Are we overconfident in executive overconfidence research? An examination of the convergent and content validity of extant unobtrusive measures

Aaron D. Hill [a,*], David A. Kern [b,1], Margaret A. White [c,2]

[a] Spears School of Business, Oklahoma State University, 211 Business Building, Stillwater, OK 74078, United States
[b] College of Business and Technology, Northeastern State University, BT 221B, Tahlequah, OK 74464, United States
[c] Spears School of Business, Oklahoma State University, 350 North Hall, 700 N. Greenwood, Tulsa, OK 74106, United States

ARTICLE INFO

ABSTRACT

Building understanding of overconfident executives is central to a growing literature that spans a number of disciplines. Much of this research has utilized unobtrusive, or indirect, measures to assess executive overconfidence from secondary data sources. We analyze the convergent and content validity of seven extant unobtrusive measures of executive overconfidence. The results of our analyses indicate that these measures do not exhibit adequate convergence, suggesting that existing measures are not measuring the same construct. Further, we administer a sort task to academic colleagues to assess whether scholars believe that the seven measures are adequately assessing the intended construct. The results of our sort task indicate that scholars did not categorize any of the seven measures as sufficient for measuring overconfidence. We conclude with suggestions for future research to address the inadequate convergent and content validity found in our assessment of extant measures of executive overconfidence.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Executive overconfidence, defined as executives' tendency to inflate or overestimate their own abilities (DeBondt & Thaler, 1995; Malmendier & Tate, 2005), is central to a growing, cross-disciplinary research stream focused on how executives affect the behaviors of organizations. Because of limitations with direct measurement when gathering data from executives (Hambrick & Mason, 1984) and the lack of a validated instrument for use in direct inquiries (Hayward & Hambrick, 1997; Hiller & Hambrick, 2005) as well as the benefits of secondary data, scholars have developed multiple "unobtrusive" measures (Webb, Campbell, Schwartz, & Sechrest, 1966) from secondary data to assess executive overconfidence (Bollaert & Petit, 2010; Hill, Kern, & White, 2012). The utilization of multiple unobtrusive measures to assess executive overconfidence presents two problems. First, it can be difficult to interpret results across studies using different measures, and comparisons may not be reliable if the measures do not exhibit adequate convergent validity or agreement with respect to the construct they are attempting to assess (Campbell & Fiske, 1959; Churchill, 1979; Jacoby, 1978;

Venkatraman & Grant, 1986). Accordingly, if measures lack convergent validity, the nomological validity, or degree to which measures of a construct exhibit the expected statistical relationship with other constructs, will likewise be inadequate (Jacoby, 1978; Lubatkin, Merchange, & Srinivasan, 1993). Second, existing measures of executive overconfidence are criticized as lacking content validity (Bollaert & Petit, 2010; Hiller & Hambrick, 2005; Jin & Kothari, 2008), or the degree to which the measures adequately assess the construct (Churchill, 1979; Hinkin, 1998).

Despite concerns associated with a lack of both convergent and content validity of unobtrusive measures of executive overconfidence, scholars have yet to provide evidence as to whether these measures exhibit adequate agreement or whether they adequately assess the construct. For the executive overconfidence research stream to continue to develop as well as to assure that we can have confidence in interpretations from existing and future research alike, it is essential to understand whether extant measures agree with respect to what they are attempting to measure, and further whether these measures are adequately measuring what they attempt to measure. The purpose of this paper is to assess the convergent and content validity of extant unobtrusive measures of executive overconfidence.

The results of our assessment of convergent validity of extant unobtrusive measures of executive overconfidence suggest that these measures do not exhibit adequate convergence. Further, we present an evaluation of content validity using scholarly raters that suggests that the extant unobtrusive measures of executive overconfidence are not adequately assessing the construct. Cumulatively, these findings

* Corresponding author. Tel.: + 1 405 744 5193.
  E-mail addresses: aaron.hill@okstate.edu (A.D. Hill), kernd@nsuok.edu (D.A. Kern), margaret.white@okstate.edu (M.A. White).
  [1] Tel.: + 1 918 444 3089.
  [2] Tel.: + 1 918 594 8434.

confirm concerns with respect to the validity of unobtrusive measures of overconfidence (Bollaert & Petit, 2010; Hiller & Hambrick, 2005; Jin & Kothari, 2008). In addition, since extant measures of executive overconfidence do not exhibit adequate convergent or content validity, interpretations within the existing research stream should be viewed with caution. Because research on executive overconfidence continues to generate interest among scholars, we conclude with suggestions for moving this line of research forward. The suggestions we offer may also be of benefit to scholars interested in utilizing unobtrusive methods to assess other psychological constructs as well.

