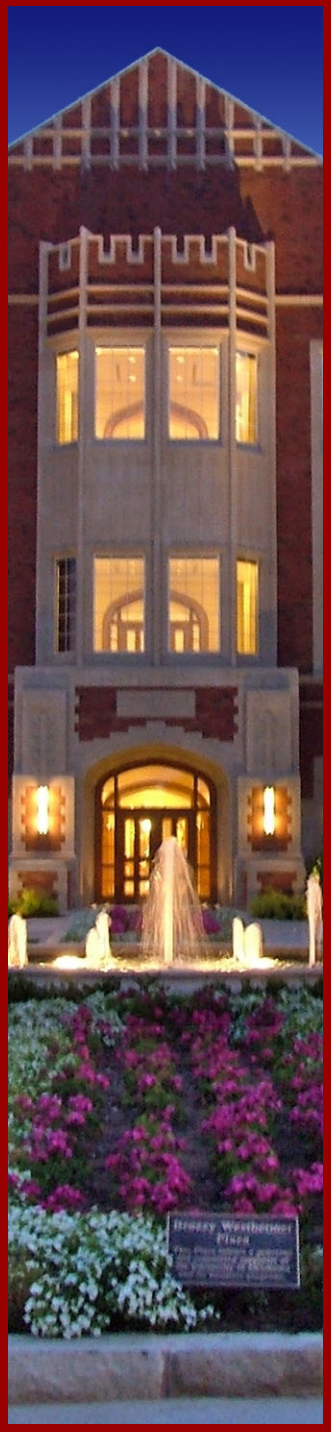
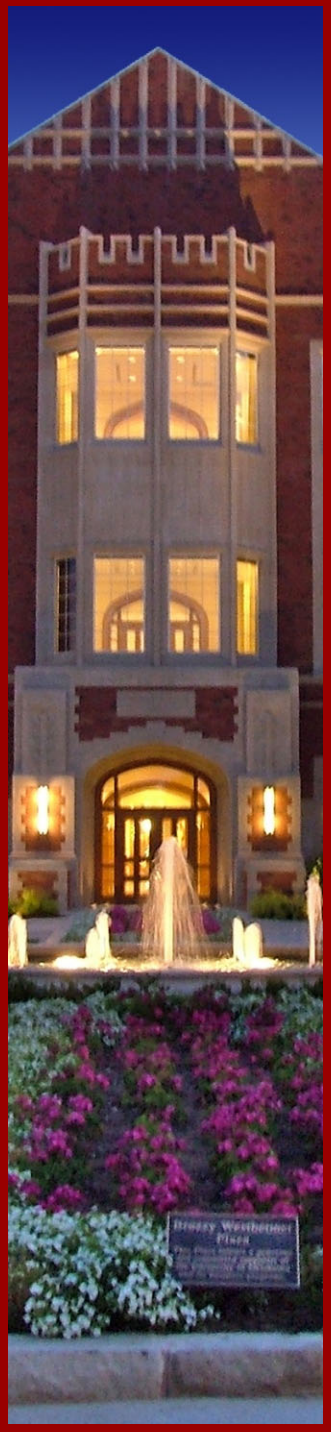


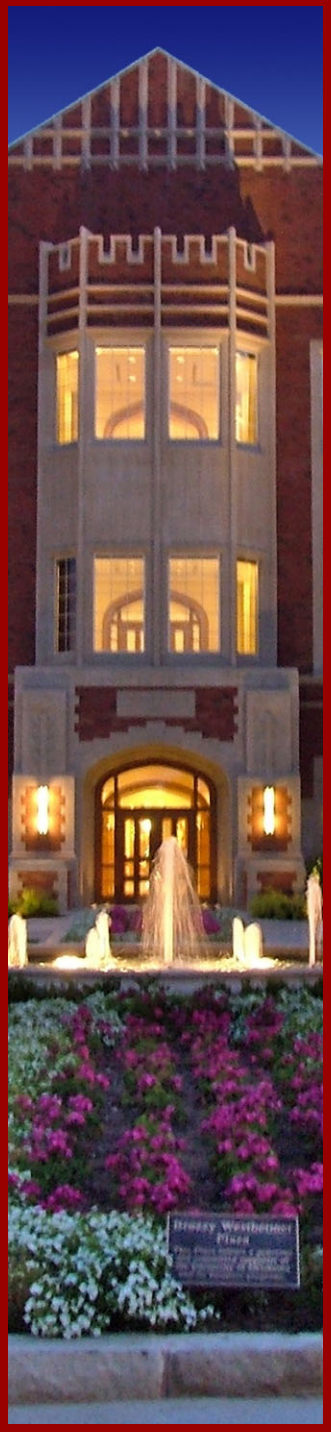
Measuring and Validating Constructs Using Computer Aided Text Analysis (CATA)

Jeremy C. Short
University of Oklahoma



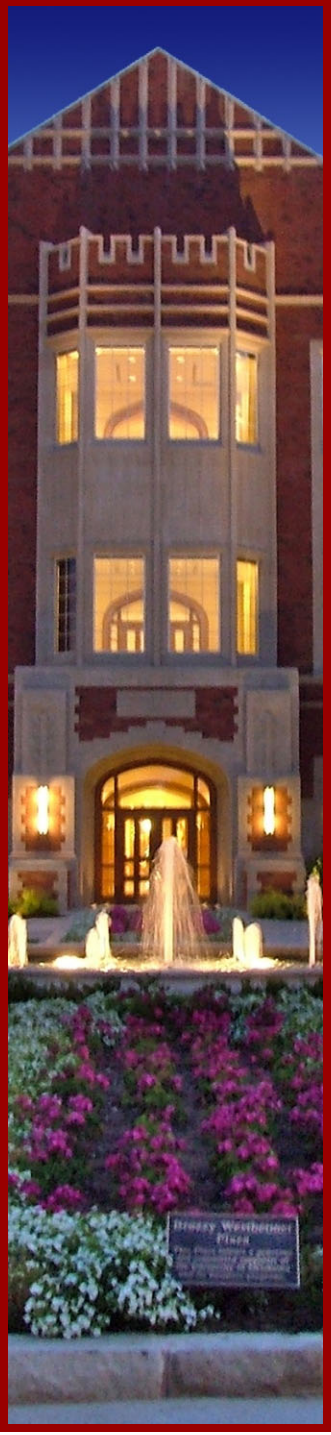
Two Views on Words

- "Words are the voice of the heart."
  - Confucius (China's most famous teacher, philosopher, and political theorist, 551-479 BC)
- "I was reading the dictionary. I thought it was a poem about everything."
  - Stephen Wright (Comedian)



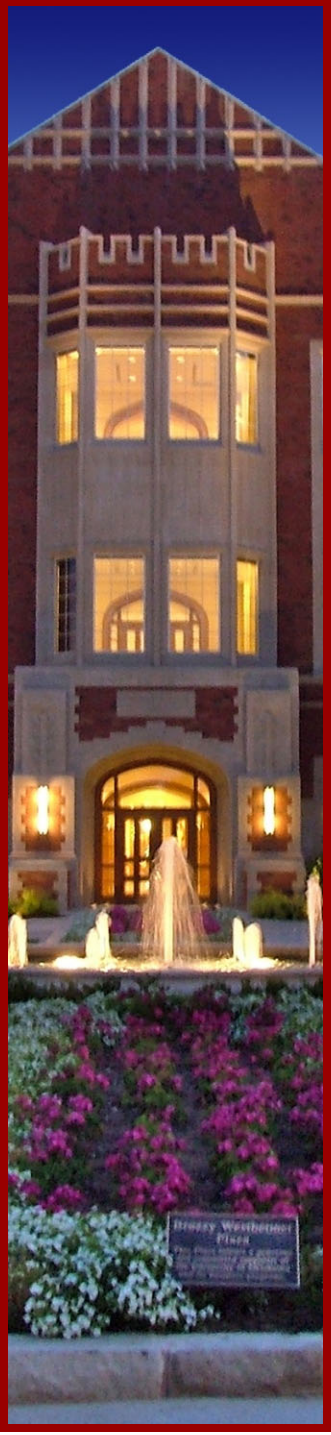
What is Content Analysis?

