Entrepreneurship Research Journal

Manuscript 1005

Studying the Emergence of New Organizations: Entrepreneurship Research Design

Jennifer L. Woolley, *Santa Clara University*

Studying the Emergence of New Organizations: Entrepreneurship Research Design

Jennifer L. Woolley

Abstract

Entrepreneurship is a fundamental phenomenon in society around the world, but empirical work to improve its understanding has struggled to gain legitimacy. This article examines the challenge of establishing the field of research by comparing six recent studies on entrepreneurship research design and methods. Consistently, scholars have defined entrepreneurship as a phenomenon of emergence; however, most entrepreneurship research has focused on questions regarding new ventures characteristics and outcomes after a new venture is started. One reason scholars continue to struggle with origins and emergence questions is the difficulty in obtaining relevant data. This paper recommends a series of tools and strategies to gather and analyze holistic data on entrepreneurship and organizational emergence. Specifically, the use of multi-level longitudinal data from several sources by multiple collection methods can provide the rich context necessary to illuminate venture emergence. To inform future studies, examples from nanotechnology entrepreneurship research are provided.

KEYWORDS: research design, mixed methods, entrepreneurship, origins, new venture creation, organizational emergence

Author Notes: The author would like to thank Renee Rottner, Tammy Madsen, and Jennifer Steffel Johnson for their invaluable feedback.

Research in the area of entrepreneurship continues to gain vitality and legitimacy as more scholars enter the field and research methods improve. However, the rigor and the use of appropriate research methods remain in question (Short et al., 2010; Ireland, Reutzel, & Webb, 2005, Busenitz et al., 2003, Chandler & Low, 2001). In one of the earliest critiques of entrepreneurship research, Low and MacMillan (1988) found that studies excessively relied upon a single method, a narrow time frame, and one level of analysis. Low and MacMillan advocated for multi-method research across a broad time frame, using multiple, clearly specified methods while Aldrich and Waldinger (1990) called for longitudinal data from many levels of analysis and social contexts. Entrepreneurship research has continued to be criticized for a wide variety of reasons, from a lack of theoretical and conceptual framework (Shane & Venkataraman, 2000; Ireland, Webb, & Coombs, 2005), to a lack of analytical rigor (Chandler & Lyon 2001; Mullen, Budeva & Doney, 2009), to a lack of methodological variety (Davidsson, 2003; Ireland, Reutzel, & Webb, 2005). While there is evidence that the trend is improving, weaknesses remain. Crook et al. (2010) find that the experts identify the lack of fit between the research design and methods and measures used as the most common problem plaguing entrepreneurship research. Since research design is critical to obtaining results appropriate for the question asked (De Vaus, 2001), which is necessary for building theory and developing a field, robust research design is critical to the advancement of the entrepreneurship field.

One area of entrepreneurship work in which this is particularly clear is the study of organizational emergence, or the time during which an organization is founded and becomes viable. Given the embeddedness of economic structures (Granovetter 1985), the emergence of organizations is highly dependent on its contextual environment (Van de Ven & Garud, 1993; Aldrich & Martinez, 2000; Aldrich & Ruef, 2006). Scholars have struggled with capturing and assessing the precursors to organization and population emergence, especially contextually significant factors. Work studying firm emergence often does so using data from after the firm has been created, usually after incorporation. However, understanding organizational emergence provides insight into the core of entrepreneurship and emergence, as well as evolution of populations, technologies, and society, all of which influence the socio-economic health of our world.

The following examines research design limitations and opportunities specific to entrepreneurship researchers. I focus on the data requirements and methods that facilitate research on new organization emergence. The goal of this article is to provide techniques for creating a thorough assessment of emergence activities and actors. I specifically examine the difficulties in emergence studies and how one can design a study with related research questions in mind. Here, I

encourage the collection of longitudinal and contextual data in entrepreneurship research, attending in particular to data *prior* to the emergence of organizations. I argue that using multiple methods to collect longitudinal data from several sources improves the quality, validity, and robustness of entrepreneurship research. Doing so not only reduces single-source bias, but also improves the quality of research, avoids left-truncation of data, provides basis for the use of mixed methods and triangulation, and improves validity and confidence in findings. Building on methods used in other areas of management research and other disciplines, this paper offers tools, techniques, and examples to enable scholars to explore a broader range of research questions to cultivate more rigorous and substantial scholarship.

