Measurement in Family Business Research: How Do We Measure Up?

Allison W. Pearson¹, and G.T. Lumpkin²

Perhaps the greatest difficulty in conducting research in organizations is assuring the accuracy of measurement of the constructs under examination.

—Hinkin (1998, p. 104)

The field of family business research is progressing forward rapidly as evidenced by both the quantity and quality of family business studies. In 2010, Family Business Review (FBR) received a record 232 submissions from authors in 33 countries (Sharma, 2011). FBR’s impact factor has experienced a meteoric rise from 2008 to 2011, increasing yearly from 0.675 to 1.357 to 1.881 to 2.426, respectively.¹ In her review of studies published in FBR in 2010, Sharma (2011) noted several interesting research trends that speak to the positive development of the field, including (a) research grounded in theory beyond agency or the resource based view; (b) the “domination of studies using quantitative methods” (p. 6), where a full 85% of the published articles were quantitative, including empirical case studies; and (c) a trend toward researchers developing and validating scales to differentiate between family-owned and family-managed firms (e.g., Holt, Rutherford, & Kuratko, 2010).

These and other trends in family business research led Moores (2009), based on Kuhn’s (1970) seven stages of evolution of a scientific discipline, to conclude that the family business domain has achieved “paradigm consensus” (Stage 2) because of “the emergence of a common body of belief” (p.168) and is now engaged in the work of Stage 3—“normal science.” In the remainder of his essay, Moores (2009) highlighted the importance of theory development to better explain the shared beliefs and insights developed in Stage 2. Along with articulating theory, the normal science stage also calls for experimentation, fact gathering, and what Kuhn labels “puzzle solving.” The purpose of this article, therefore, is to highlight the role of empirical testing and, in particular, precise and accurate measurement as a way to advance family business research.

Demonstrating, supporting, or revising theory helps us better understand the unique attributes and behaviors of family enterprises. Measurement of observed variables allows researchers to test theory. Therefore, valid and reliable measurement is essential to the development of the “normal science” of family firms and our field’s ability to contribute to knowledge development in the broader context of organizational studies.

So what does measurement consist of and why does it matter? Measurement of phenomena in social sciences such as family business research refers to assessing both observed and latent variables. Observable variables—objects such as number of people in a room or events such as number of trips to the doctor—can readily be measured. But many phenomena cannot be directly observed. For example, we cannot observe or directly measure family intentions for the firm, or altruism, or agency-related issues. Therefore, we rely instead on latent variables that are inferred from proxies or other observables that can be measured to serve as indicators. What such measures assess are constructs—that is, abstract ideas that cannot be directly observed. A construct is “a broad mental configuration of a given phenomenon” (Bacharach, 1989, p. 500) and includes “terms which though not observational either directly or indirectly, may be applied or even defined on the basis of observables” (Kaplan, 1964, p. 55). The majority of research in the social sciences, which includes organizational research, relies on constructs (Pedhazur &

¹Mississippi State University, Starkville, MS, USA
²Syracuse University, Syracuse, NY, USA

Corresponding Author:
Allison W. Pearson, College of Business and Industry, Mississippi State University, Starkville, MS 39762, USA
Email: allison.pearson@msstate.edu
Schmelkin, 1991). To test and further refine theory, it is necessary to find valid and reliable methods to measure constructs of interest.

Construct validity (how well a measurement device measures what it intends to measure) and reliability (how consistent responses are) then become the cornerstones of sound empirical research (Nunnally & Bernstein, 1994). However, as Hinkin (1998) notes, measurement of constructs is one of the greatest challenges of organization research. Without adequate construct validity and reliability in measures used in empirical studies, a variety of problems can and do arise—that can cause confusion and slow progress toward theory development. Without construct validity we are not measuring what we purport to measure. As a result, we run the risk of drawing conclusions about relationships that do not really exist. Furthermore, the use of poorly validated measures can result in research with contradictory findings. As Hinkin (1995) concludes, when we use poorly validated and unreliable measures, “often scholars are left with the uncomfortable and somewhat embarrassing realization that results are inconclusive and that very little may actually be known about a particular topic” (p. 967).