- Content Analysis is a research method that uses a set of procedures to classify or categorize communication.
- Content analysis allows for an unobtrusive method to gather attributions, cognitions, or other organizational projections.
- Commonly analyzed communications include shareholder letters and organizational mission statements.



Computer-Aided Text Analysis

- Computer-Aided Text Analysis (CATA) is a specific form of content analysis.
- CATA often proceeds with the assumption that word choices provide valuable information in the context of a particular organizational narrative.
- CATA is advantageous as it can allow for the processing of hundreds of documents quickly with extremely high reliabilities.
- Despite benefits, less than 25% of Content Analysis studies analyzed by Duriau, Reger, & Pfarrer (2007, ORM) used CATA.

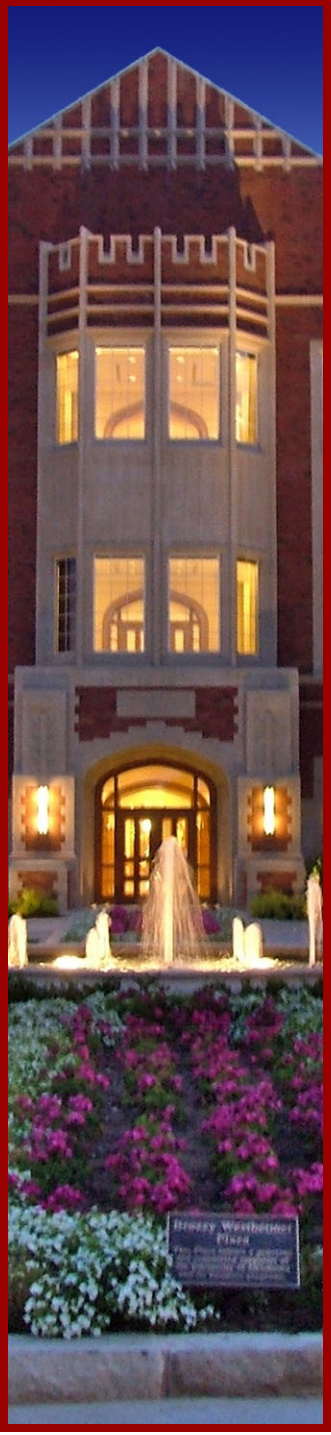


Zachary, McKenny, Short, & Payne (2011)

- **Title:** Family business and market orientation: Construct validation and comparative analysis
- **Journal:** Family Business Review
- **Variables:** Market Orientation (Customer Orientation, Competitor Orientation, Interfunctional Coordination, Long-term Focus, Profitability)
- **Narrative:** CEO Letters to Shareholders
- **Findings:**
  - Family businesses espouse a lower market orientation than non-family businesses.
  - Market orientation is positively related to firm performance.

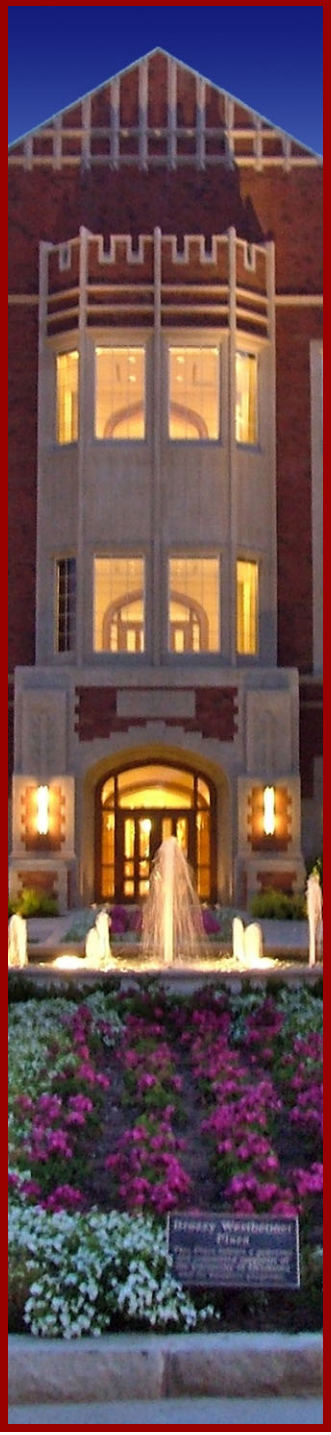
Table 1. Word List for Market Orientation Behavioral Components

Market Orientation Dimension	Content Analysis Words With Expert Validation
Customer orientation	Attendee, buyer, buying, client, clientele, consume, consumer, customer, emptor, habitué, market, marketer, patron, patronage, patronize, patronized, purchase, purchased, purchaser, purchasing, shopper, spectator, subscribe, subscribed, subscriber, subscribing, user, vend, vended, vendee, visitor
Competitor orientation	Adversary, adverse, aggression, aggressions, aggressive, ambition, ambitions, ambitious, antagonist, antagonize, antagonized, aspirant, aspire, aspired, aspires, assail, assailant, assailants, assailed, barricade, barricaded, battle, battled, battler, battles, beat, beaten, beating, bid, bidded, bidder, block, blockade, blockaded, blocked, blocks, challenge, challenged, challenger, challenges, challenging, clash, clashed, clashes, clashing, collide, collided, collides, colliding, combat, combated, combating, combative, combats, compete, competed, competer, competes, competing, competition, competitive, competitor, competitors, conflict, conflicted, conflicting, conflicts, confront, confrontation, confrontational, confrontations, conquer, conquered, conquering, conquers, contend, contender, contending, contentious, contest, contestant, contestants, counteraction, counteractions, counteractive, cutthroat, cutthroats, disputant, dispute, disputed, disputes, disputing, enemies, enemy, engage, engaged, engagement, engagements, engages, engaging, entrant, fight, fighting, fights, foe, foes, formidable, fought, grappled, grapple, grapples, grappling, jockey, jockeys, jockied, match, matched, matches, matching, opponent, oppose, opposed, opposers, opposing, opposition, oppositionist, oppositionists, oppositions, out bid, outclass, outclassed, outclassing, outmatch, outmatched, outmatches, outmatching, outrank, outranked, outranking, outranks, outrate, outrated, outrates, outrating, participant, participants, participate, participated, resist, resistance, resistant, resistants, resisted, resisting, rival, rivals, spar, sparing, sparred, spars, strive, strived, strives, striving, struggle, struggled, struggles, struggling, superior, surpass, surpassed, surpasses, surpassing, vied, vying, war, warring, aggressor, combatant, imitator, advantage, advantages



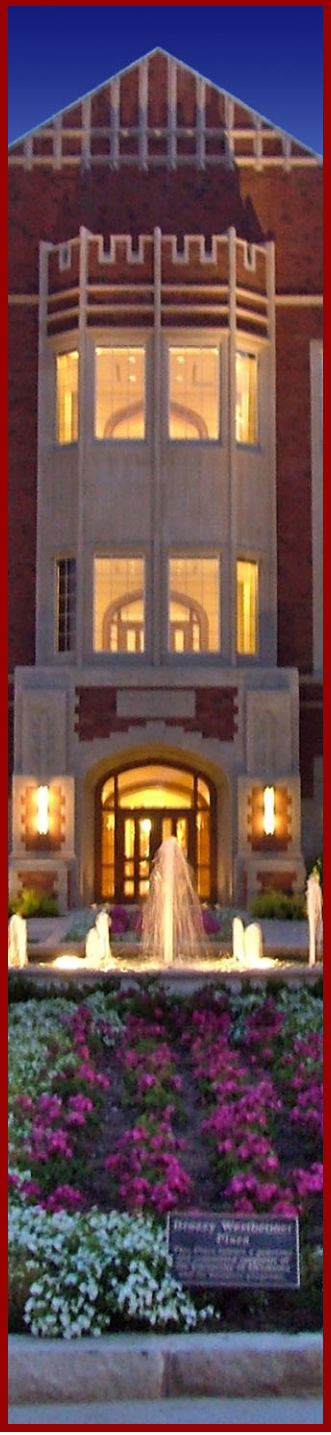
How CATA Works

- Dictionary-based coding
  - Completely automated, computer does the coding
  - Dictionaries (lists of words) are created and validated prior to analysis
- Computer looks for words from the dictionary in the narratives being analyzed
  - When it finds a word, it increments the value for that dictionary by 1