The article proceeds as follows. First, I look at the types and levels of data currently used in organization emergence and entrepreneurship research. Next, I look at ways to expand current research by using several sources of data from multiple levels to observe organizational emergence both at founding and the conditions prior to founding. I propose a range of tools that can improve research design and lead to more robust results and comparability among studies. Concrete examples from a study of nanotechnology emergence and entrepreneurship are provided throughout (Woolley, 2010; Woolley & Rottner, 2008; Woolley, 2007; Youtie & Shapira, 2008).

CURRENT ENTREPRENEURSHIP LITERATURE

Entrepreneurship is a relatively young area of study (Cooper, 2003; Short, Ketchen, Combs, and Ireland, 2010) and as such, has suffered from a learning curve around research methods. As the field has matured, this situation has improved, but many of the methodological criticisms remain valid (Crook et al., 2010). In the next section, I compare the findings of six recent studies that look at research design and methods over time: Chandler & Lyon (2001), Ireland, et al., (2005), Dean et al., (2007), Mullen et al., (2009), Crook et al., (2010), and my own survey. Table 1 summarizes the results of these studies. The results tell a story of how the field of entrepreneurship research has evolved over time and the weaknesses that remain.

TABLE 1
Summary Characteristics of Entrepreneurship Articles

	Updated	Crook, et al. 2010	Mullen, et al., 2009	Dean, et al., 2007	Ireland, et al., 2005	Chandler & Lyon, 2001
Empirical articles analyzed	47	238	478	592**	50	291
DATA SOURCES						
Primary data	51%	52%	57%		94%	75%
Secondary data	77%	48%	32%		58%	25%
Multiple source types	26%	7%				
Surveys	19%	51%	54%		48%	52%
Interview	23%	3%			40%	19%
Observations	4%				6%	1%
Experimental	4%	3%	3%			3%
Longitudinal	62%	31%				7%
Cross-section	38%	63%		86%		80%
Retrospective case				14%		13%
Not reported		9%				
LEVEL OF ANALYSIS						
Single level	77%					89%
Multiple level	23%			3%		11%
Individual	26%	26%		31%	20%	35%
Group / Team	0%	4%		4%	4%	4%
Innovation / Project						2%
Organization / Firm	60%	58%		60%	68%	53%
Industry	13%			1%	4%	9% (both)
Environment	19%			1%	4%	
Other	0	14%				
METHOD OF ANALYSIS						
Qualitative	15%		10%	40%	16%	19%
Quantitative	79%		90%	60%	90%	81%
Mixed	6%				6%	3%

* Derived; ** Differs from reported, derived here; NA= Not analyzed

Each study has its own idiosyncrasies, using different journals or years. Chandler & Lyon (2001) reviewed 291 entrepreneurship articles published between 1989 and 1999 in *Entrepreneurship Theory and Practice* (ETP), *Journal of Business Venturing* (JBV), *Strategic Management Journal* (SMJ), *Journal of Management* (JOM), *Academy of Management Journal* (AMJ), *Academy of Management Review* (AMR), *Organization Science* (OS), *Management Science* (MS), and *Administrative Science Quarterly* (ASQ). Ireland et al. (2005) reviewed 50 entrepreneurship articles published in AMJ between 1963 and 1999. Dean et al. (2007) reviewed a random sample of 592 entrepreneurship articles published in ETP (1976-2004) and JBV (1985-2004). Mullen et al. (2009) reviewed 478 entrepreneurship articles published in the *Journal of Small Business Management* (JSBM), ETP, and JBV from 2001 through 2008. Crook et al. (2010) reviewed all 182 entrepreneurship articles published in ETP and JBV from the years 2000-02 and 2005-07 as well as a sample from AMJ (12), ASQ (4), JOM (3), MS (13), OS (4), and SMJ (20) for a total of 238 articles. As each study used different keywords to identify entrepreneurship articles, one caveat of comparing the studies is the inherent distinctions in each. However, while the methods vary, samples overlap. Thus, general trends are examined here, not statistical variance.

To provide an addition perspective and evaluate progress, I also reviewed entrepreneurship articles published between 2005 and 2009 in top-tier management journals[1]. To select the sample and ensure the highest quality articles, I examined the journals with the lowest acceptance rates for all submissions and in which less than a third of articles were on entrepreneurship. By restricting the sample in this way, the articles observed have the highest standards of research, even in comparison with more established domains. Using four journals—AMJ, ASQ, OS, and SMJ—I analyzed 47 empirical articles.

