Computer-Aided Text Analysis, or: How I Learned to Stop Worrying and Love the Computer

Aaron F. McKenny

UNIVERSITY OF CENTRAL FLORIDA

Why I love the computer

S-1 1 ds1.htm FORM S-1

Table of Contents

As filed with the Securities and Exchange Commission on April 29, 2004
Registration No. 333-

SECURITIES AND EXCHANGE COMMISSION
Washington, D.C. 20549

FORM S-1
REGISTRATION STATEMENT
Under
The Securities Act of 1933

GOOGLE INC.
(Exact name of Registrant as specified in its charter)

Delaware	7375	77-0493581
(State or other jurisdiction of incorporation or organization)	(Primary Standard Industrial Classification Code Number)	(I.R.S. Employer Identification Number)

1600 Amphitheatre Parkway
Mountain View, CA 94043
(650) 623-4000
(Address, including zip code, and telephone number, including area code, of Registrant's principal executive offices)

Eric Schmidt
Chief
1600 Am
Mount
(
(Name, address, including zip code, and telephone number, including area code, of agent for service)

267 pages

	Copies to:	
Larry W. Sonsini, Esq. David J. Segre, Esq. Wilson Sonsini Goodrich & Rosati, P.C. 650 Page Mill Road Palo Alto, California 94304-1050 (650) 493-9300	David C. Drummond, Esq. Jeffery L. Donovan, Esq. Anna Itoi, Esq. Google Inc. 1600 Amphitheatre Parkway Mountain View, CA 94043 (650) 623-4000	William H. Hinman, Jr., Esq. Simpson Thacher & Bartlett LLP 3330 Hillview Avenue Palo Alto, California 94304 (650) 251-5000

Approximate date of commencement of proposed sale to the public: As soon as practicable after the effective date of this Registration Statement.

If any of the securities being registered on this Form are being offered on a delayed or continuous basis pursuant to Rule 415 under the Securities Act of 1933, as amended (the "Securities Act"), check the following box. ☐

If this Form is filed to register additional securities for an offering pursuant to Rule 462(b) under the Securities Act, please check the following box and list the Securities Act registration number of the earlier effective registration statement for the same offering. ☐

If this Form is a post-effective amendment filed pursuant to Rule 462(c) under the Securities Act, check the following box and list the Securities Act registration number of the earlier effective registration statement for the same offering. ☐

If this Form is a post-effective amendment filed pursuant to Rule 462(d) under the Securities Act, check the following box and list the Securities Act registration statement number of the earlier effective registration statement for the same offering. ☐

If delivery of the prospectus is expected to be made pursuant to Rule 434, check the following box. ☐

CALCULATION OF REGISTRATION FEE

Why I love the computer

Table 2
N for Small, Medium, and Large ES at Power = .80 for α = .01, .05, and .10

	α								
	.01			.05			.10		
Test	Sm	Med	Lg	Sm	Med	Lg	Sm	Med	Lg
1. Mean dif	586	95	38	393	64	26	310	50	20
2. Sig *r*	1,163	125	41	783	85	28	617	68	22
3. *r* dif	2,339	263	96	1,573	177	66	1,240	140	52
4. *P* = .5	1,165	127	44	783	85	30	616	67	23
5. *P* dif	584	93	36	392	63	25	309	49	19
6. χ^2									
1*df*	1,168	130	38	785	87	26	618	69	25
2*df*	1,388	154	56	964	107	39	771	86	31
3*df*	1,546	172	62	1,090	121	44	880	98	35
4*df*	1,675	186	67	1,194	133	48	968	108	39
5*df*	[illegible]	[illegible]	[illegible]	1,293	143	51	1,045	116	42
6*df*	[illegible]	[illegible]	[illegible]	1,36[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
7. ANOVA									
2g[a]	586	95	38	393	64	26	310	50	20
3g[a]	464	76	30	322	52	21	258	41	17
4g[a]	388	63	25	274	45	18	221	36	15
5g[a]	336	55	22	240	39	16	193	32	13
6g[a]	299	49	20	215	35	14	174	28	12
7g[a]	271	44	18	195	32	13	159	26	11
8. Mult *R*									
2k[b]	698	97	45	481	67	30			
3k[b]	780	108	50	547	76	34			
4k[b]	841	118	55	599	84	38			
5k[b]	901	126	59	645	91	42			
6k[b]	953	134	63	686	97	45			
7k[b]	998	141	66	726	102	48			
8k[b]	1,039	147	69	757	107	50			