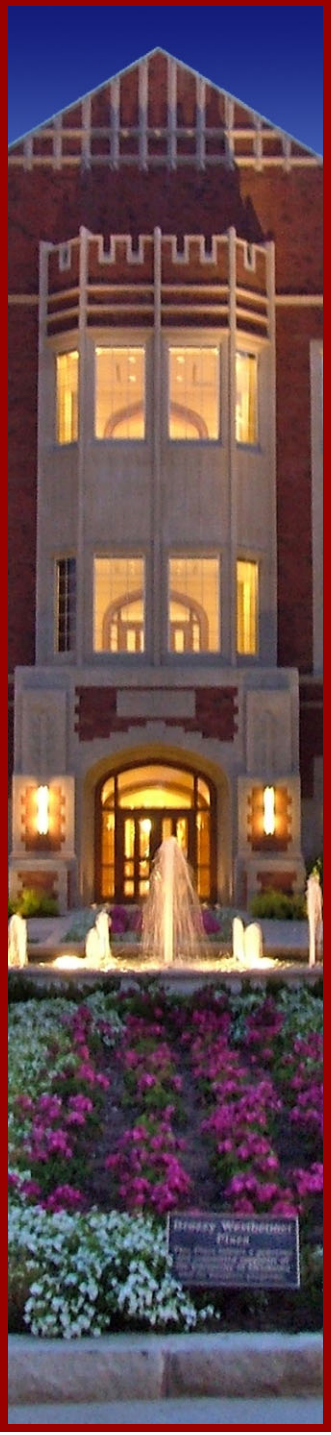
Example of Dictionary-based coding

- Simple dictionary: *Innovativeness*
  - "Innovative" "Innovation" "Innovate" "Research" "Inventions" "Inventive" "Creative" "Creativity"
- Simple 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***."
- Computer-aided text analysis result:
  - Innovativeness: 4



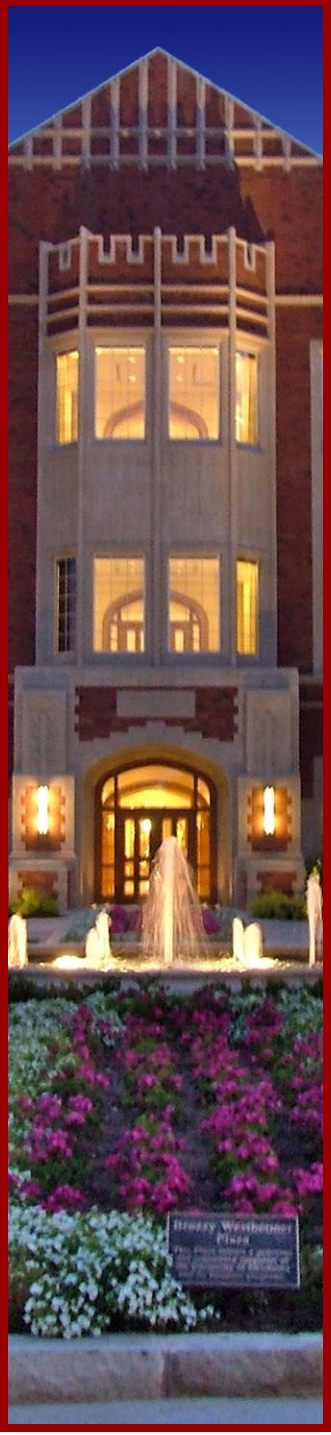
Available CATA Tools

- DICTION 6.0 - http://www.dictionsoftware.com/
- Standard dictionaries
  - 5 master dictionaries (Activity, Optimism, Certainty, Realism, Commonality)
  - 31 base dictionaries
- Custom dictionaries
  - Can handle phrases and individual words
  - Cannot handle root-words
- Cost
  - $179 (Educational); $229 (Standard)



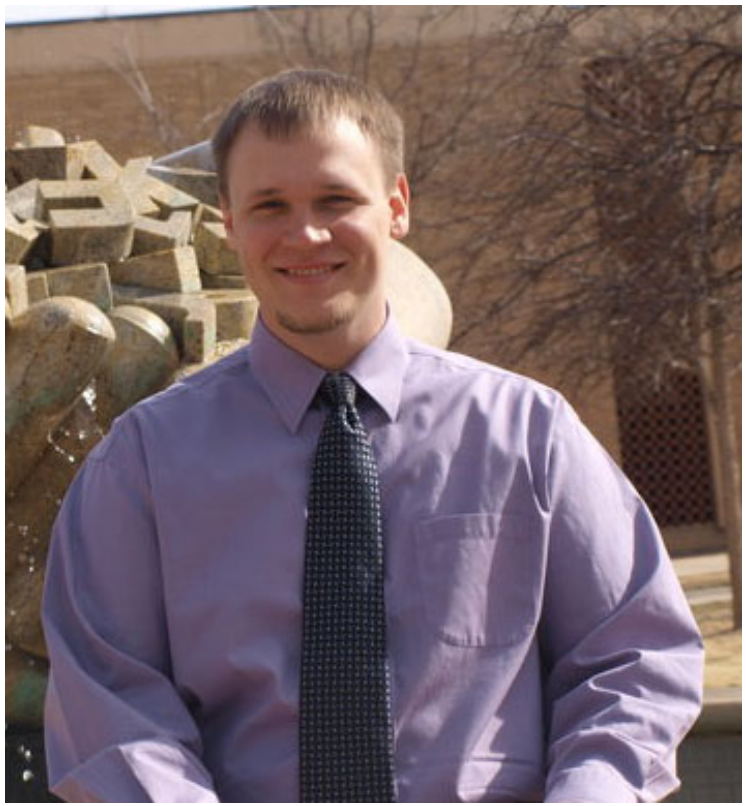
Commercial Tool Comparison

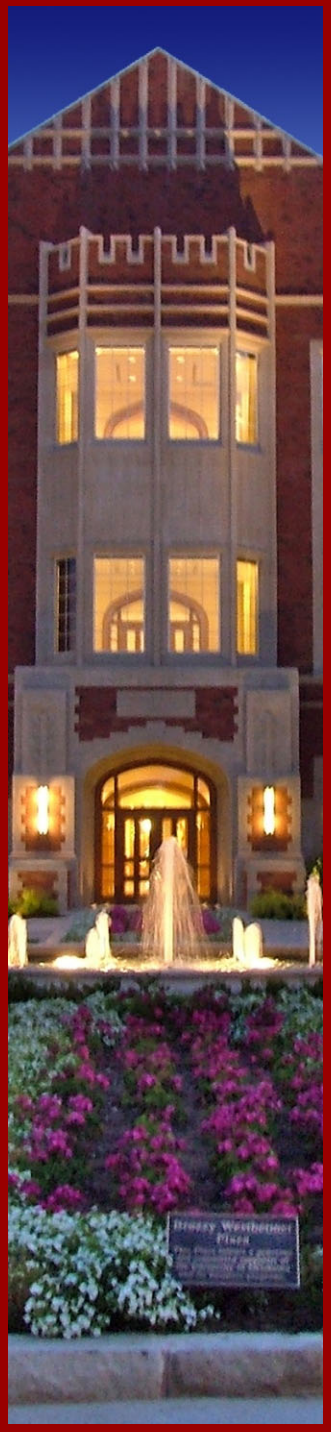
	DICTION 6	LIWC2007	LIWC2007 (Lite)	CAT Scanner
File types	PDF MS Word ASCII text	MS Word ASCII text	ASCII text	ASCII text
Speed	Moderate	Fast	Fast	Slow
Standard Dictionaries	36	70+	70+	N/A
Words	X	X	X	X
Phrases	X			X
Word-Roots		X	X	X
Inductive Word List Generation	X			X
Cost	$179	$90	$30	Free



CAT Scanner Tool

- Simple, slower, but it's free
- Three tools in one
  - Text file cleaner
  - Inductive word list generator
  - Computer-aided text analysis
  - http://www.amckenny.com/CATScanner/