Findings

The results of these studies show that in the last two decades, research design has changed. Overall, entrepreneurship work has shifted from primary data (mainly surveys) to secondary data sources. Specifically, the use of secondary data sources (mainly archival data) shifted from less than half of the studies published through 1999 to over three-quarters of the studies less than ten years later. In contrast, I found that less than 20 percent of the entrepreneurship articles published between 2005 and 2009 in the top four management journals use surveys. The use of observations and experiments remains low. Surprisingly, the

[1] ISI Web of Science database search with the topics and keywords limited to "entrepren*" and "new venture*" and publication between 2005-2009. These keywords were chosen to identify only articles specifically related to entrepreneurship.

proportion of qualitative studies has decreased slightly, but this may be due to the increase in mixed-method pieces. However, the percentage of mixed method studies remains low (six percent). Overall, the percentage of multi-level studies has increased and there is a greater proportion of articles that include macro-environment data (industry and environment).

However, three problems remain, especially for studies on emergence. First, organizational research on entrepreneurship tends to use only one data collection technique and one analytical method. As summarized in Table 1, a minority of the recent entrepreneurship studies use data collected through more than one technique (26 percent) or use mixed methods of analysis (six percent). Using a single source, technique, or method introduces the possibility of bias and may reduce the validity of the data (Singleton & Straits, 2005).

Second, entrepreneurship research remains focused on the firm or organizational level of analysis (86 percent), to the neglect of industry, field, or environmental levels. And while some studies include more than one level (23 percent), two-thirds of these were conducted at the individual-firm levels. The context beyond the immediate firm, such as competition and industry, rarely is an object of independent consideration and merely appears as a control variable in most studies. These findings echo the earlier studies that argued that entrepreneurship studies in management were increasingly stripping away the context of the study and focusing on either the entrepreneur or the firm to the exclusion of other levels of analysis (Davidsson & Wiklund 2001; Zahra, 2007). Ireland and Webb (2007) warn that "an outcome of these highly and tightly focused yet contextually specific research designs is that scholars tend to ignore theoretical and methodological insights that are embedded in other disciplines when completing their entrepreneurship-related work" (p. 892). Consequently, if entrepreneurship research in organization and management journals does not expand into other methods and sources of data, theoretical progress will be limited.

Third, I find that the proportion of longitudinal studies has substantially increased (from seven percent to about 68 percent); however, half of these papers were left-truncated (48 percent). While several studies used datasets that included the beginnings of the firm, only a few contained data from *before* the founding of the organizations (10 percent) and only one before the founding of an industry. If entrepreneurship is about emergence—whether of new organizations or economic activity (Davidsson & Wicklund, 2001), scholarship and theory can only advance through tracing that emergence. This necessitates using data that capture the emergence process over time from the beginning. While this may appear obvious, the lack of longitudinal studies that include pre-founding data indicates that entrepreneurship research has not adequately addressed the questions of emergence or creation mechanisms. To advance the study of entrepreneurship,

we must design research studies to overcome these three limitations. It is to this opportunity which I now turn.

DESIGNING RESEARCH WITH EXTENSIVE LONGITUDINAL DATA AND MULTIPLE METHODS

To answer the call for richer studies of entrepreneurship, scholars must heed the advice of Low and MacMillian (1988), Aldrich and Baker (1997), Aldrich and Martinez (2000), among others, and design studies that give maximal views of the phenomenon. One way to accomplish this is to use multiple methods to collect several sources of longitudinal data at more than one level. A plurality of method and data is desirable when examining complex systems and processes (Jick, 1979; Klein & Kozlowski, 2000; Yauch & Steudel, 2003; Singleton & Straits, 2005). Given the complex contextual environment of organizations, multi-level, longitudinal data are necessary. Below, tools and techniques to achieve these ambitions are offered. I start by discussing optimal characteristics of data to examine entrepreneurship, where we might find them, and how we might analyze them to arrive at a rich synthesis.