2. The construct of overconfidence

Overconfidence is a heuristic bias defined as individuals' tendency to overestimate their own abilities (e.g., DeBondt & Thaler, 1995; Hiller & Hambrick, 2005; Malmendier & Tate, 2008). The upward bias in the assessment of their own abilities affects how individuals view and respond to the situations they face (Busenitz & Barney, 1997; Malmendier & Tate, 2005). While psychologists have long utilized the pejorative term "overconfidence" in reference to the tendency of individuals to inflate estimates of their abilities (Meehl, 1957; Oskamp, 1965), scholars also refer to the construct as "hubris" (Hayward & Hambrick, 1997; Roll, 1986), utilizing a term originating from ancient Greece. The traditional definition of hubris utilized by ancient Greek scholars is that hubris is an individual trait defined by "a disposition of overconfidence" (Cairns, 1996, p. 1; see also Dickie, 1984). Given this view of hubris, it is perhaps not surprising that several scholars employ the terms "overconfidence" and "hubris" interchangeably (Hayward, Shepherd, & Griffin, 2006; Li & Tang, 2010; Malmendier & Tate, 2008) or that scholars who use the term "hubris" implicitly link to overconfidence by phrasing the definition of hubris as exaggerated self-confidence (Hayward & Hambrick, 1997; Hiller & Hambrick, 2005; Li & Tang, 2010). Indeed, as Li and Tang (2010, p. 45) note, both overconfidence and hubris refer to the tendency of "decision makers to overestimate their own abilities" and research by Hill et al. (2012, p. 188) has confirmed Li and Tang's view by providing examples of the equivalent use of the terms in the literature, concluding that both terms are employed in reference to "the tendency of individuals to overestimate their abilities." While we employ the term "overconfidence" exclusively for clarity, we highlight the practice of referring to overconfidence as hubris to link to the literature that employs the latter term. Further, to assess convergent and content validity of executive overconfidence measures adequately, we need to include measures referring to both overconfidence and hubris because in practice the two terms are used to refer to the same construct.

3. Method

3.1. Identifying extant measures

To identify measures of executive overconfidence, we conducted a literature review using computerized search tools and the keywords "overconfidence" and "hubris." We then reviewed the references of the identified articles to locate additional research. Our search resulted in over 250 research papers. The papers were reviewed to identify empirical tests where executive overconfidence was assessed using unobtrusive measurement; our review produced seven unobtrusive measures that provide the basis of our study. Two extant articles in the executive overconfidence/hubris literature identify the same seven measures (Bollaert & Petit, 2010; Hill et al., 2012); thus, we have confidence that our search resulted in a reasonably comprehensive list of extant measures.

3.2. Sample

Since data requirements differ across unobtrusive measures of executive overconfidence, we chose a sample that would eliminate as many confounds as possible. For several reasons, we selected publicly traded firms operating in the United States as our sample. First, because we need company data to calculate certain unobtrusive measures of executive overconfidence, publicly traded firms were essential since private firms are not required to disclose this information. Second, one measure requires comments made by media members in reference to an executive. Since large firms systematically receive more coverage in the media (Gans, 2005), using large firms in concert with smaller firms may constitute a naturally occurring selection bias. To avoid the concern that the media-based measure of executive overconfidence will systematically vary by firm size and subsequently bias measurement and analyses, we selected only firms with a minimum revenue threshold of $1 billion dollars. Third, because uncertainty in a firm's industry may affect various overconfidence measures, we used only firms that operate in industries that exhibit relatively little dynamism to enhance comparability (Cameron, Whetten, & Kim, 1987). We identified 103 firms in the year 2000 meeting our requirements; incomplete data on nine firms result in a final sample of 94 firms.

For each of the unobtrusive measures of executive overconfidence, we replicated the methodology of the original authors with the CEO as the focal executive. Given that our first purpose is to assess convergent validity of extant measures, we remained agnostic on the quality of any measure and rather replicated prior approaches. As such, we withheld comments on the strengths and weaknesses of each measure until after our analyses were complete.

3.3. Measures of executive overconfidence

3.3.1. Executive language use

Using letters to shareholders during the focal year and two preceding years, a count of the number of sentences that included personal statements in reference to the organization was divided by the total number of sentences to measure executive language use (Rovenpor, 1993).

3.3.2. Media comments

We replicated Hayward and Hambrick (1997) and coded how the media viewed the CEO using nationally distributed newspaper and magazine articles about the CEO. Specifically, the rating scale we employed is: 3 = the article is unequivocally favorable; 2 = the balance of the article was favorable but included some unfavorable remarks; 1 = the article was neither favorable nor unfavorable; −1 = the balance of the article was unfavorable but included some favorable remarks; −2 = the article was unequivocally unfavorable; and 0 = the CEO receives no media coverage. We summed the scores for all articles to arrive at our measure based upon media comments. As in the method we replicate, we omitted articles that quote or name the CEO only. The Dow Jones Factiva database for magazines and newspapers, which included more articles than either Lexis/Nexis or ABI, served as our source for articles because the broader source may alleviate some concern over selection bias of articles.

3.3.3. Recent organizational performance

We replicated the Hayward and Hambrick (1997) measure using such recent organizational performance as stockholder returns (stock price appreciation over the preceding year plus the dividend yield) divided by the initial stock price.

3.3.4. Organizational investments

Malmendier and Tate (2005, p. 2661) argue and find empirical support for the notion that overconfident CEOs "overestimate the returns to their investment projects" and thus that the investment level of the

Table 1
Descriptive statistics and correlations.

Variable	Mean	s.d.	1	2	3	4	5	6
1. Executive language use	0.03	0.04						
2. Media comments	2.43	4.83	.17					
3. Recent organizational performance	25.99	38.10	−.02	.38**				
4. Organizational investment	0.12	0.09	.13	.39**	.35**			
5. Relative compensation	1.92	1.01	.08	.07	−.01	.12		
6. Stock option exercise	10.96	8.20	−.07	.51**	.55**	.23*	−.14	
7. Stock purchases	0.49	2.27	−.04	.02	.02	−.01	.15	−.04

* $p < .05$.
** $p < .01$.

organization may provide insight into the CEO's overconfidence. We replicated their measure of organizational investment and used capital expenditures divided by beginning of year plant, property, and equipment; we then divided this value by industry-adjusted averages (see also Campbell, Johnson, Rutherford, & Stanley, 2009).