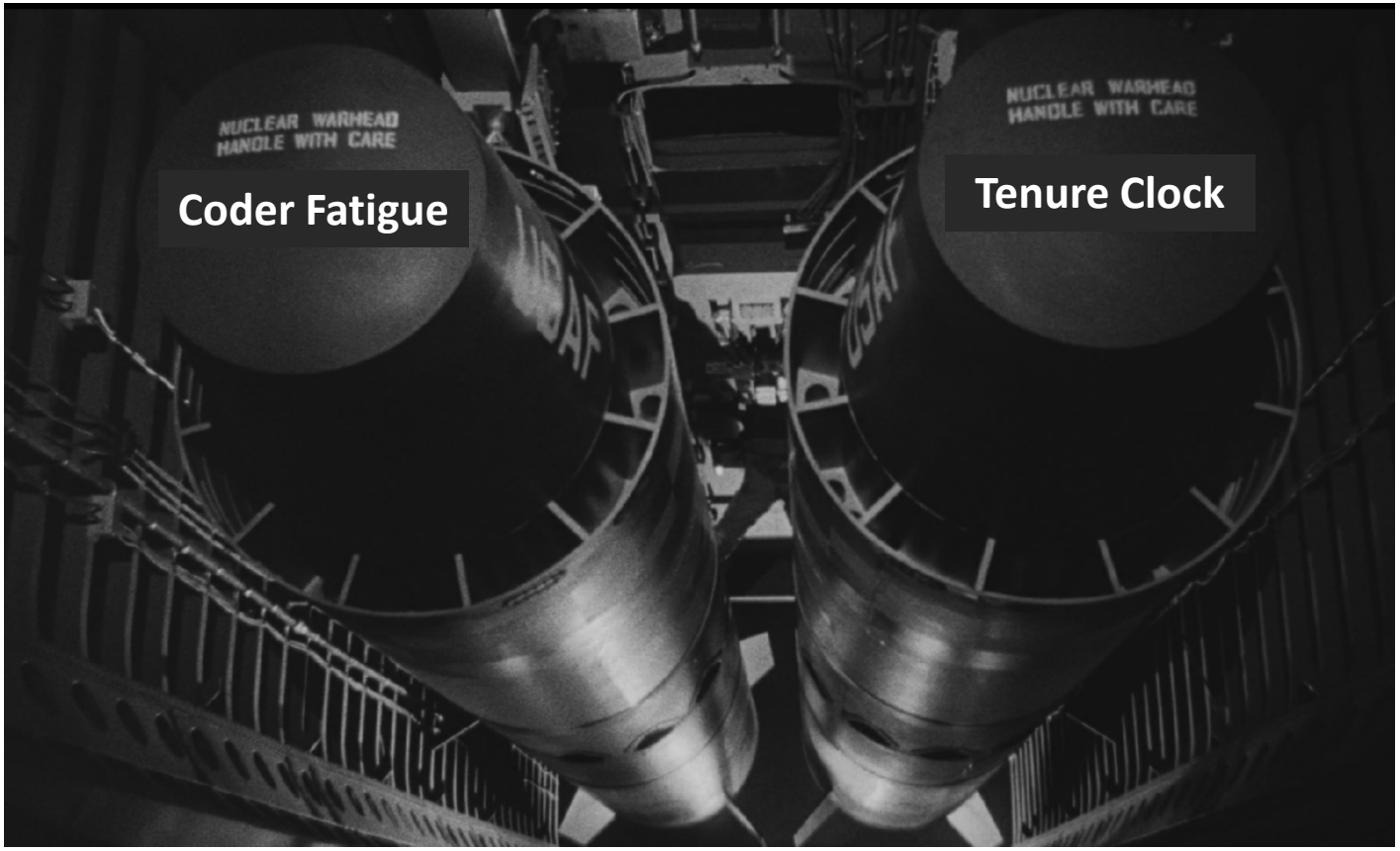
Note. ES = population effect size, Sm = small, Med = medium, Lg = large, diff = difference, ANOVA = analysis of variance. Tests numbered as in Table 1.
[a] Number of groups. [b] Number of independent variables.

n = 91-126 documents

(Table from Cohen, 1992)

Manual Coding…

25,000-35,000 pages!