Data Requirements

To answer research questions about the origins and emergence of organizations requires data with specific characteristics. In addition to the customary requirements for validity and reliability, the entrepreneurship researcher must consider three factors in the selection of a research design: the data must be *longitudinal*, start *before* the organization's emergence, and include the *socio-economic context*. Due to the time involved in gathering and studying this amount of data, these requirements are difficult to fulfill. This may explain why published works rarely use such an extensive data collection process, and why very little research on the origin or creation of organizations or industries has been published. However, it is only through data that are longitudinal, prior to emergence, and contextual that we can address questions regarding organizational emergence.

First, research on organizational emergence necessitates longitudinal data as entrepreneurship does not take place in one static instant but is a process that occurs over time. It is not possible to gain a full understanding of a process by observing a cross section of the event as this only provides a snapshot of a phenomenon in the middle of change. Rather, researching a process requires the direct or indirect observation of multiple stages during its progression. Van de Ven and Garud (1994) argue that "social evolutionary processes are better viewed as a cumulative progression of numerous interrelated acts of variation, selection,

and retention over an extended period of time" (p. 427). Thus, to study emergence in entrepreneurship, the researcher must locate reliable longitudinal data, whether archival or through direct observation of the process over time.

Second, the study of organizational emergence requires data on the events and factors that take place during and *before* founding. This may seem counterintuitive, as how can we gather data from before something exists? The simple response is that researchers must gather contextual data to better understand the conditions under which the entity (be it organization, organizational form, industry, community, field, etc.) emerges. Entrepreneurship research has been criticized for using left-truncated data (Gartner, 2001), resulting in a body of literature focused on successful entrepreneurs and existing organizations that does not capture organizational origins. As seen above, this problem continues. Effective examination of organizational emergence requires longitudinal data of the phenomenon from its earliest days since this is the time frame of interest. As stated, to better understand organizational emergence, a thorough understanding of the context in which the organization is arising is necessary. Temporally, this means a solid comprehension of the socio-economic-institutional context at the time of an organization's origins and founding. This can only occur with data from before the founding of the organization itself.

Building on this, to understand the emergence of an economic structure such as an organization or firm, one must gather data not only on the organization itself, but also on the context in which it arises. Organizations are economic actors embedded in a social context (Granovetter, 1985; Stinchcombe, 1965). At the simplest level, every organization has customers and suppliers with which it must interact. Furthermore, organizations emerge under different social conditions in which financiers, regulators, and voters can offer or withhold their support for that organization or industry. Without examining the social, economic, and institutional contexts in which an organization is situated, one cannot adequately understand the influences on entrepreneurship.

To fulfill these requirements, researchers should use a variety of methods to collect several sources of longitudinal data from more than one level. It is important to use data from a variety of sources as it allows a researcher to gain "distance" from the phenomenon to maintain objectivity and prevent single-informant bias (Strauss & Corbin, 1998: 44). The use of multiple data sources also improves the validity of the results as this allows for triangulation, or the inclusion of two or more dissimilar data sources or collection instruments that do not have the same methodological weaknesses and strengths (Jick, 1979). Triangulation enables the researcher to compare and contrast findings, thereby improving validity. By utilizing different data sources and collection methods, the confidence and strength of results increases (Singleton & Straits, 2005). Additionally, more than one level of data provides greater insights into context

and a richer understanding than one can obtain from a single level of data (Jick, 1979). In the next section, I review the data collection methods, data sources, and analytical tools by which triangulation can be achieved in entrepreneurship research.

Data Collection Methods

Collecting data from several sources can be accomplished by using one of three techniques. First, the researcher can focus on a single data collection method, but collect it from a variety of distinct sources. This ensures that diverse perspectives are considered. Second, the researcher can use multiple types of data collection methods, but focus on one source. This technique allows a more in-depth examination of a single setting and greater reliability. And third, the researcher can use multiple data collection methods from several sources. I offer examples on each of these methods from recent research.

Multiple sources, one type of data collection method

Researchers can use one type of data collection method (archival, interview, survey, etc.) from several sources (e.g., the top management team of a firm, their venture capital team, industry analysts, and the research and development department). For example, I use several sources of archival data in the study of nanotechnology entrepreneurship as illustrated in Table 2. Each of these sources provides a distinct perspective on new ventures using nanotechnology and is valuable to understanding venture origins. Consider the funding of new firms. By using multiple data sources, I found that both government agencies and venture capital firms are important funding sources, but government agencies started to fund nanotechnology firms more than five years earlier than venture capital firms in the U.S., and in other countries, this lag was almost ten years. Collecting only venture capital firm data would thus overlook the dynamics of support structures for the earliest firms. Additionally, venture capital firms tend to invest in firms that have some track record of viability and success. Using only venture capital data would provide a view of only those firms existing (success bias) and those seeking venture capital funding (size bias). By collecting data from both funding sources, a fuller view of nanotechnology entrepreneurship emerges.