3.3.5. Relative compensation

Hayward and Hambrick (1997) argue that a CEO's belief that or his or her abilities are superior and hence that he or she is more valuable is manifest in the CEO's compensation relative to the next highest paid executives. We replicated their use of relative compensation as an indicator. Specifically, we measured the relative compensation of the CEO to the next highest paid executive as the ratio of the cash compensation of the CEO to the highest paid non-CEO officer (Hayward & Hambrick, 1997).

3.3.6. Stock option exercise

Because stock options cannot legally be traded or short-sold and holding exercisable options increases risk, executives typically exercise options immediately as they vest (as long as they have a positive value). As Malmendier and Tate (2008, p. 24) argue, "one interpretation of failure to exercise is overconfidence." We measured overconfidence based upon stock option exercise using the estimated value of each CEO's exercisable stock options that were unexercised (Malmendier & Tate, 2005, 2008).

3.3.7. Stock purchases

As with failure to exercise in-the-money exercisable options, purchasing additional stock in the company suggests that CEOs are confident in their ability to increase the value of the stock. We replicated the Malmendier and Tate (2005) measure using stock purchases by calculating the net change in CEOs' ownership of the stock in their employing organizations, excluding options and grants.

4. Analysis

4.1. Convergent validity

We first analyze whether extant measures exhibit convergent validity (Campbell & Fiske, 1959; Venkatraman & Grant, 1986). Although assessments of reliability typically precede convergence, this may be inappropriate here for two reasons (Cortina, 1993 p. 103): first, if measures do not converge on one factor, reliability is not appropriate since it does not provide evidence that measures "are measuring the construct or constructs they are intended to measures" and therefore, it may be necessary to "make sure there are no large departures from unidimensionality" prior to assessing reliability; second, given that scholars utilize extant measures individually as regressors to represent independent variables in their analyses, not as items in a summated measure, unidimensionality has been assumed — thus, alpha may be inappropriate. We do find, however, a Cronbach's alpha of 0.51 and an average inter-item covariance of 0.13, suggesting poor reliability of the measures if we treat them as a summated measure. We also

note that scholars use multiple measures as formative indicators of overconfidence; but in each case, they use the measures as individual regressors as well. We highlight each measure as an individual measure here given that scholars both: a) use the measures as individual regressors to represent overconfidence, and b) suggest that formative measurement of psychological constructs such as overconfidence may be problematic (Diamantopoulous, Riefler, & Roth, 2008; Edwards, 2011). We return to formative measurement in our discussion.

Two methods of assessing convergent validity are commonly employed (Brahma, 2009; Lubatkin et al., 1993). The first, correlational analysis (Nunnally, 1978), assesses convergence by analyzing the correlations between variables (Chatterjee & Blocher, 1992; Lubatkin et al., 1993). Descriptive statistics and correlations for each of the seven measures are provided in Table 1. Nonsignificant correlations suggest that the items do not measure the same construct (Churchill, 1979; Nunnally, 1978). As Table 1 indicates, several measures are uncorrelated with other measures: only six of the 21 possible relationships are significant, and three of the measures (Executive Language Use, Relative Compensation, and Stock Ownership) are uncorrelated with any other measure. While some of the measures do correlate and thus may be assessing a similar construct, as a whole the results suggest that at least some of the measures that scholars utilize to measure overconfidence are not measuring the same construct (Churchill, 1979; Nunnally, 1978).

A second method for assessing convergence is factor analysis. We conduct an exploratory factor analysis (EFA) for three reasons. First, scholars suggest that EFA should precede confirmatory factor analysis (CFA) because if the measures do not converge with EFA, there is no theoretical justification for moving to CFA (Brahma, 2009; Ford, MacCallum, & Tail, 1986; Hinkin, 1998). Second, if the results of EFA justify the contention that each of the measures assesses a common construct, a second sample can be utilized to confirm this contention with CFA (Kline, 2005). Third, because the results of correlational analysis suggest that these measures do not exhibit agreement, EFA is the preferable technique as it is a less stringent test than CFA; if these measures are to converge, they are more likely to do so with EFA (Ford et al., 1986; Hinkin, 1998). We utilize the principle axis EFA to maximize the potential for convergence (Ford et al., 1986); the results are presented in Table 2. As our EFA indicates, extracting eigenvalues greater than

Table 2
Results of exploratory factor extraction.

Component	Initial Eigenvalues		
	Total	Percent of variance	Cumulative percent of variance
1	2.21	31.50	31.50
2	1.22	17.49	48.99
3	1.12	16.02	65.00
4	.88	12.62	77.62
5	.69	9.80	87.41
6	.54	7.69	95.11
7	.34	4.89	100.00

1.0 suggests a three-factor solution. The fact that we find a three-factor solution suggests that the seven measures of overconfidence do not converge on a common factor and thus are not measuring the same construct. If the measures were assessing the same construct, the EFA would result in a one-factor solution because theoretically each of the measures should assess the same underlying construct (overconfidence). Our finding suggests that while some of the measures may assess overconfidence, at least some of the measures employed are assessing something other than overconfidence, as is indicated by the failure of the measures to converge onto a single factor.