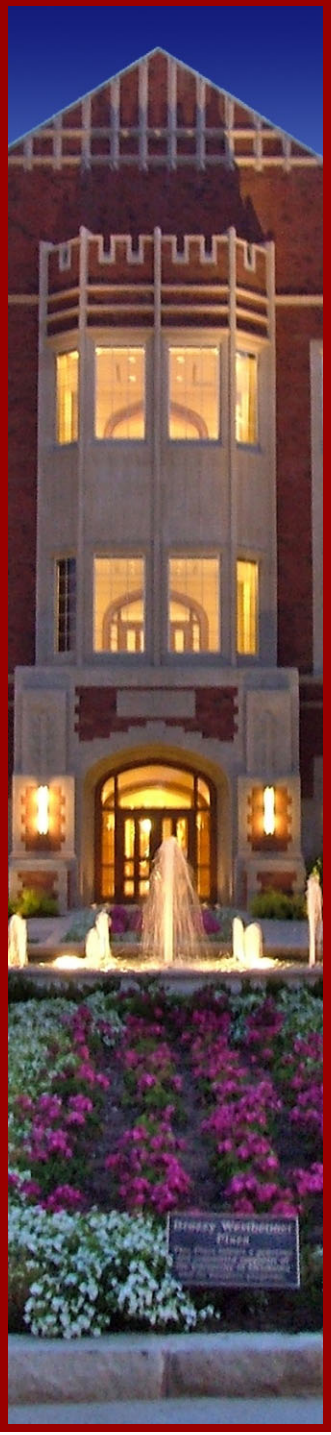




Text File Cleaner

- The problem:
  - Copying text from PDFs can result in garbage characters in the text file
  - This causes issues in CATA
    - False negatives
    - Causes software to crash (ie. DICTION)

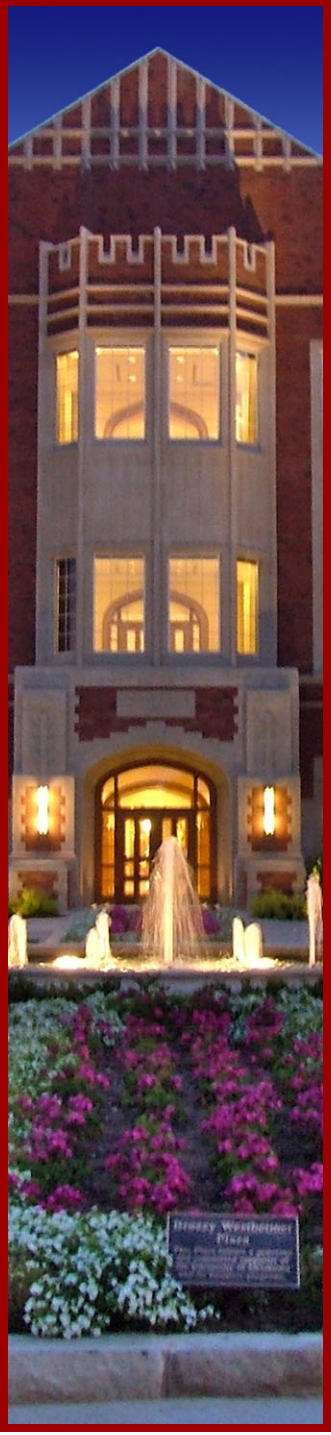
```
PAST, PRESENT AND FUTURE
The second half of fiscal 2009 and a promising
future wouldn t have been possible without
tremendous work by talented teams fully engaged in
```



…but you still need to be careful

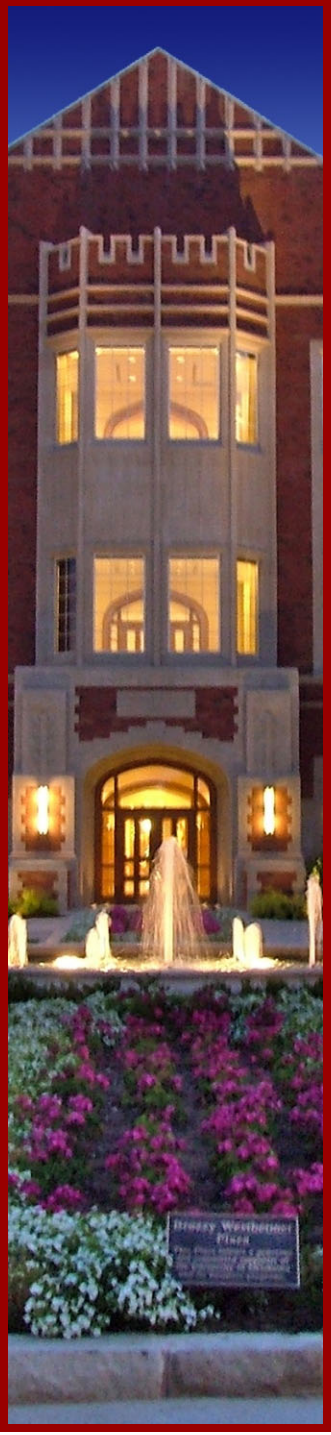
- Optical Character Recognition can still produce bogus (but garbage) text

> Ink- growing sales aluhle-digit rates in distant currencies, oui gl< >bal c< msumei products d< illai sales increase. by for the year with dollar pre-tax earnings growing by over 2H".«. Our Sally and Beauty Systems Group businesses — now with over 2.400 stores in the t\S„ Canada and overseas — again repeated strong double-digit sales and profit growth.



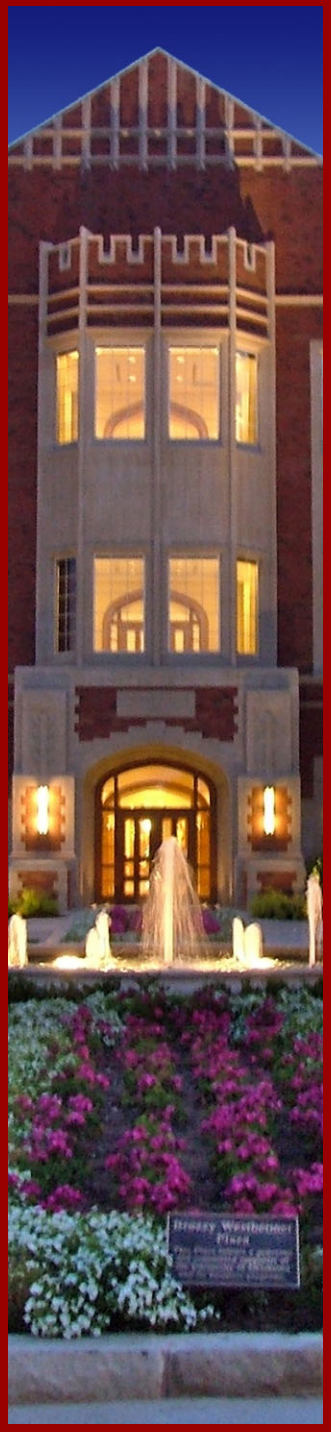
Computer-Aided Text Analysis
(With CAT Scanner)

- Runs the main analysis based on custom dictionaries you select
- Includes 18 custom dictionaries from the management literature
  - Entrepreneurial orientation (Short et al., 2010)
  - Market orientation (Zachary et al., 2011)
  - Organizational virtue orientation (Payne et al., 2011)
  - Ambidexterity (Uotila et al., 2009)



Dictionary Development and Validation

- Pretty much all of my thoughts are ‘best practices’ lifted from the following:
- 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.