TABLE 2
Archival Data in a Nanotechnology Emergence Study by Level of Analysis and Source

Level of Analysis	Organizations	Sources	Types	Examples
Environment/ Field	Government Agencies	Funding sources, Initiatives	Calls for proposals, research reports, grants	National Nano-technology Initiative Annual Report, Grant calls,
		Regulation committees	Meeting memos, minutes	EPA memos
		Standardization boards	Reports, meeting summaries	NNI reports: NNI Initiative and its Implementation Plan; NIST reports
	Other Resource Providers	Venture capital firms & Investment banks	Funding announcements, firm listings, reports	Harris & Harris nanotechology funding list
	Educational Facilities	Institutes	New venture competitions	Berkeley Nano Opportunity Challenge
		Technology transfer offices	Directories, meetings, reports	Caltech's Office of Technology Transfer firm listings
	Third Party Research	Research firms & Consultants	Reports, directories	Reports by Lux Research
		Media firms	Reports, new articles, press releases	Reports by NanoTechWire
Market / Industry	Competitors, Suppliers, & Customers	Directly and indirectly related firms	SEC documents, press releases, etc.	Veeco SEC documents
	Associations & Institutes	Commercial associations	Directories	NanoBusiness Alliance directory
		Scientific associations & Technical groups	Articles, reports, memos, conferences, membership	IEEE Nanotechnology Council reports
		Non-profit institutes	Reports	Foresight Institute
Organization	Focal Firms	New ventures	Internal documentation, press releases, website data	Zyvex press releases
Team	All of the above	Top management teams, R&D teams, product development	Meeting notes, Group reports, email conversations	Top management team of nanotechnology firm
Individual	All of the above	CEO's, Employees, etc.	Published interviews, biographies	Corporate website biography, CEO resumes

The value of the collection of several sources of data is strengthened when the data include more than one level of analysis—such as individuals, groups, organizations or institutions—since each level offers a unique perspective. Multilevel research is "complex, rigorous, and able to capture much of the nested complexity of real organizational life" (Klein and Kozlowski, 2000: 211). Not only do data from more than one level allow the researcher to formulate a more comprehensive understanding of new venture creation, but multiple levels of analysis provide insight into the context of and interactions between new ventures. Without examining an organization's social structure, one cannot adequately understand its embeddedness. This is particularly apparent when examining new venture origins.

Data from more than one level are also useful for identifying factors that influence the formation of new organizations and industries before their emergence. Baum and Haveman (1997) have argued that "organizational attributes cannot be used as explanatory variables in analyses of founding because they cannot be observed for organizations that do not yet exist" (p. 304). Thus, it is important to collect data from a range of levels to appropriately capture the dynamics of the emergence process. For example, scholars will find that in the quest for understanding the emergence of new firms, they must first gain an understanding of the firm's industry, market, and field. In the case of Woolley (2007), environmental-level archival data from technical groups that held conferences about the use of nanotechnology and non-profit institutes that produced reports concerning the application and safety of nanotechnology provided a foundation of insider information about the technology and field participants. Table 2 provides specific examples of data sources from each level of analysis. By starting with environmental-level data, the researcher can gain insight into multiple stakeholders in firm emergence that would not have been possible using data from the firms themselves.

Multiple collection methods, one or more data sources

The second and third techniques emphasize the strengths of using of multiple data collection methods, which increase the possibilities for triangulation. With the second technique, the researcher uses multiple types of data collection methods, but focuses on one source (e.g. archival, survey, and interview data from a single organization). This technique is most appropriate when a detailed account of an internal process is sought, such as case studies regarding organizational creation (Eisenhardt, 1989). This approach focuses the study on the mechanisms of the focal entity, but provides opportunity to gain insight not available from one data collection method since each provides a distinct perspective (Table 2). However,

this technique generally limits the researcher to one level of data, thus limiting the contextual insight that may be useful to understanding the phenomenon.