Cumulatively, the results of correlational and factor analyses suggest that existing unobtrusive measures of executive overconfidence do not exhibit adequate convergent validity. Inadequate convergence of measures used to represent the same construct has significant implications for existing research. Specifically, although an extensive literature has been developing regarding the effects of overconfident executives on the behavior of their organizations, the fact that the measures do not converge inhibits our ability to compare results across studies (Churchill, 1979; Nunnally, 1978) and further suggests that some studies may not be measuring the same construct of overconfidence. Our finding of inadequate convergence also has implications for the design of future research assessing executive overconfidence, specifically that researchers interested in utilizing unobtrusive measures to assess executive overconfidence first need to establish a valid instrument. With this need in mind, we next address whether any of the seven extant measures of executive overconfidence exhibit adequate content validity and thus may be viable for further use.

4.2. Content validity

A technique to assess the content validity of various measures is utilizing expert raters (Cronbach, 1971). Researchers employ both scholars (Hinkin, 1998; Schriesheim, Powers, Scandura, Gardiner, & Lankau, 1993) and practitioners (Hambrick, 1981; Subramaniam & Venkatraman, 2001) as expert raters to evaluate whether measures adequately assess an intended construct. We conducted three item-sort tasks using 15 independent scholars as raters for each sort task, for 45 raters. Scholars were recruited from the personal contacts of the researchers. At the time of the sort task, each of the raters who participated either: a) held a doctoral degree, or b) had been enrolled in a doctoral program for a minimum of one year and completed a seminar in which psychological constructs were a focal topic.

The first sort task asked these raters to classify measures into one of two categories that best represents the measure: CEO overconfidence or something other than CEO overconfidence. To prevent confusion regarding the intent of the sort task or the construct of interest, we provided a brief introduction of the task and a definition of the construct to each rater (Hinkin & Schriesheim, 1989; MacKenzie, Podsakoff, & Fetter,

1991; Schriesheim et al., 1993). The statement provided to each rater was as follows.

The purpose of this task is to determine whether each of the statements accurately reflects CEO overconfidence. CEO overconfidence is defined as the tendency of these individuals to inflate estimations of their own abilities. Please assign each of the following statements to one of two categories that you feel the statement best reflects: 1. Overconfidence − the statement best reflects CEO overconfidence; 2. Other − the statement best reflects something other than overconfidence.

To guard against the possibility of ordering biased responses, we randomly ordered statements presented to respondents (statements appear in Table 3). We used the second and third sort tasks to assess whether sorting varies by raters' interpretations of "overconfidence" rather than "hubris" or a broader term such as "self-confidence level." In the second sort task, raters were administered the exact same sort task as the first group with the exception that the term "overconfidence" was replaced by "hubris." In the third sort task, raters were administered a sort task with two changes: 1) the focal construct was labeled "CEO self-confidence level" rather than overconfidence, and 2) the definition provided was adjusted to "CEO self-confidence level is defined as the level at which these individuals estimate their own ability."

Results of the three sort tasks appear in Table 3. MacKenzie et al. (1991) suggest that measures that raters assign to the hypothesized category 80% or more of the time exhibit content validity since this threshold suggests that scholars agree with the notion that a measure adequately assesses the intended construct. None of the seven measures were deemed sufficient to assess the focal construct of the CEO (i.e., overconfidence, hubris, self-confidence level) using the 80% threshold. The highest score achieved just over 50% (0.53) and most achieved much less, including instances in which only one of the 45 raters surveyed thought the measures were an adequate reflection of the construct. Cumulatively, the results of our three sort tasks suggest that scholars do not believe that current unobtrusive measures adequately assess executive overconfidence. As such, these measures do not exhibit adequate content validity. The use of measures that lack content validity causes concerns as to whether this literature contains biased or even incorrect findings. That is, because researchers are utilizing measures that may not adequately assess the construct of interest, it is possible that inaccurate conclusions have been drawn from analyses generated using these measures. Despite researchers' best efforts to measure executive overconfidence then, the current literature may both contain and be building upon inaccurate conclusions.

5. Discussion

Validating measures of a construct is a continuous process of investigation and development. Our study provides evidence that scholars need to continue this process for unobtrusive measures of executive

Table 3
Expert analysis of measures using sort task.

Measure	Statement	Overconfidence	Hubris	Self-confidence level
1. Executive language use	First-person singular pronouns such as I, me, mine, and my are used in documents produced by the organization such as letters to shareholders and proxy statements.	.47	.47	.53
2. Media comments	Nationally prestigious publications with high circulation rates publish articles that are favorable of the CEO.	.13	.07	.13
3. Recent organizational performance	Over the previous few years, the organization in which the CEO is employed has had positive performance.	.00	.07	.00
4. Relative compensation	The CEO receives a high level of compensation relative to other executives in the organization.	.07	.00	.00
5. Organizational investment	Over the previous few years, the organization in which the CEO is employed has had a high rate of capital expenditures relative to other organizations.	.07	.13	.13
6. Stock option exercise	The CEO has stock options that have a positive value. These options could be exercised resulting in additional income for the CEO, but the CEO has yet to exercise them.	.47	.40	.47
7. Stock purchases	The CEO has recently purchased stock in his or her company.	.53	.53	.53