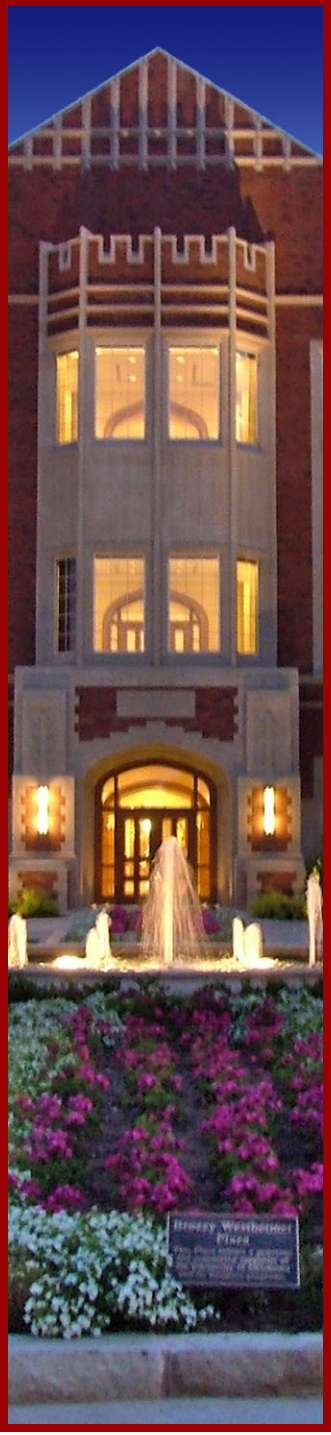


Confusion about Content Analysis

- Krippendorff (2004) suggests:
  - ‘‘Deductive and inductive inferences are not central to content analysis’’ (p. 36).
- Neuendorf (2002) suggests:
  - You may use standard dictionaries (e.g., those in Hart’s program DICTION) or originally created dictionaries. When creating original dictionaries, be sure to first generate a frequency list from your text sample, and examine for key words and phrases. (p. 50)
- Management researchers using content analytic methods generally incorporate both deductive and inductive approaches when using content analysis (Doucet & Jehn, 1997; L. Doucet, B. Kabanoff, & T. Pollock, personal communications, December 14, 2008)

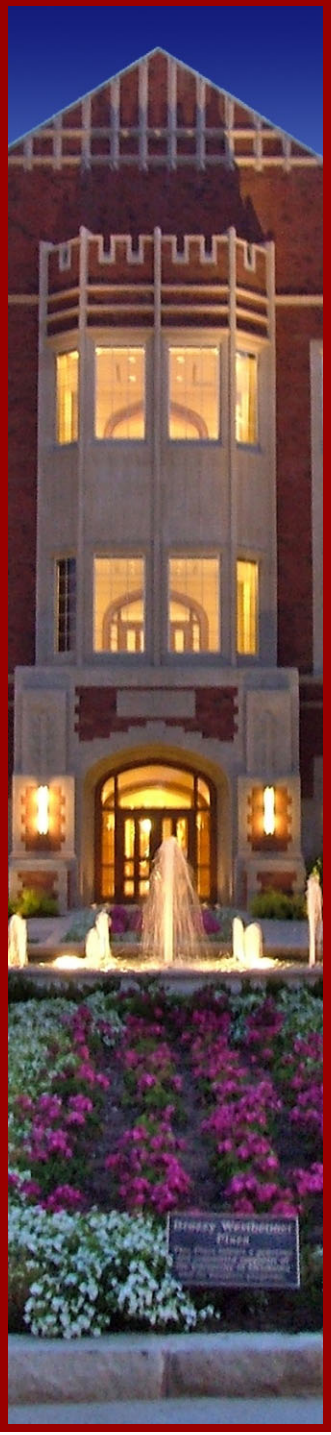
Table 1
Examination of Validity in Content Analysis Research

	Content Analysis Studies Used by Duriau et al. (2007)	Entrepreneurial Orientation Studies Using Content Analysis
Number of studies	98	9
Deductive approach	79 of 98 (81%)	9 of 9 (100%)
Evaluated content validity of word dictionaries	5 of 98 (5%)	0 of 9 (0%)
Computer-aided text analysis (CATA)	24 of 98 (25%)	0 of 9 (0%)
Assessment of multiple samples	12 of 98 (12%)	0 of 9 (0%)
Assessment of construct dimensionality	5 of 98 (5%)	0 of 9 (0%)
Assessment of predicative Validity	41 of 98 (42%)	9 of 9 (100%)
Citation support	36 of 98 (37%)	9 of 9 (100%)



Content Validity

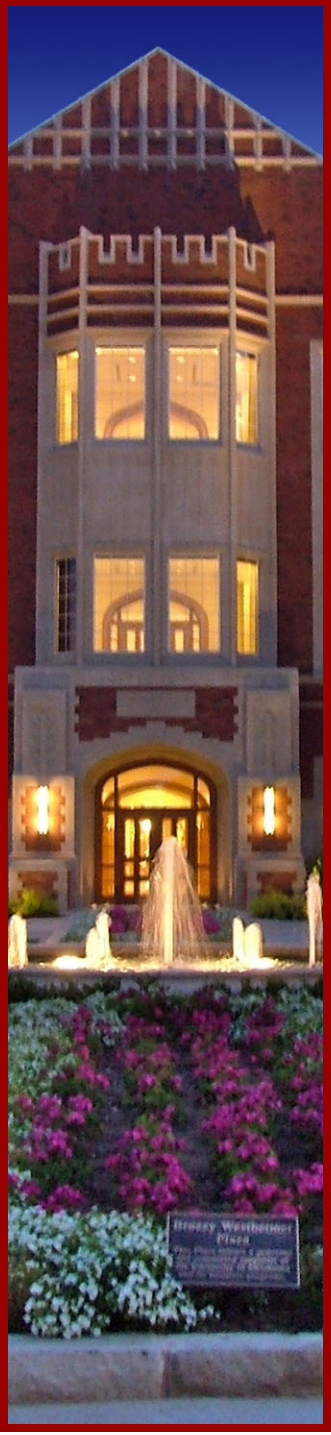
- Content validity involves an assessment examining a match between theoretical definition and empirical measurement (Nunnally & Bernstein, 1994).
- Our approach to CATA relies on single words as the unit of analysis.
- Key steps involve content validity, external validity, reliability, assessment of dimensionality, and predictive validity.
- I illustrate our approach using the construct of entrepreneurial orientation.



Two-Step Deductive+Inductive Approach

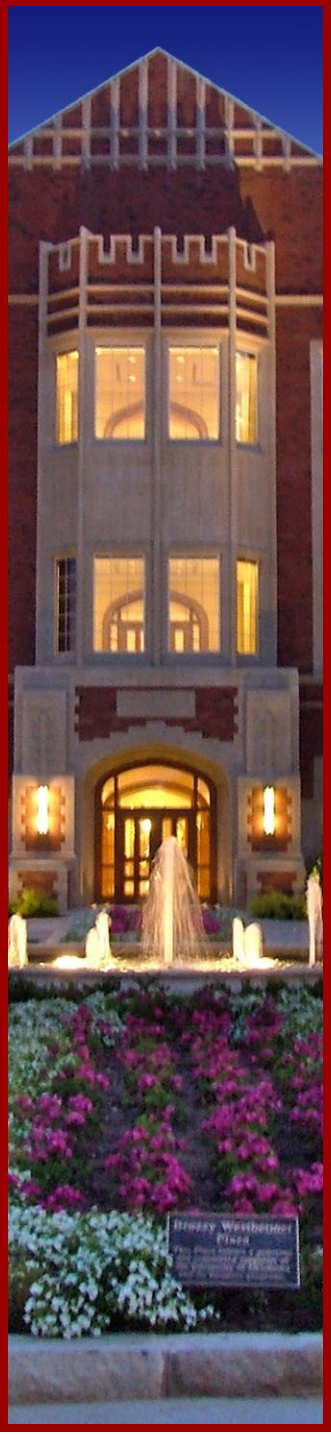
- Maximizes content validity
  - Deductive word list: Identify words based on the definition of the construct
  - Inductive word list: Identify words from the narratives of interest

- Combined, they help to ensure entire content domain of the construct is covered.