The third technique is a combination of the first two: the researcher collects multiple types of data from several sources for a more robust representation. This technique is the most comprehensive, but the most arduous. As such, I recommend that the researcher break this technique into steps based on data collection method or source. For example, one could break the data collection into steps by type of data and source starting with the most pertinent to the research question.[2] After gathering data from each major step, it is important to take stock of what has been made clearer as well as the questions that remain. Such reflection provides guidance for the next step. More often than not, new questions will arise unexpectedly that are critical for understanding of the overall research question. Additionally, by reflecting on the data collected and conducting iterative analysis, the researcher can gain a sense of data saturation and limit the amount of time and energy spent collecting redundant data.

In the study of nanotechnology entrepreneurship introduced above, observation and interview data supplemented the archival data. Observations of nanotechnology's emergence came in the form of conference attendance. I visited eleven conferences related to nanotechnology over five years. The conferences varied from highly technical and scientific to commercial applications. I also conducted semi-structured interviews with community experts. I chose representatives from the different levels of the nanotechnology field, including professors of nanotechnology; founders, CEOs, and employees of nanotechnology firms; current and former government employees directly involved in nanotechnology policy; and consultants specializing in nanotechnology applications and commercialization. The interviews started with a set of open-ended questions and progressed to free dialogue. Interviews lasted between 20 minutes and three hours and covered the topics of entrepreneurship, nanoscience and nanotechnology creation, commercialization, and contemporary activities. When permitted, the interviews were digitally recorded and detailed notes were made. Overall, more than 300 pages of interview notes and transcriptions were produced. Observational and interview data provided contextual information on the new nanotechnology organizations, as well as the critical actors in their social environment. Data from the firms alone could not have provided such a holistic picture.

[2] For example, if the research question focused on the strategic role of founders during organization emergence, one could start with gathering archival data about each founder and the firm itself. Next, one could interview each founder, and then move on to interviews of other present at the organization during founding. Third could be interviews of stakeholders outside of the organization such as competitors, funding sources, industry experts and the like. These steps can occur simultaneously as logical.

Longitudinal data can be collected for each data type. Due to the vast amounts of documentation that organizations produce, longitudinal archival data are one of the easiest to collect (depending on the age of the firm and recording technology available at the time of founding). Surveys and interviews can also be administered over time and replication is not difficult. One difficulty is the change of personnel over time that may influence the size and strength of your sample. Another difficulty is obtaining access to employees at multiple times without being disruptive to the organization. As a researcher must actively collect observational and experimental data, these tend to be more difficult to collect over a long period of time. Nevertheless, it is worthwhile for researchers to engage in multiple iterations of all data collection types, regardless of time, as otherwise understanding will be limited. Additionally, those studies with a richer depth of data contribute more and have a larger impact on the field.

Analytical Tools

Data collection from several sources from more than one level often requires multiple methods of analysis. The key to using such an extensive data collection technique is to apply the proper analysis method for the data and the research questions and then triangulate the findings. This section illustrates the application of various analytical tools using concrete examples from research on the emergence of nanotechnology entrepreneurship.

The most commonly used methods of analysis and the types of data for which they are appropriate are summarized in Table 3. The first two methods, historical analysis and case studies, are typically considered qualitative methods while the remaining five are considered more quantitative methods. It is important to note that interviews, surveys, and archival data can be both qualitative and quantitative and contribute to both types of analyses. Qualitative and quantitative methods can be used together to produce more robust results than using one type alone (Yauch and Steudel, 2003).