overconfidence. Specifically, we first find that extant measures do not exhibit convergence and hence are not measuring consistent constructs. We then find that scholars do not believe that extant measures exhibit adequate content validity and thus do not sufficiently assess the construct. As such, our results provide evidence confirming concerns regarding the validity of unobtrusive measures of executive overconfidence (Bollaert & Petit, 2010; Jin & Kothari, 2008), suggesting that the executive overconfidence literature is potentially compromised. That is, the lack of convergent and content validity of extant measures implies that our knowledge may be obfuscated by measures that "lack rigorous psychological and methodological grounding" (Hiller & Hambrick, 2005, p. 298). However, this does offer scholars a promising opportunity to pursue improved measurement practices in future investigations. To this end, we offer suggestions to aid researchers in improving the measurement of executive overconfidence. While we focus on concerns in this literature stream, our suggestions are relevant to a broad array of research streams given that many employ different measures of a single construct (cf. Hill et al., 2012).

5.1. Recommendations

5.1.1. Pursue validation of measures

Because the development of valid construct measures is an essential step for empirical research, our first recommendation is for scholars to pursue construct validation of measures as part of future inquiries. Ultimately, as Hinkin (1995, p. 971) notes in reviewing measure development, "a necessary prerequisite" in the measure development is "establishing a clear link between items and their theoretical domain." That is, we must begin with "a thorough understanding of the theoretical foundations" of the construct we are attempting to measure (Hinkin, 1998, p. 106).

In the case of executive overconfidence, our findings with respect to the lack of content validity of extant measures may mean that researchers moved too quickly to empirical research without first developing sound theory linking the measures to construct. The fact that executives are difficult to access may have led researchers to focus on convenient or coarse measures of the constructs as an initial step (Hambrick & Mason, 1984). While the focus on convenient and coarse measures is perhaps understandable, our findings suggest that we need to first return to theory and develop greater understanding of overconfidence and hubris prior to developing more fine-grained measures that can allow us build knowledge of these constructs. Are they the same, as they have been treated in the literature (Hill et al., 2012), or are they different and possibly malleable or developed over time, implying a syndrome or trait-like properties that in turn will impact measurement (cf. Chen, Gully, Whiteman, & Kilcullen, 2000)? In sum, further development is necessary with respect to the dimensionality and uniqueness of these constructs. To that end, we need to return to theory to understand the construct more clearly prior to developing a valid unobtrusive measure.

In moving toward measurement validation, we wish to note three points. First, scholars may benefit from utilizing validated psychometric instruments designed for use in direct assessments to help cross-validate unobtrusive measures as well. Both Busenitz and Barney (1997) and Simon and Houghton (2003) surveyed non-executive managers; the instruments utilized in their studies may be useful both in assessing executives and in cross-validating unobtrusive measures. However, validated instruments for direct assessments of overconfidence/hubris are also lacking (Hayward & Hambrick, 1997; Hiller & Hambrick, 2005). Thus, prior to cross-validating unobtrusive measures with instruments designed for direct assessment, scholars should ensure that these instruments exhibit adequate psychometric properties. That is, using measures with poor psychometric properties to cross-validate other measures does not improve measurement. Thus,

scholars should take heed with the selection of measures in cross-validation.

Second, alternative forms of investigation may compliment the current reliance on unobtrusive measures. Although gaining access to assess complex phenomena in the depth required to conduct alternative assessments can be problematic, the insight gained from various methods could add to our existing knowledge in ways that traditional large-scale samples analyzed with quantitative methodologies could not. Further, as with cross-validating unobtrusive measures with survey instruments, utilizing multiple methodologies can enhance the robustness of research by overcoming limitations associated with any one methodology or measure and inherent difficulties associated with assessing unobservable constructs (Campbell & Fiske, 1959; Venkatraman & Grant, 1986).

Third, every measure has strengths and limitations. Our finding that extant measures of overconfidence are not adequately assessing the construct as intended implies that the weaknesses of each measure may outweigh the strengths and as such that the measures, if they are not assessing overconfidence, are assessing something else. To clear up these issues, we conducted two surveys of faculty colleagues and doctoral students using the same statements and procedures as in the prior sort tasks. Insights with respect to the strengths and limitations of each measure were gathered from 15 individuals who participated in the prior sort tasks, while 15 individuals who did not participate in prior sort tasks were asked to state what they believe each measure *best* assesses. Table 4 highlights the aggregated responses from these surveys and the logic employed by the original author(s) to link the construct to the measure.

Thirteen of 15 respondents suggested that the strength of the measures is that they allow for assessment of constructs that may otherwise be hard to measure. In particular, unobtrusive approaches such as those reviewed may be a necessity in executive research because of difficulties with obtaining responses from executives (Finkelstein, Hambrick, & Cannella, 2009; Hambrick & Mason, 1984). With respect to the measurement of overconfidence/hubris specifically, the lack of rigorously validated instruments for direct measurement further contributes to the use of unobtrusive measures (Hayward & Hambrick, 1997; Hiller & Hambrick, 2005). These measures also have the benefits of enabling easier extension of studies longitudinally, taking advantage of the masses of available data and avoiding specific limitations of direct measurement such as nonresponse and social desirability (Finkelstein et al., 2009; Webb et al., 1966).