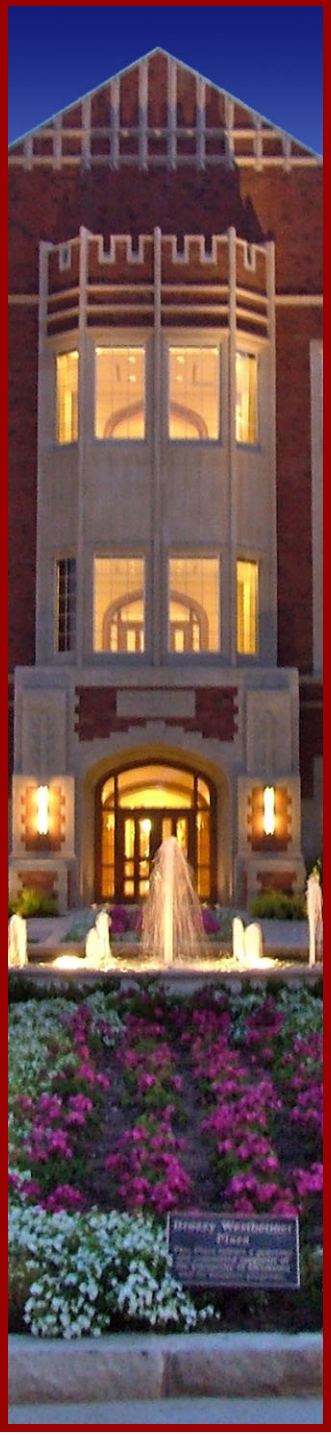
Deductive Content Validity

1. Create a working definition of the construct of interest using a priori theory when possible
   - We begin by identifying a formal definition of entrepreneurial orientation offered by Lumpkin and Dess (1996 AMR), who define the construct as the, "processes, practices, and decision-making activities that lead to new entry" (p. 136).



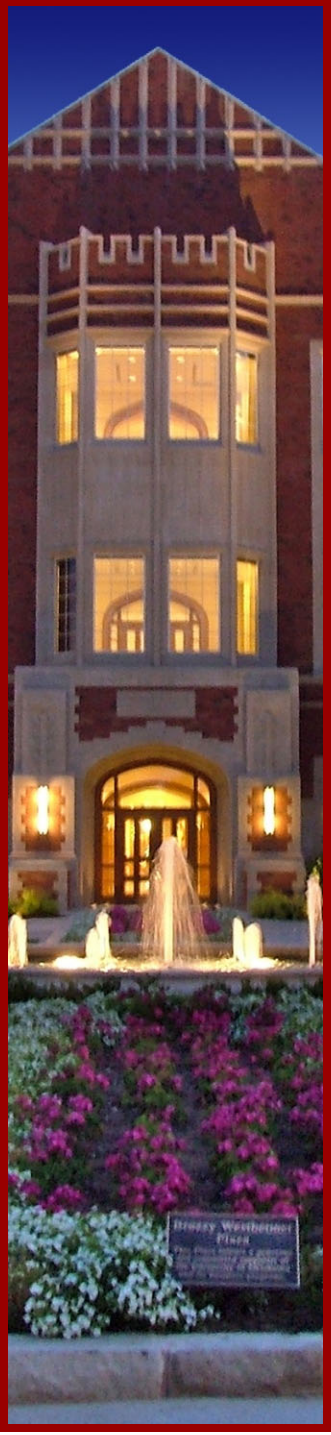
Deductive Content Validity

2. Conduct initial assessment of construct dimensionality
   - Five independent but related dimensions:
   - Autonomy
   - Competitive aggressiveness
   - Innovativeness
   - Proactiveness
   - Risk taking



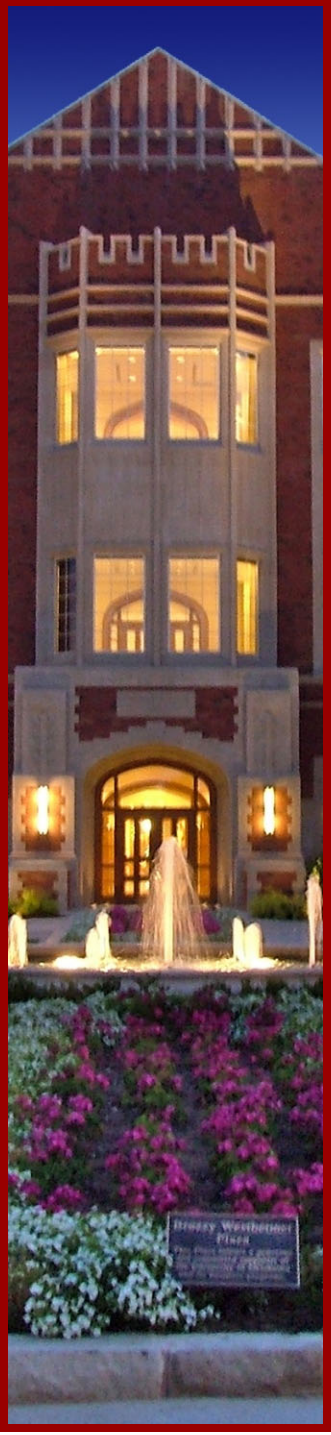
Deductive Content Validity

3. Develop initial list(s) of words.
   - One list per dimension
   - Identify 2-3 core words that are representative of each dimension
   - Use a synonym finder (e.g., Rodale's) to identify words associated with those core words and their variants



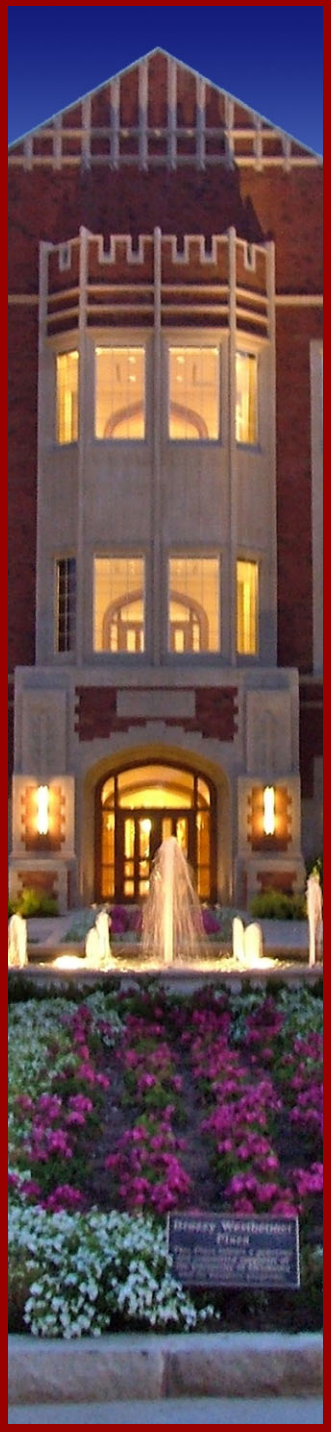
Deductive Content Validity

4. Validate word lists using content experts to assess rater reliability.
   - We provide an Excel tool to facilitate this at: http://www.amckenny.com/CATScanner/resources.php
   - We use Holsti (1969): $PA_O = 2A/n_A + n_B$
     - PAO = Proportion Agreement Observed
     - A = Number of agreements between two raters
     - n_A and n_B are the number of words coded by the two raters
   - No generally accepted rules-of-thumb
     - Riffe, Lacy, and Fico (2005) and Krippendorff (2004) suggest interpreting values greater than .80.



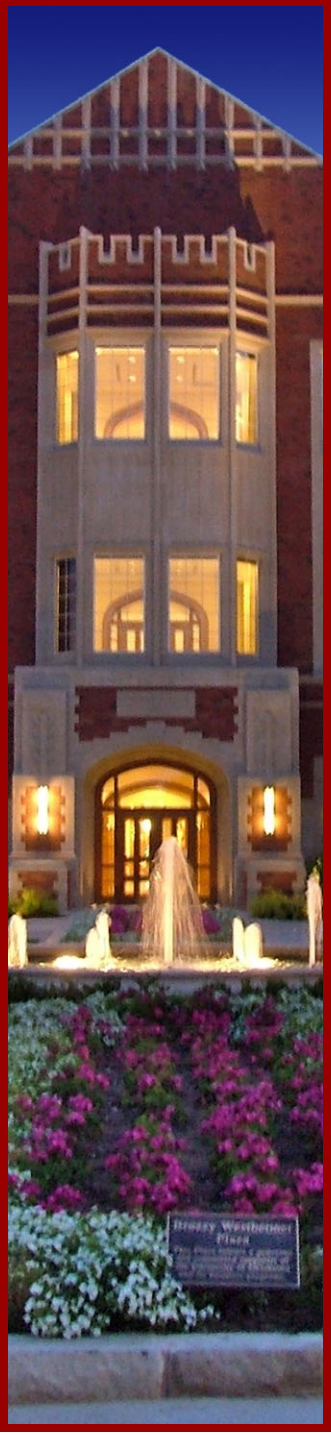
Inductive Content Validity

1. Identify commonly used words from narrative texts of interest
   - DICTION
   - CAT Scanner
2. Identify or create a working definition of the construct of interest to guide word selection.