Computer-Aided Text Analysis

Organizational Ambidexterity

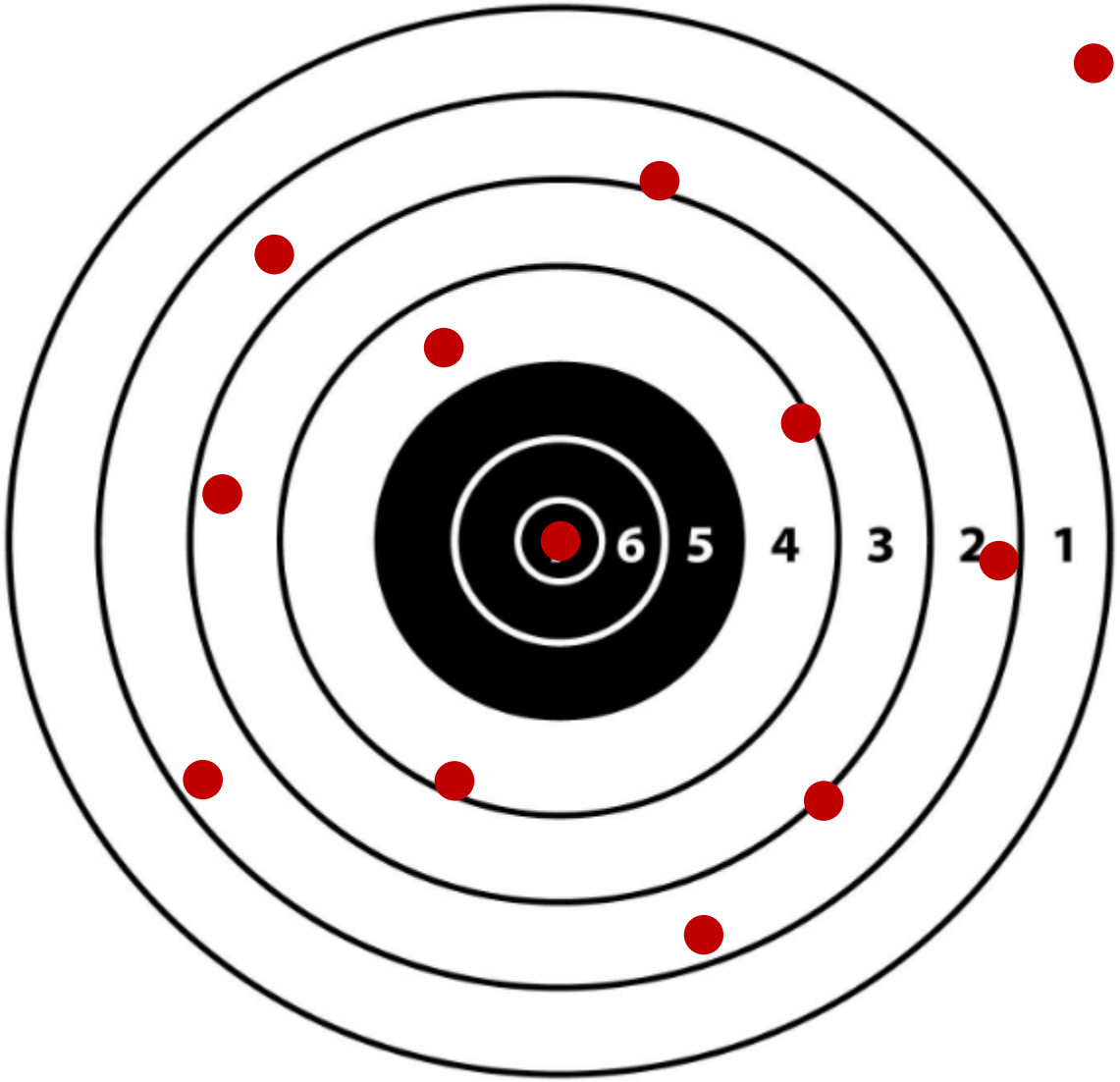
How it works

- Sample Dictionary: *Innovativeness*
  - "Innovative" "Innovation" "Innovate" "Research" "Inventions" "Inventive" "Creative" "Creativity"

- Sample narrative to analyze:
  - "The creativity of our research and development team make this organization one of the most innovative in the industry, with patents on over 2,300 inventions."