Based on guidelines of the American Psychological Association (1985) as well as those provided by measurement scholars (e.g., Nunnally & Bernstein, 1994), measures should demonstrate basic forms of validity and reliability prior to use in substantive theory testing (Schwab, 1980). Three basic forms of validity are content validity, construct validity, and criterion validity (Hinkin, 1998; McIntire & Miller, 2005). Content validity addresses whether or not the measure comes from a theoretical domain and is appropriately defined. Construct validity involves assessing the structural form of the measure and whether or not the scale items actually represent what was intended to be measured. Criterion validity includes both concurrent validity, which occurs when assessments correlate with measures that have previously been validated, and predictive validity, which is demonstrated when a measure predicts what it is theoretically related to, also referred to as the nomological network (Cronbach & Meehl, 1955). The reliability or internal consistency of measures assesses how well a set of measurement items holds together—that is, whether the items reflect the construct in a similar manner. Although not all forms of validity (e.g., convergent, discriminant, etc.) and reliability are necessary to demonstrate sound measurement, basic content and construct validity, as well as reliability, are considered essential. As our field of family business research progresses, we should see these basic forms of validity and reliability prominently featured in our published research. But do we?

To assess the current state of measurement in the field of family business research, we reviewed measurement techniques in the articles deemed most influential to our field (Chrisman, Kellermanns, Chan, & Liano, 2010), since they are more prominent, more widely read, and therefore serve as seminal pieces that will likely influence other research in the field. Using citation analysis to identify the most influential works in the field, 25 of the most cited family business studies were identified by Chrisman et al. (2010). Although these works do not necessarily represent exemplars of empirical research and may have been published during a period when conceptual development was more critical to the field than strong methods, they nevertheless represent important bellwethers in terms of assessing measurement in family business research.

Empirical pieces totaled 12 of the 25 articles identified as most influential. From those 12 empirical pieces, 8 relied solely on single-item measures and 10 provided no evidence of validity or reliability of the measures used. These results demonstrate an urgent need to pay greater attention to measurement issues if the field is to make scientific progress. Furthermore, in a recent survey (Litz, Pearson, & Litchfield, 2011), 80% of family business scholars surveyed reported that they believed the field had no understanding or limited understanding of measurement issues. These strong collective scholarly beliefs suggest that family enterprise scholars have an opportunity to develop rigorous scientific measures to capture the constructs of interest.

We might also ask, “Are these results unusual for an emerging field?” If we compare our findings with those of other fields in organizational science, the answer is clearly “no”. Emerging fields often struggle with similar measurement issues: Mintzberg (1977) criticized management research for being too prescriptive and lacking in sophistication of measurement. Schwab (1980) chas-tised organizational behavior researchers for weak construct development practices that sometimes led to contradictory results in theory testing. In reviews of strategy research, scholars have criticized the field for extreme reliance on nominal and single-item scales, as well as for inadequate assessments of reliability and validity; the researchers also concluded that most strategy research could not adequately satisfy the requirements for content validity, internal consistency,
convergent validity, discriminant validity, and/or criterion validity (Boyd, Gove, & Hitt, 2005; Venkatraman & Grant, 1986). The field of entrepreneurship has fared little better as it moved through its emergent stage of development (Busenitz et al., 2003). A thorough content analysis of published entrepreneurship studies led researchers to conclude that entrepreneurship has also relied too heavily on single-item measures whose validity and reliability cannot be assessed (Crook, Shook, Morris, & Madden, 2010). As an emerging field, family business research appears to be no different—we have opportunities to improve our measurement of abstract constructs and improve our science. What is the path forward to improve measurement in empirical studies?

To assess validity and reliability, measurement of a construct should consist of multi-item measures whenever possible. Although some phenomena can be assessed validly with single-item measures (e.g., observable variables), single-item measures can never be tested for reliability (Nunnally, 1978). Content validity is still required, yet often lacking, to demonstrate that the single-item measure is a valid proxy for the theoretical construct of interest. Furthermore, the single-item measurement often assumes that there is no measurement error and all variance is attributed to the construct of interest, leading to a reduction in power (for a full explanation, see Boyd et al., 2005).