Limitations outlined include a number of factors that may cause measurement error, affecting both the degree to which the measures converge as well as whether the measures serve as valid indicators. For example, language use may contain errors associated with impression management. In addition, errors may be caused by the fact that company documents are typically at least partially ghost written and media comments may lack objectivity because of media bias or the perceptional biases associated with the author of the article portraying the executive. Likewise, respondents noted that measures might be capturing a variety of phenomena unrelated to overconfidence (organizational performance attributable to industry factors, stock option exercise and purchase decisions attributable to personal wealth management). The results generated by our survey may provide insight to scholars as they move towards the validation of measure. Similarly, while the results of our survey presented in Table 4 may be a starting point for determining what is actually measured by the seven unobtrusive measures utilized to assess overconfidence in extant studies, the literature would benefit from rigorous analyses to place these measures in the nomological network so that we know what they appear to be measuring and thus what conclusions can be drawn from studies that utilize those measures.

5.1.2. Revisit earlier studies

Given the results we find in our cumulative analyses, we also suggest that scholars revisit earlier studies in an effort to replicate or reinterpret

Table 4
Assessment of extant unobtrusive measures of overconfidence.

Measure	Logic for, and strengths of, measure	Limitations of measure	Statement best measures
1. Executive language use	• Word usage reflects speaker characteristics (Rovenpor, 1993) • Allows for indirect assessment	• May be ghost written and thus not assess psychology of the executive • Subject to impression management and contextual effects • Statements may reflect various constructs	• Desired posture or identity of firm at time of release • Various factors related to the style and psychology of the author(s) • Not confident it measures intended construct
2. Media comments	• Provides insight on how others perceive the individual (Malmendier & Tate, 2008) • May be self-reinforcing such that praise by media drives overconfidence (Hayward & Hambrick, 1997) • Allows for indirect assessment	• Lack of source objectivity (media bias, perceptional biases) • Subject to impression management and contextual effects • Limited information creates selection bias	• Various factors related to the perceptions of CEOs and/or their organizations in eyes of the author(s) and their publications' belief about what will sell articles/advertising • Not confident it measures intended construct
3. Recent organizational performance	• Organizational phenomena reflect and reinforce characteristics of the executive (Hayward & Hambrick, 1997) • Allows for indirect assessment	• Reflects variety of other factors • Limited information creates selection bias	• Market conditions • Whatever the performance measure assesses • Factors related to the measure utilized • Managerial ability
4. Organizational investment	• Organizational phenomena reflect and reinforce characteristics of the executive (Malmendier & Tate, 2005) • Allows for indirect assessment	• Reflects variety of other factors • Limited information creates selection bias	• Strategy or focus of firm • Market conditions • Not confident it measures anything of substance
5. Relative compensation	• Organizational phenomena reflect and reinforce characteristics of the executive (Hayward & Hambrick, 1997) • Allows for indirect assessment	• Reflects variety of other factors • Limited information creates selection bias	• Market conditions • Ability, value, and/or power differences • Various factors related to perceptions of those setting compensation
6. Stock option exercise	• Assesses actions directly attributable to CEO and provides insight into characteristics driving action (Malmendier & Tate, 2005, 2008) • Allows for indirect assessment	• Action may reflect various constructs or contextual factors • Limited information creates selection bias	• Various factors related to the psychology and personal situation of the CEO • Pressure from board and/or desire to signal • Not confident it measures anything of substance
7. Stock purchases	• Assesses actions directly attributable to CEO and provides insight into characteristics driving action (Malmendier & Tate, 2005, 2008) • Allows for indirect assessment	• Action may reflect various constructs or contextual factors • Limited information creates selection bias	• Various factors related to the psychology and personal situation of the CEO • Pressure from board and/or desire to signal • Not confident it measures anything of substance

findings. That is, since our findings suggest that measures employed to assess executive overconfidence are not assessing the same construct and are inadequate in doing so, we cannot be sure what we actually know. For instance, two studies reexamine the work of Amihud and Lev (1981) on divestiture and fail to find support for the conclusions drawn in the study; both cite measurement error as a possible cause of the discrepancies (Boyd, Gove, & Hitt, 2005; Lane, Cannella, & Lubatkin, 1998). While we are not alone in advocating such inquiries (Singh, Ang, & Leong, 2003; Tsang & Kwan, 1999), we are left to wonder whether current literatures that face measurement concerns similar to those of diversification and overconfidence research contain inaccurate conclusions. To both confirm that we really know what we think we know and to continue building our understanding of complex strategy phenomena it is essential that scholars revisit prior work with particular focus on areas that may be plagued by measurement concerns.

5.1.3. Employ methods to account for measurement error

Another avenue for scholars to pursue moving forward is to account for measurement error of unobservable constructs and assess reliability using multiple measures. One methodology that is amenable to accounting for measurement error is structural equation modeling (SEM). Although SEM is not widely used in strategy research, the technique is flexible and has the advantage of accounting for measurement error. (See Shook, Ketchen, Hult, & Kacmar, 2004, for insight on using SEM in strategy research.) Similarly, utilizing factor-analytic methods to obtain a composite measure can be beneficial in providing information on measure validity. An alternative approach to utilizing a single measure is to use multiple measures and compare results across them. Hayward and Hambrick (1997) provide an example of both methods. They develop three measures of executive overconfidence and include these in a separate model, assuming measurement without error, and interpret results across the measures. Further, they also develop

a composite measure utilizing factor analysis. Both multiple measures and CFA could be employed in future investigations as well.

