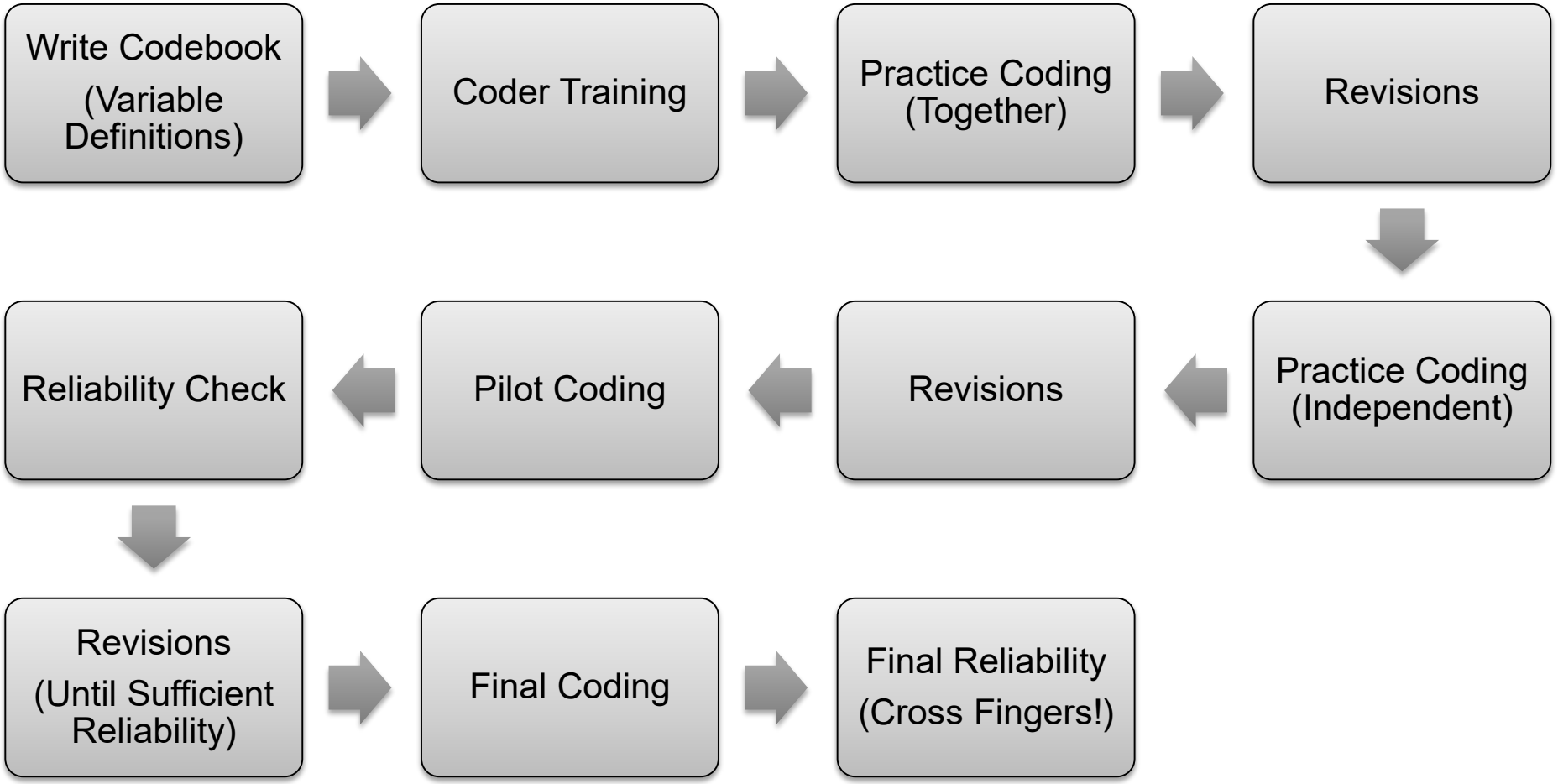
Why Reliability?

- Validity of coding scheme
  - Results are not the idiosyncratic to rater subjectivity
- Allows for the use of multiple coders
- Replication over time

Reviewers are going to ask for it!

Reliability Flowchart

(Neuendorf, 2002; Weber, 1990)



Reliability Measures

http://dfreelon.org/utils/recalfront/

Measure	Type	Best for	More than 2 coders?
Percent Agreement	Agreement	Nominal	No
Holsti's Method	Agreement	Nominal	No
Scott's Pi	Agreement (w/ chance)	Nominal	No
Cohen's Kappa	Agreement (w/ chance)	Nominal	Yes
Krippendorff's Alpha	***Agreement (w/ chance)***	***Any***	***Yes***
Spearman Rho	Covariation	Ordinal	No
Pearson Correlation	Covariation	Interval/ratio	No
Lin's Concordance	Covariation	Interval/ratio	No

Other Thoughts

- Codebook and form - make the set so complete and unambiguous as to eliminate individual coder differences
- At least **2 coders**, 10% overlap ranging between **50-300 observations** depending on sample size
- Reliability can be low when coding subjective into objective, thus cut-offs can be lower (.67-.80)
- **Blind coding is preferable**
- Consensus needed when training/building, but not in final coding
- What to do with variables that are **not reliable**?
  - Redefine variable, split variable, re-train coders , drop variable, drop coder, integrate non-content analytic data
- Need **separate reliability for each measure**
- With CATA, reliability always = 1, but validity still an issue

Tips

- Habits & Routines
  - Code daily, but avoid fatigue
- Spend time up front
  - Familiarize self with content texts and theory
  - Invest in training!

Revise early and revise often!

References

- Practical guide – Neuendorf, K. A. 2002. *The content analysis guidebook*. Sage.
- More robust guide – Krippendorff, K. 2012. *Content analysis: An introduction to its methodology.* Sage.
- A classic (greenie) – Weber, R. P. 1990. *Basic content analysis*. Sage.
- From MGMT – Duriau, V. J., Reger, R. K., & Pfarrer, M. D. 2007. A content analysis of the content analysis literature in organization studies: Research themes, data sources, and methodological refinements. *Organizational Research Methods*, 10: 5-34.
- http://www.terry.uga.edu/management/contentanalysis/