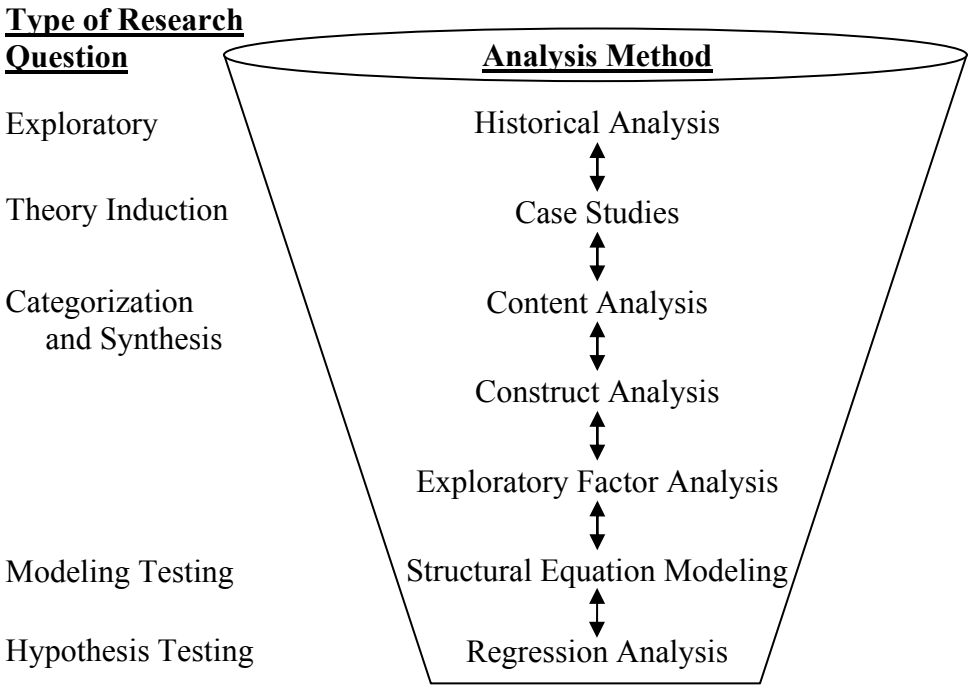
In the review of entrepreneurship work described earlier in this paper, I examined the types of analysis methods used and the type of research question asked. By referencing methods work (Singleton and Straits, 2005; Pedhazur and Schmelkin, 1991), I organized the results according to scope of research question and method[3]. The results are summarized in Figure 1. The methods are depicted in a funnel, filtering research questions from the most exploratory to the most specific.

[3] Although there was no evidence of structural equation modeling, the method of analysis was included in the figure due to its reference and its potential use. I refer the reader to Kelloway (1998) and Shook, Ketchen, Hult, & Kacmar (2004) for a more complete analysis of this method.

TABLE 3
Analytical Tools and Data Collection Methods in Organizational Research

Data Collection Methods							
Analytic Tools	**Observations**	**Interviews**	**Surveys**	**Archival**	**Experiments**	**Uses**	**Resources**
Historical Analysis	X	X	X	X		Build context understanding	Ventresca and Mohr, 2002
Case Studies	X	X	X	X		Induce theory	Eisenhardt, 1989; Yin, 2003
Content Analysis		X	X	X		Reduce data into categories	Singleton & Straits, 2005
Construct Analysis		X	X	X		Identify patterns	Lee, 1999; Miles & Huberman, 1984
Exploratory Factor Analysis		X	X	X		Detect underlying structure	Suhr, 2003; Kline, 1998
Structural Equation Modeling		X	X	X		Model testing	Kline, 1998
Regressions		X	X	X	X	Test hypotheses	Greene, 2008

FIGURE 1.
Research Question Types and Analysis Interaction



At the beginning of any research agenda, a researcher should use a wide range of methods and data to hold the convergent and divergent viewpoints, as well as the changing context over time. Starting with more exploratory methods (at the top of the funnel) allows the researcher to peer into the phenomenon of interest from a higher level than the focal level of analysis. This initial exploration may investigate how the phenomenon resides in its context, its relationship with other social and economic actors, or its relative importance among other economic and social phenomena. Once a theoretical framework has been developed, more specific research questions can be approached using more narrow methods. Thus, at the beginning of a research project, it is useful for the researcher to use exploratory methods before examining more fine-grained research questions.

It is important to note that there is not a linear relationship between research methods and the specific order of use. Upon entering the research setting, often the order of research collection and analysis is recursive and the examination of one question opens several others. Similar to the value of

reflection after each step of data collection, each type of analysis can inform others in refining or cultivating more research questions. It is the interaction of the analyses that provides the most rigorous investigation of the phenomenon.

In the study of nanotechnology entrepreneurship emergence, I collected observational, archival, and interview data that were analyzed using six of the seven methods shown in Figure 1. The following section presents those six methods of data analysis from most exploratory to most specific research questions: 1) historical analysis 2) case studies 3) content analysis, 4) construct analysis, 5) exploratory factor analysis, and 6) regressions. However, as mentioned earlier, these methods need not be followed in a particular linear pattern. Also, please note that there was a great deal of recursiveness in this agenda as many techniques were used multiple times. For instance, a content analysis will provide additional insight into case studies and historical analyses. In fact, as the research progresses to more specific questions, the researcher understanding of broader questions increases. The following is an in-depth description of this multilevel, multi-source data collection and analysis process.