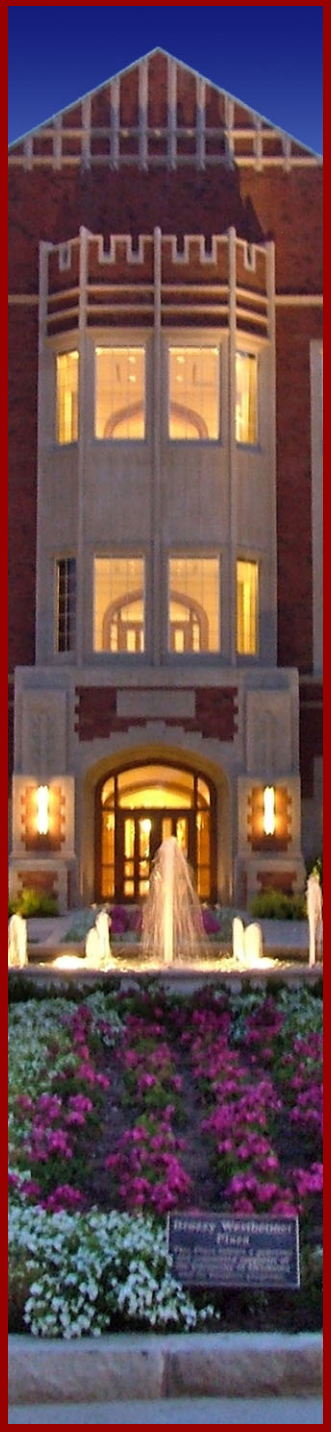
Inductive Content Validity

3. Have judges identify words that match the construct of interest.
4. Establish interrater reliability.
5. Refine and finalize word lists.



Final Word List for Autonomy

At-liberty, authority, authorization, autonomic, autonomous, autonomy, decontrol, deregulation, distinct, do-it-yourself, emancipation, free, freedom, freethinking, independence, independent, liberty, license, on-one's-own, prerogative, self-directed, self-directing, self-direction, self-rule, self-ruling, separate, sovereign, sovereignty, unaffiliated, unattached, unconfined, unconnected, unfettered, unforced, ungoverned, unregulated



External Validity

1. Select appropriate samples and relevant narrative texts to examine construct of interest
   - We chose shareholder letters since entrepreneurial orientation has been conceptualized as a firm level construct.
2. Compare two relevant samples when possible.
   - We compare the S&P 500 with high growth firms from the Russell 2000.

Table 4
Evidence of Language Representing Entrepreneurial Orientation Dimensions in Shareholder Letters of Firms in S&P 500 and Russell 2000®

	S&P 500 Sample				Russell 2000® Sample			
	N	Mean	*SD*	*t* Test	*N*	Mean	*SD*	*t* Test
Autonomy	453	1.01	2.11	10.23*	205	.65	1.07	8.70*
Competitive aggressiveness	453	3.13	2.99	22.25*	205	2.04	2.17	13.46*
Innovativeness	453	12.76	8.94	30.38*	205	9.13	7.75	16.86*
Proactiveness	453	2.95	3.20	19.60*	205	3.07	3.62	12.14*
Risk taking	453	1.11	2.05	11.50*	205	.78	2.02	5.49*
Inductively derived entrepreneurial orientation words	453	18.94	13.46	29.85*	205	15.22	10.85	20.03*

Note: The results of this table were based on computer-aided text analysis using the word lists for entrepreneurial orientation presented in Table 3.

*$p < .01$.

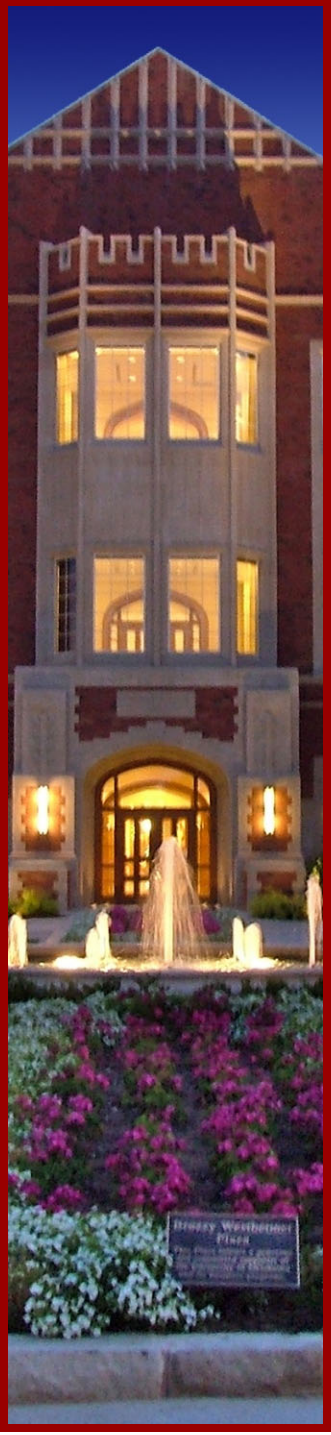
Table 5
ANOVA Comparisons of S&P 500 to Russell 2000® Firms on Entrepreneurial Orientation Dimensions

Entrepreneurial Orientation Dimension	S&P 500 Sample ($N = 205$)	Russell 2000® Sample ($N = 205$)	F Test
Autonomy	.61	.57	.16
Competitive aggressiveness	1.87	1.64	2.43
Innovativeness	7.31	7.15	.15
Proactiveness	1.67	2.61	14.68**
Risk taking	.76	.53	4.00*
Inductively derived entrepreneurial orientation words	11.42	14.06	9.16**
Total entrepreneurial orientation	20.54	21.90	3.05

Note: ANOVA = analysis of variance. The results of this table were based on computer-aided text analysis using the word lists for entrepreneurial orientation presented in Table 3. Entrepreneurial orientation dimensions are standardized by the number of words in the shareholder letter.

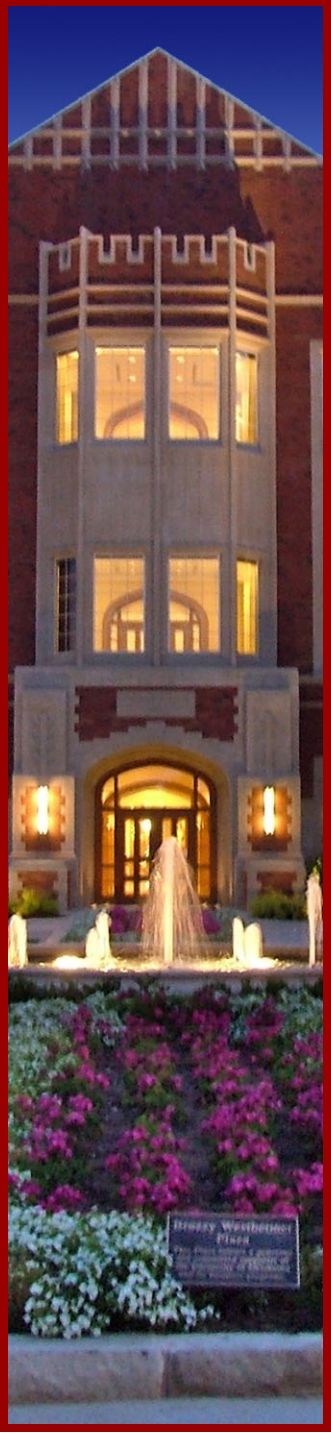
*$p < .05$.