Scholars wishing to utilize multiple measures and/or CFA should take caution, however, to ensure that the indicators included in the factor analysis are correctly specified as either formative or reflective. Because some measures rely on formative logic while others rely on reflective logic, it is foreseeable that a researcher could unintentionally mis-specify a factor-analytic model and thus bias the results (Kline, 2005). Further, researchers interested in utilizing formative measurement should consider whether this technique is appropriate in this context (cf. Diamantopoulous et al., 2008; Edwards, 2011). Additionally, theoretical work as to the nature of the constructs, as noted above, will direct scholars with respect to the appropriate measurement approach for conceptualizing the construct (i.e., dimensionality and uniqueness of these constructs).

5.1.4. Acknowledge limitations and alternative interpretations of measures

A fourth recommendation building upon measurement error is to carefully interpret results and acknowledge the limitations of, and alternative interpretations associated with, measures. Although scholars often are confronted with limited space with which to discuss their research, a more thorough discussion of limitations and alternative interpretations of the measures and subsequently of associated results will benefit the literature. Likewise, scholars should take care with research design to ensure that they define their samples in a way that eliminates as many confounds and alternative explanations as possible. For instance, accounting for size differences that may drive the occurrence of phenomena is an important step, as would be several other concerns with respect to sample selection that can drive measurement error (Cook & Campbell, 1979). Similarly, the summary offered in Table 4 may be a starting point for suggesting limitations and alternative interpretations of findings.

6. Conclusion

We find that existing unobtrusive measures of executive over-confidence lack both convergent and content validity. When existing measures of the same construct do not exhibit convergence, it is impossible to reliably interpret results across these measures because they do not exhibit agreement regarding what they assess (Churchill, 1979; Jacoby, 1978). Therefore, knowledge that has been accumulated using disparate measures that do not converge is not reliable. As a result, the executive overconfidence research stream may contain inaccurate conclusions that cannot be replicated in future investigations, and the foundation for the building of knowledge and understanding in these research streams will be weak. Our results also suggest that existing measures of executive overconfidence may not accurately reflect the construct of interest. That is, our assessment of content validity using scholarly raters suggests that these measures may reflect something other than the intended construct, and therefore our knowledge of executive overconfidence may not be valid. In sum, if measures do not agree and yet a literature is treating them as representations of the same constructor – or worse, these measures do not adequately assess the construct – the conclusions drawn from studies using these measures are not reliable. Put another way, our findings suggest that we are overconfident in our knowledge of executive overconfidence. The "knowledge" we are accumulating regarding the effect of overconfident executives is on a shaky foundation. In response, scholars need to revisit what is meant by overconfidence, develop valid measures of executive overconfidence, and revisit earlier studies that have utilized the measures tested herein to confirm that results are actually valid. Good scientific discipline requires this if knowledge and understanding are to be built on a solid foundation.

References

Amihud, Y., & Lev, B. (1981). Risk reduction as a managerial motive for conglomerate mergers. *Bell Journal of Economics, 12*, 605–617.
Bollaert, H., & Petit, V. (2010). Beyond the dark side of executive psychology, current research and new directions. *European Management Journal, 28*, 362–376.
Boyd, B., Gove, S., & Hitt, M. (2005). Consequences of measurement problems in strategic management research: The case of Amihud and Lev. *Strategic Management Journal, 26*, 367–375.
Brahma, S. (2009). Assessment of construct validity in management research: A structured guideline. *Journal of Management Research, 9*, 59–71.
Busenitz, L., & Barney, J. (1997). Differences between entrepreneurs and managers in large organizations: Biases and heuristics in strategic decision-making. *Journal of Business Venturing, 12*, 9–30.
Cairns, D. L. (1996). Hybris, dishonour, and thinking big. *Journal of Hellenic Studies, 116*, 1–32.
Cameron, K., Whetten, D., & Kim, M. (1987). Organizational dysfunctions of decline. *Academy of Management journal, 30*, 126–138.
Campbell, D., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56*, 81–105.
Campbell, T., Johnson, S., Rutherford, J., & Stanley, B. (2009). *CEO confidence and forced turnover.* Working paper. : Texas A&M University.
Chatterjee, S., & Blocher, J. (1992). Measurement of firm diversification: Is it robust? *Academy of Management Journal, 35*, 874–888.
Chen, G., Gully, S., Whiteman, J., & Kilcullen, R. (2000). Examination of relationships among trait-like individual differences, state-like individual differences, and learning performance. *Journal of Applied Psychology, 85*, 835–847.
Churchill, G. (1979). Paradigm for developing better measures of marketing constructs. *Journal of Marketing Research, 16*, 64–73.
Cook, T., & Campbell, D. (1979). *Quasi-experimentation, design and analysis issues for field settings.* Chicago: Rand McNally.
Cortina, J. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98–104.
Cronbach, L. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (pp. 443–507). Washington, DC: American Council on Education.
DeBondt, W., & Thaler, R. (1995). Financial decision-making in markets and firms: A behavioral perspective. In R. Jarrow (Ed.), *Handbooks in operations research and management* (pp. 385–410). Amsterdam: Elsevier.
Diamantopoulous, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research, 61*, 1203–1218.