Innovativeness: 4

What worried me



Origins of my recommendations

Organizational Research Methods
Volume 13 Number 2
April 2010 320-347

10.1177/1094428109333949
http://orm.sagepub.com
hosted at
http://online.sagepub.com

Construct Validation Using Computer-Aided Text Analysis (CATA)

An Illustration Using Entrepreneurial Orientation

Jeremy C. Short
Texas Tech University
J. Christian Broberg
Wichita State University
Claudia C. Cogliser
Keith H. Brigham
Texas Tech University

Construct validity continues to pose challenges in the organizational sciences. To capture difficult-to-measure constructs of interest, researchers have often relied on content analysis. One content analysis technique, computer-aided text analysis (CATA), is particularly attractive because of the ability to process large samples with high speeds and reliabilities. Unfortunately, inconsistent guidance exists to guide researchers through the use of this tool in a manner compatible with accepted methods used to validate constructs in a rigorous manner. The authors review research using content analysis to examine the extent to which such studies integrate methods for assessing content, external, discriminant, and predictive validity. To provide direction for organizational researchers interested in using CATA to measure theoretically based constructs relevant to the management field, they suggest a number of possible procedures to enhance construct validity. They illustrate these procedures using the construct of entrepreneurial orientation.

Keywords: *computer-aided text analysis; construct validation; content analysis; DICTION; entrepreneurial orientation*

Validity refers to evaluating inferences drawn from measures of concepts of interest and considers the extent to which a measure accurately represents that focal concept (Cronbach, 1971). The importance of rigor in construct measurement cannot be minimized. Kerlinger and Lee (2000) argued that the ability to demonstrate appropriate measures of theoretical notions of interest through construct validity "is one of the most significant advances of modern measurement theory and practice" (p. 670).

Despite advancements in the organizational sciences vis-à-vis construct validation, reviews of construct validity in the management field have yielded generally bleak conclusions about

Authors' Note: The authors would like to thank Tyge Payne for his thoughtful comments on previous drafts of this manuscript. Please address correspondence to Jeremy C. Short, Area of Management, Texas Tech University, Lubbock, TX 79408; e-mail: jeremy.short@ttu.edu.

Article

Organizational Research Methods
16(1) 152-184

DOI: 10.1177/1094428112459910
orm.sagepub.com
SAGE

Using Computer-Aided Text Analysis to Elevate Constructs: An Illustration Using Psychological Capital

Aaron F. McKenny[1], Jeremy C. Short[1], and G. Tyge Payne[2]

Abstract
Applying individual-level constructs to higher levels of analysis can be a fruitful practice in organizational research. Although this practice is beneficial in developing and testing theory, there are measurement and validation concerns that, if improperly addressed, may threaten the validity and utility of the research. This article illustrates how computer-aided text analysis might be utilized to facilitate construct elevation while ensuring proper validation. Specifically, we apply a framework to develop organizational-level operationalizations of individual-level constructs using the psychological capital construct as an example.

Keywords
content analysis, level of analysis, multilevel, psychological capital, construct measurement, computer-aided text analysis

Macro organizational research often "borrows" constructs and associated theories from micro disciplines to investigate how they might apply at higher levels of analysis (Whetten, Felin, & King, 2009). There seems to be good reason for this practice as several micro-level theories and constructs have helped explain phenomena occurring at higher levels (Staw, 1991). For example, the concept of "learning," once reserved for individuals (e.g., Gagné, 1965), has been expanded by organizational theorists to apply to teams (e.g., Brooks, 1994), organizations (e.g., Huber, 1991), networks (e.g., Knight, 2002), and even institutions (e.g., Siebenhüner & Suplie, 2005).