**$p < .01$.



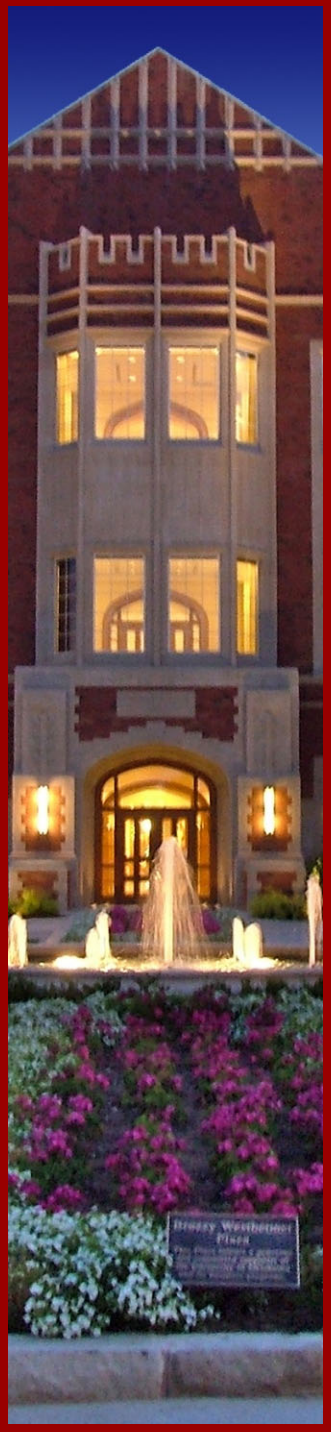
Reliability

- Assure reliability by analyzing texts using a computer-aided technique.
  - CAT Scanner
  - DICTION
  - LIWC



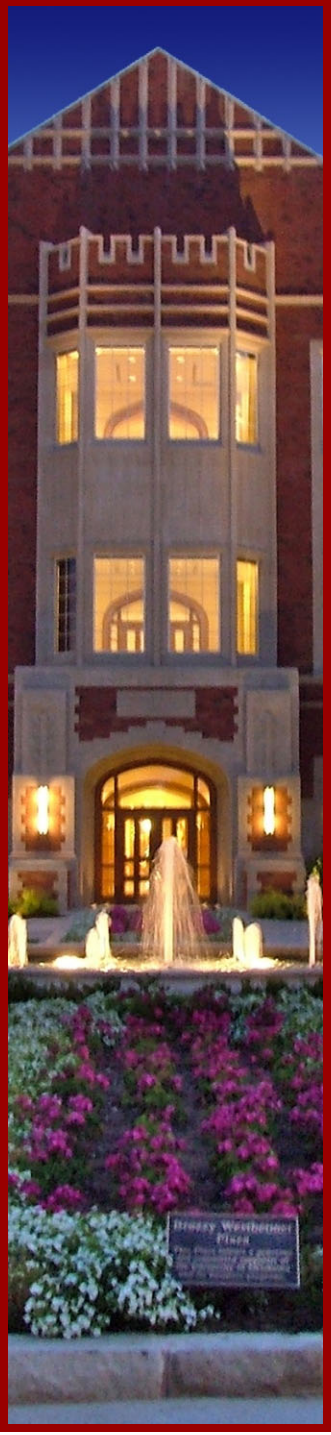
Dimensionality

- Assess construct dimensionality using visual inspection of the correlation matrix.
- If dimensions are uncorrelated, they might be assessing different constructs and dimensions might exhibit problems of convergent validity.
- If dimensions are correlated over .5, the construct may not be multidimensional. If dimensions are too highly correlated, consider collapsing subdimensions to form a single measure (or fewer subdimensions)



Correlation Matrix

1. Autonomy	*.80*				
2. Competitive Aggressiveness	.15**	*.75*			
3. Innovativeness	.17**	.38**	*.88*		
4. Proactiveness	.10*	.20**	.30**	*.85*	
5. Risk Taking	.11*	.21**	.06	.09*	*.83*

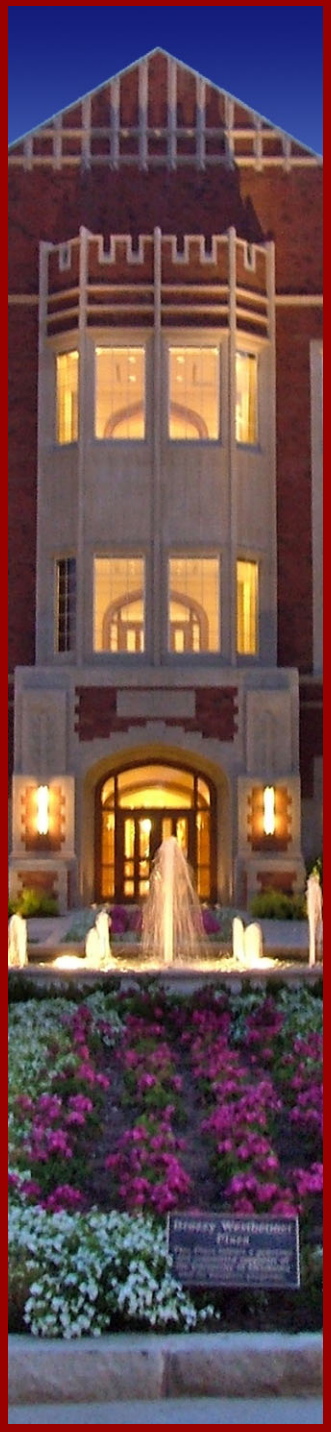


Predictive Validity

- Examine ability to predict theoretically related variables not captured via content analysis using regression or other appropriate multivariate technique.

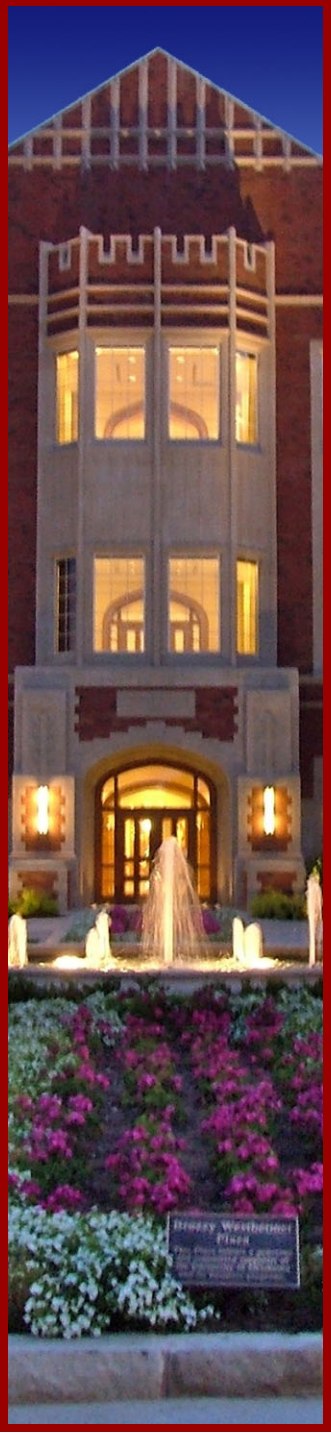
Table 7
Hierarchical Regression Analyses for Deductively Defined Entrepreneurial Orientation Measures Predicting Tobin's *Q*: S&P 500 Firms and Russell 2000® Firms

	S&P 500 Sample		Russell 2000® Sample	
	1	2	1	2
Firm size (logarithm of employees)	−.09	−.13**	−.18**	−.14*
Autonomy		−.03		.02
Competitive aggressiveness		−.08		−.11
Innovativeness		.19**		.15*
Proactiveness		.12*		.18*
Risk taking		−.21**		−.15*
R^2	.01	.12	.03	.09
F for change in R^2	3.46*	11.09**	6.05*	3.97**



Discriminant and Convergent Validity

- Discriminant validity involves the extent to which a construct is distinct from other constructs (Campbell & Fiske, 1959).
- Convergent validity examines the extent to which a measure captures the same/similar variance as other measures of the same/similar construct (Campbell & Fiske, 1959).
- Can be assessed by examining correlations



Elevating Construct Level of Analysis

- McKenny AF, Short JC, Payne GT. (In press). Using Computer-aided text analysis to elevate constructs: An illustration using psychological capital. Organizational Research Methods.
- In Press at *Organizational Research Methods*
  - How to elevate a construct to an aggregate level of analysis using CATA
  - Our example: Psychological Capital (Luthans, Youssef, & Avolio, 2007)
- Additional considerations when changing level of analysis