Dickie, M. (1984). Hesychia and hybris in Pindar. In D. Gerber (Ed.), *Greek poetry and philosophy, studies in honour of Leonard Woodbury* (pp. 83–109). Atlanta: Scholars.
Edwards, J. (2011). The fallacy of formative measurement. *Organizational Research Methods, 14*, 370–388.
Finkelstein, S., Hambrick, D., & Cannella, A. (2009). *Strategic leadership, theory and research on executives, top management teams, and boards.* New York: Oxford.
Ford, J., MacCallum, R., & Tail, M. (1986). The applications of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology, 39*, 291–314.
Gans, H. (2005). *Deciding what's news, a study of CBS Evening News, NBC Nightly News, Newsweek, and Time.* Evanston, IL: Northwestern.
Hambrick, D. (1981). Environment, strategy and power within top management teams. *Administrative Science Quarterly, 26*, 253–275.
Hambrick, D., & Mason, P. (1984). Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review, 9*, 193–206.
Hayward, M., & Hambrick, D. (1997). Explaining the premiums paid for large acquisitions: Evidence of CEO hubris. *Ad Sci Quar, 42*, 103–127.
Hayward, M., Shepherd, D., & Griffin, D. (2006). A hubris theory of entrepreneurship. *Management Science, 52*, 160–172.
Hill, A., Kern, D., & White, M. (2012). Building understanding in strategy research: The importance of employing consistent terminology and convergent measures. *Strategic Organization, 10*, 187–200.
Hiller, N., & Hambrick, D. (2005). Conceptualizing executive hubris: The role of (hyper-) core self-evaluations in strategic decision-making. *Strategic Management Journal, 26*, 297–319.
Hinkin, T. (1995). A review of scale development practices in the study of organizations. *Journal of Management, 21*, 967–988.
Hinkin, T. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Org Res Met, 1*, 104–121.
Hinkin, T., & Schriesheim, C. (1989). Development and application of new scales to measure the French and Raven (1959) bases of social power. *Journal of Applied Psychology, 74*, 561–567.
Jacoby, J. (1978). Consumer research: A state of the art review. *Journal of Marketing, 42*, 87–96.
Jin, L., & Kothari, S. (2008). Effects of personal taxes on managers' decision to sell their stock. *Journal of Accounting and Economics, 46*, 23–46.
Kline, R. (2005). *Principles and practice of structural equation modeling.* New York: Guilford.
Lane, P., Cannella, A., & Lubatkin, M. (1998). Agency problems as antecedents to related mergers and diversation: Amihud and Lev reconsidered. *Strategic Management Journal, 19*, 555–578.
Li, J., & Tang, Y. (2010). CEO hubris and firm risk taking in China: The moderating role of managerial discretion. *Academy of Management Journal, 53*, 45–68.
Lubatkin, M., Merchange, H., & Srinivasan, N. (1993). Construct-validity of some unweighted product-count diversification measures. *Strategic Management Journal, 14*, 433–449.
MacKenzie, S., Podsakoff, P., & Fetter, R. (1991). Organizational citizenship behavior and objective productivity as determinants of managerial evaluations of salespersons performance. *Organizational Behavior and Human Decision Processes, 50*, 123–150.
Malmendier, U., & Tate, G. (2005). CEO overconfidence and corporate investment. *Journal of Finance, 60*, 2661–2700.
Malmendier, U., & Tate, G. (2008). Who makes acquisitions? CEO overconfidence and the markets reaction. *Journal of Financial Economics, 89*, 20–43.
Meehl, P. (1957). When shall we use our heads instead of the formula. *Journal of Counseling Psychology, 4*, 268–273.
Nunnally, J. (1978). *Psychometric theory.* New York: McGraw-Hill.
Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology, 29*, 261–265.
Roll, R. (1986). The hubris hypothesis of corporate takeovers. *Journal of Business, 59*, 197–216.
Rovenpor, J. (1993). The relationship between four personal characteristics of chief executive officers (CEOs) and company merger and acquisition activity (MAA). *Journal of Business and Psychology, 8*, 27–54.
Schriesheim, C., Powers, K., Scandura, T., Gardiner, C., & Lankau, M. (1993). Improving construct measurement in management research: Comments and a quantitative approach for assessing the theoretical content adequacy of paper-and-pencil survey-type instruments. *Journal of Management, 19*, 385–417.
Shook, C., Ketchen, D., Hult, G., & Kacmar, K. (2004). An assessment of the use of structural equation modeling in strategic management research. *Strategic Management Journal, 25*, 397–404.
Simon, M., & Houghton, S. M. (2003). The relationship between overconfidence and the introduction of risky products: Evidence from a field study. *Academy of Management Journal, 46*, 139–149.
Singh, K., Ang, S., & Leong, S. (2003). Increasing replication for knowledge accumulation in strategy research. *Journal of Management, 29*, 533–549.
Subramaniam, M., & Venkatraman, N. (2001). Determinants of transnational new product development capability: Testing the influence of transferring and deploying tacit overseas knowledge. *Strategic Management Journal, 22*, 359–378.
Tsang, E., & Kwan, K. (1999). Replication and theory development in organizational science: A critical realist perspective. *Academy of Management Review, 24*, 759–780.
Venkatraman, N., & Grant, J. (1986). Construct measurement in organizational strategy research: A critique and proposal. *Academy of Management Review, 11*, 71–87.
Webb, E., Campbell, D., Schwartz, R., & Sechrest, L. (1966). *Unobtrusive measures, nonreactive research in the social sciences.* Chicago: Rand McNally.