There is excellent guidance available on how to create valid, reliable, multi-item measures of constructs of interest. Perhaps one of the most comprehensive and easy to follow explanations is Hinkin (1998). For developing multi-item measures, he recommends a multistep process, including (a) item generation based on theory and sound construct definition, (b) initial survey administration to pretest items, (c) item reduction to purify a measure, and (d) confirmatory factor analysis to demonstrate the construct validity of the measure; for more thoroughly developed scales, additional steps need to be taken: (e) convergent and discriminant validity tests to show that the measure is unique from other measures, yet performs similarly to related measures, and (f) replication to demonstrate that the measure holds up in other settings, other organizations, and other countries and cultures. Keeping these measures in mind, these steps should be completed prior to substantive theory testing and with unique data sets (Schwab, 1980).

In family business studies, researchers often borrow and adapt scales that are already established. When using existing multi-item measures, rather than developing new measures, the researcher should follow best practices established by Crook et al. (2010) and Hinkin (1995) to acceptable validity and reliability of the adapted or borrowed measures, prior to use in theory testing. These established steps for using existing measures include the following:

1. Provide clear construct definitions grounded in theory.
2. Provide the sources, items, or examples of items used to measure each construct.
3. Assess and report construct validity in the form of exploratory or confirmatory factor analysis.
4. Assess and report reliability of all measures used.
5. Use multiple data sources when appropriate (validation and theory testing should be conducted on separate samples or hold-out subsamples).
6. Adopt analysis techniques that measure error rather than assume there is none.
7. Provide a clear explanation of all study measures, validity, and reliability so future researchers can replicate and build on what is known.

In addressing the question of what else we can do to improve measurement in published research, Boyd et al. (2005) also challenged editors and reviewers to adopt rigorous standards of measurement. As the leading family-business-specific publication outlet, FBR has the responsibility of setting high standards for construct measurement. And increasingly, we are noticing the high expectations of FBR reviewers and editors on this front. Notable examples of family enterprise–related measurement include the development of the Family Climate Scales (Bjornberg & Nicholson, 2007), the adaptation of existing scales to measure the constructs of stewardship theory (Davis, Allen, & Hayes, 2010), and the creation of innovative measures of market orientation developed from content analysis of letters to shareholders (Zachary, McKenny, Short, & Payne, 2011). These studies are encouraging examples of how family business scholars are providing quality measurement, as well as meeting the call to give back to our sister disciplines (Zahra & Sharma, 2004). Additionally, we are a discipline with family business concepts that are in need of valid, reliable measures, for example,
familiness (Habbershon & Williams, 1999), socioemotional wealth (Gomez-Mejia, Haynes, Nunez-Nickel, Jacobson, & Moyano-Fuentes, 2007), and what it means to be a “business family” rather than family business (Moores, 2009).

Without progress in developing psychometrically sound constructs and measures, we risk the credibility of our field as a whole, for as Peter (1979) notes, “If the measures used in a discipline have not been demonstrated to have a high degree of validity, that discipline is not a science” (p. 6). As authors, reviewers, and editors, we need to commit to uphold the best practices of measurement in empirical studies in order to continue the rapid advancement of the science of family business research.

Acknowledgments
The authors thank Pascual Berrone, Justin Craig, David Gras, Rob Nason, Trish Reay, Carlo Salvato, Pramodita Sharma, and Jeremy Short for their helpful comments on an earlier draft of this article.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interests with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes
1. By comparison, the 2011 impact factors for other journals were the following: Academy of Management Journal, 5.25; Journal of Management, 3.747; Strategic Management Journal, 3.583; Entrepreneurship Theory and Practice, 2.272; and Journal of Business Venturing, 2.149. These impact factors indicate the successful progress of FBR, yet they also indicate additional room for advancement.
2. For a more detailed explanation of theoretical contributions in family business research, see Reay and Whetten (2011).

References


Bios
Allison W. Pearson is the Jim and Julia Rouse Professor of Management, Department of Management, Mississippi State University, Starkville, Mississippi, USA. Email: allison.pearson@msstate.edu

G.T. Lumpkin is the Chris J. Witting Chair of Entrepreneurship, Department of Entrepreneurship and Emerging Enterprises, Whitman School of Management, Syracuse University, Syracuse, New York, USA. Email: lumpkin@syr.edu