Using MTurk to Improve Content Analyses

Content Analysis PDW
Tim Hubbard, Ph.D.
You just collected thousands of media articles about a sample of firms…

How do you feel confident they are actually about the firm?
Why Should We Care?

- Stronger connection between construct and proxy
- Reduces measurement error
- Increases the likelihood of finding results
- Thoroughness of independent ratings are more valid and easy to report
Our Situation

- Study of celebrity startup firms
- Reviewer requested we collect more media data
- Needed to analyze general media’s tenor, volume, and non-conformity of sample firms
- Needed to do it thoroughly—and fast
Amazon’s MTurk Helped

- Human coding
- Gives access to multiple individual raters per article
- Very fast
- Imports data using CSV files
- Easy payment
General Process

1. Media collection from LexisNexis
2. MTurk Ratings and Evaluation
3. Linguistic Analysis (LIWC 2015)
4. Statistical Analysis (Stata)

Custom code (Text files to CSV) → CSV Out & In → CSV Import
Parameters

- 6,260 articles from LexisNexis
- Presented only the first page of the articles
- 2 raters per article (12,520 ratings)
  - Disagreement or unclear, added one more (795)
- 7 hours per run
- $0.08/rating
### Instructions (Click to expand)

**Is ${company} the primary focus of this article?**

<table>
<thead>
<tr>
<th>Value</th>
<th>Guidance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Select this if the article primarily about the company ${company}.</td>
</tr>
<tr>
<td>Unclear</td>
<td>Select this if it is unclear to you whether the article is primarily about ${company}.</td>
</tr>
<tr>
<td>No</td>
<td>Select this if the article is not primarily about the ${company}. For example, it might be an article about Microsoft that mentions ${company}.</td>
</tr>
</tbody>
</table>

**Article Title: ${title}**

**Article Body: ${body}**

**Company: ${company}**

**Reference: ${identifier}**

**Company Ticker: ${ticker}**
Moving from LexisNexis to MTurk

- Custom developed software to move from LexisNexis text files to a fully populated CSV file
- That file can then go up to MTurk
- Developed by Rampant Strategy OÜ
LIWC 2015

- Can import CSV files now
- Straight from MTurk
- Can export to CSV which is easy to import into Stata
General Process

1. Media collection from LexisNexis
2. MTurk Ratings and Evaluation
3. Linguistic Analysis (LIWC 2015)
4. Statistical Analysis (Stata)

Custom code (Text files to CSV)
CSV Out & In
CSV Import
Questions & Discussion

thubbard@nd.edu