[1]Price College of Business, University of Oklahoma, Norman, OK, USA
[2]Rawls College of Business, Texas Tech University, Lubbock, TX, USA

Corresponding Author:
Jeremy C. Short, Division of Management & Entrepreneurship, Price College of Business, 307 W. Brooks, Rm. 206, University of Oklahoma, Norman, OK 73019, USA.
Email: jeremy.short@ou.edu

Journal of Management
Vol. XX No. X, Month XXXX 1–25
DOI: 10.1177/0149206316657594

Reprints and permissions:
sagepub.com/journalsPermissions.nav

What Doesn't Get Measured Does Exist: Improving the Accuracy of Computer-Aided Text Analysis

Aaron F. McKenny
University of Central Florida
Herman Aguinis
George Washington University
Jeremy C. Short
Aaron H. Anglin
University of Oklahoma

Computer-aided text analysis (CATA) is a form of content analysis that enables the measurement of constructs by processing text into quantitative data based on the frequency of words. CATA has been proposed as a useful measurement approach with the potential to lead to important theoretical advancements. Ironically, while CATA has been offered to overcome some of the known deficiencies in existing measurement approaches, we have lagged behind in regard to assessing the technique's measurement rigor. Our article addresses this knowledge gap and describes important implications for past as well as future research using CATA. First, we describe three sources of measurement error variance that are particularly relevant to studies using CATA: transient error, specific factor error, and algorithm error. Second, we describe and demonstrate

Acknowledgments: We thank Fred Oswald, two Journal of Management *anonymous reviewers, and Regan Stevenson for providing highly constructive and useful feedback that allowed us to improve our manuscript substantially. Also, we thank the following three author teams for sharing texts used in their studies: (a) Moss, T. W., Payne, G. T., & Moore, C. B. 2014. Strategic consistency of exploration and exploitation in family businesses.* Family Business Review, *27: 51-71; (b) Short, J. C., Broberg, J. C., Cogliser, C. C., & Brigham, K. H. 2010. Construct validation using computer-aided text analysis (CATA): An illustration using entrepreneurial orientation.* Organizational Research Methods, *13: 320-347; and (c) Zachary, M. A., McKenny, A. F., Short, J. C., & Payne, G. T. 2011. Family business and market orientation: Construct validation and comparative analysis.* Family Business Review, *24: 233-251.*

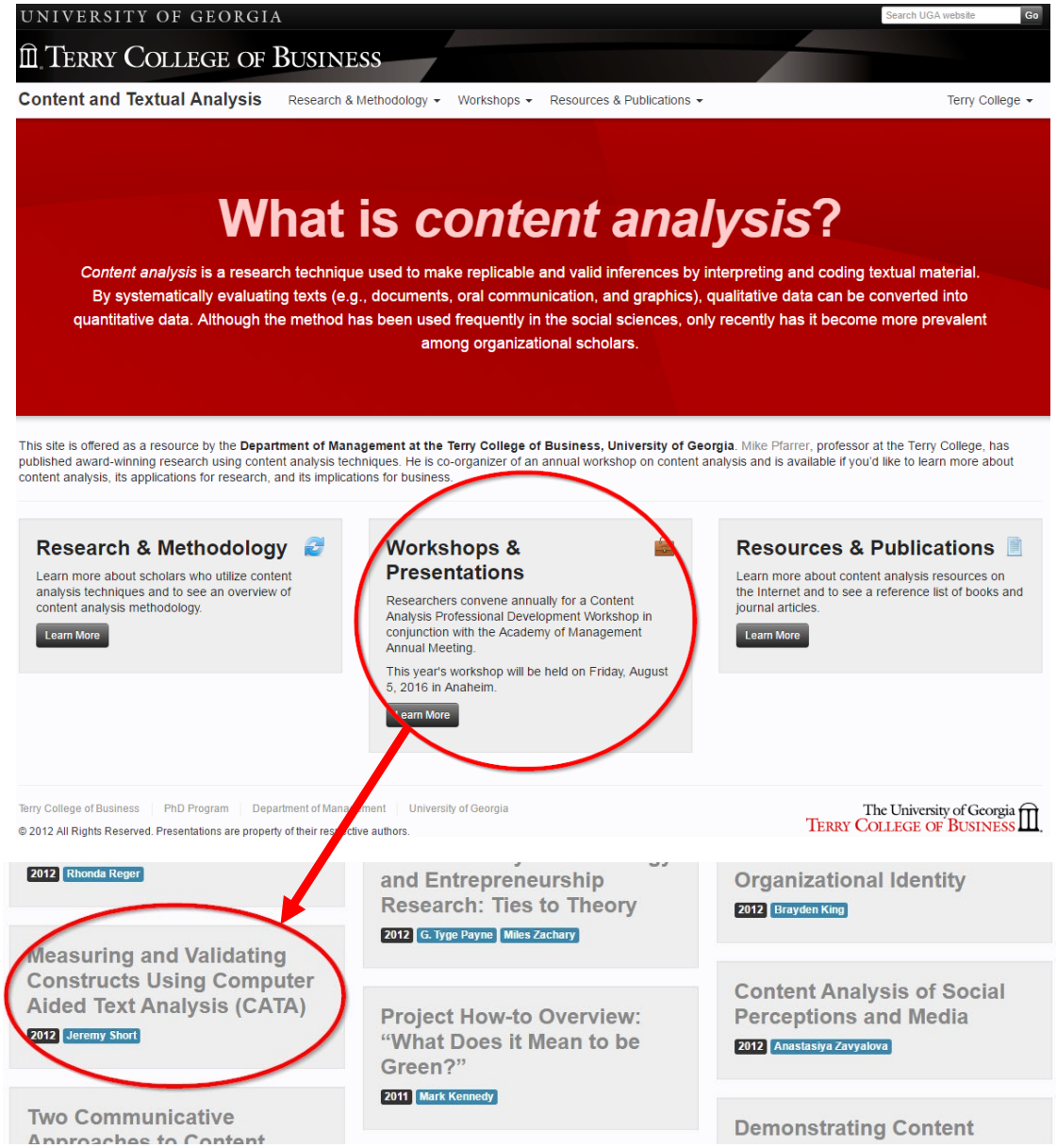
Supplemental material for this article is available at http://jom.sagepub.com/supplemental

Corresponding author: Aaron F. McKenny, Department of Management, College of Business Administration, University of Central Florida, P. O. Box 161400, Orlando, FL 32816-1400, USA.

E-mail: aaron.mckenny@ucf.edu

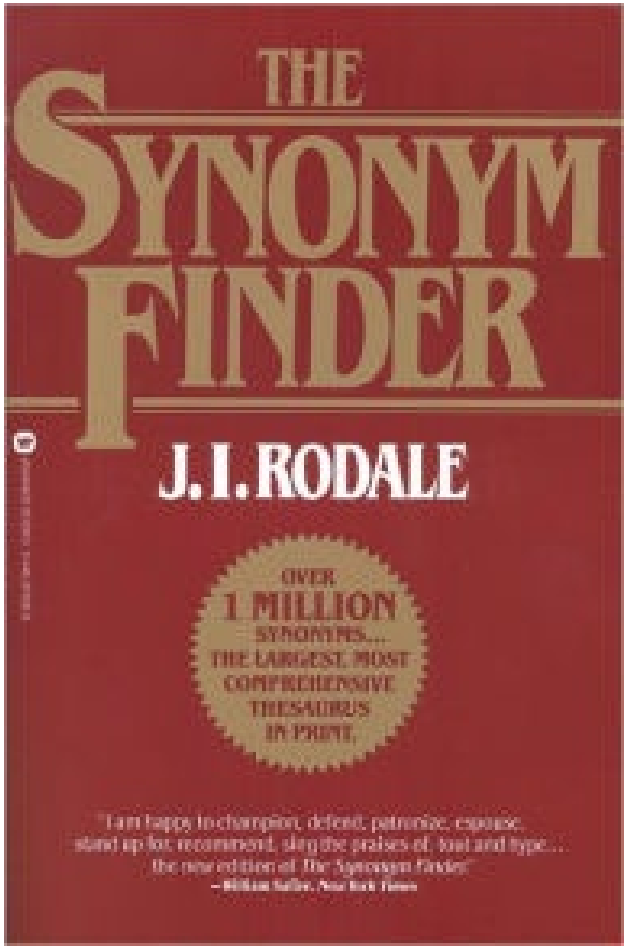
Validity

- Content
- External
- Discriminant
- Predictive



Two-step process

- Deductive word list
  - From theory
  - From existing word lists
  - From the synonym finder
- Inductive word list
  - From sampled texts



Measurement Error

- ~~Random Response~~

  Something can change while writing the text.

  e.g., get a journal rejection in the middle of writing an email

- Transient
- Specific Factor
- Algorithm

Measurement Error

- ~~Random Response~~
- Transient

  Something can change between writing of sequential texts.

  e.g., Stock market tanks

- Specific Factor
- Algorithm

Measurement Error

- ~~Random Response~~
- Transient
- Specific Factor

  Contents of the word list involves judgment: omission/selection of words

  e.g., Does "risk" always indicate risk taking?

- Algorithm

Measurement Error

- ~~Random Response~~
- Transient
- Specific Factor
- Algorithm

  Differences in CATA tool algorithms

  e.g., Word list includes phrases but your CATA tool cannot process phrases

Estimating Reliability

Measurement Error Source	Reliability Estimate	Calculation Guidelines
Transient Error	Test-Retest Reliability	Collect 2 sequential texts and use CATA to analyze them. Calculate the correlation.
Specific Factor Error	Parallel Forms Reliability	Manually code 10% of sample. Calculate the correlation.
Algorithm Error	Interrater Agreement	Run CATA using two tools. Calculate Krippendorff's alpha.

Is there error to worry about?

Error Source	Type of Reliability Estimate	Entrepreneurial Orientation Dimension	Reliability Estimate	Percent of Variance Due to Measurement Error[a]
Transient error	Test-retest	Autonomy	.32	68
		Competitive aggressiveness	.43	37
		Innovativeness	.52	48
		Proactiveness	.55	45
		Risk taking	.71	29
		Mean (Test-retest)	**.51**	**49**
Specific factor error	Parallel forms	Autonomy	.29	71
		Competitive aggressiveness	.51	49
		Innovativeness	.35	65
		Proactiveness	.72	28
		Risk taking	.30	70
		Mean (Parallel forms)	**.43**	**57**
Algorithm error	Krippendorff's alpha	Autonomy	.90	10
		Competitive aggressiveness	.89	11
		Innovativeness	.89	11
		Proactiveness	.88	12
		Risk taking	.90	10
		Mean (Krippendorff's alpha)	**.89**	**11**

Is that significant?

- Engelen, Neumann, & Schmidt (2016)
  - EO→Tobin's Q: r = .24
  - Without measurement error: r = .34

  - Regression coefficient with error: 0.59
  - Regression coefficient without error: 0.83
  - Difference: 0.24

For a Business Week 1000 firm with $10B in assets, the difference is an increase of $2.4B in market value

How you too can stop worrying: Transient Error

- Is the construct theoretically stable over time?
- Investigate the possibility of shocks
- Decrease lag between texts
- Text selection
  - Frequency
  - Content consistency (managerial attention)
  - Standardization
  - Author (did it change?)

How you too can stop worrying: Specific Factor Error

Refine the dictionaries

- Deductive/Inductive (Short *et al.*, 2010)
- Manual/CATA (McKenny *et al.*, in press)
- Avoid word roots (e.g., Refin*)
- Include all relevant conjugations
- Replace words with phrases (e.g., “risk” vs “risk taking”)
- Ensure that words are relevant to your context

How you too can stop worrying: Algorithm Error

- Identify CATA tool limitations
- Run CATA using a third tool
- Recreate CATA analysis manually

Thank you

"Based on the findings of this presentation, I hope your conclusion is that measurement error is not a deterrent for reasons which at this moment must be all too obvious."

~Dr. Strangelove (*almost*)

Resources

- CAT Scanner CATA software (Free)
  - http://www.catscanner.net/
- University of Georgia Content Analysis site
  - http://www.terry.uga.edu/management/contentanalysis
  - Particularly Relevant Presentations
    - Custom Dictionaries in Computer-Aided Text Analysis (McKenny, 2012)
    - Measuring and Validating Constructs Using Computer-Aided Text Analysis (Short, 2012)

Resources

- McKenny AF, Aguinis H, Short JC, Anglin AH. In Press. What doesn't get measured does exist: improving the accuracy of computer-aided text analysis. *Journal of Management*. doi: 10.1177/0149206316657594.
  - Supplemental Materials

- McKenny AF, Short JC, Payne GT. 2013. Using CATA to elevate constructs in organizational research: validating an organizational-level measure of psychological capital. *Organizational Research Methods* **16**(1): 152-184.

- Short JC, Broberg JC, Cogliser CC, Brigham KH. 2010. Construct validation using computer-aided text analysis (CATA): an illustration using entrepreneurial orientation. *Organizational Research Methods* **13**(2): 320-347