Historical analysis

Historical analysis is the in-depth examination of data to produce a rich description of a phenomenon (Ventresca & Mohr, 2002). Historical analysis is a rigorous analytic method relying on formal methodologies, not biased storytelling. Ventresca and Mohr (2002) explain that historical analysis relies on formal analytic methodology, focuses on the social context rather than the organizations themselves, examines relationships rather than attributes, studies social processes of shared forms of meaning, and explores configuration logics that tie together organized activity. Unfortunately, few organization researchers have used historical analysis to examine entrepreneurship-related research questions. Scholars using historical analysis notably include Leblebici, Salancik, Copay, and King (1991), Padgett and Ansell (1993), and Hargadon and Douglas (2001). (See Ventresca and Mohr, 2002 for a more complete review.)

Historical analysis enabled me to explore the interaction between the events and actors leading to the creation of nanotechnology entrepreneurship. The outcome was a complex and robust description of the creation of the nanotechnology community, starting with the conceptualization of its fundamental science and technology through the commercialization of products, founding of firms, and emergence of industries.

Case studies

The next phase of research digs deeper than the historical analysis to create case studies for firms. Several excellent guides detail the case study process including Eisenhardt (1989) and Yin (2003). For the building of case studies for the emergence of nanotechnology firms, I designed a five-case analysis of the first companies founded to exploit the new technology. Using archival and interview data, I followed Eisenhardt's detailed description of a step-wise approach to case analysis.

Content analysis

Next, I used content analysis to refine my understanding of nanotechnologies history and the emergence of firms. Content analysis is a set of methods used to systematically reduce large amounts of data into categories enabling the researcher to more easily identify sequences, patterns, and relationships (Singleton & Straits, 2005). This is especially useful when the data are mainly textual. In content analysis, data preparation is time consuming but essential. To analyze the nanotechnology data, I first reduced over 9,000 pages of archival data and 29 interviews into a manageable form. To do so, I created a database of every textual account related to the research question. This can be done using data mining software, textual coding software, or simple brute force (i.e., copy and paste into a spreadsheet). To enhance the reliability of this data, a second researcher also coded the information in the database. Intercoder reliability is the level of agreement between two data coders, using the same instrument (Singleton & Straits, 2005). To measure intercoder reliability, two coders separately compiled lists of events and constructs from one data source and the two lists were then compared for agreement. The agreed-upon list is used for subsequent research.

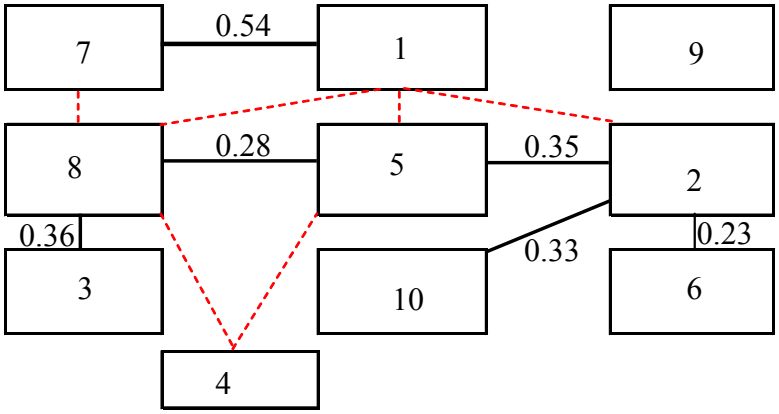
From the content analysis, I identified patterns of activity leading up to the founding of the first nanotechnology firm, such as the technical and scientific advancements of incumbent firms that were researching nanotechnology. I also was able to determine which institutions were active in the funding of nanotechnology research (mainly large governmental organizations in the U.S.).

Construct analysis

To help identify patterns and themes in qualitative data, Miles and Huberman (1984) suggest consolidating several variables into a smaller number of meaningful groups. To consolidate, I coded each of the events in the database using a pre-determined, theoretically driven list of constructs. I formed these constructs by analyzing core concepts in the data and grouping them by theoretical commonality (Lee, 1999). Next, I conducted a construct analysis (Miles & Huberman, 1984) using a correlation matrix of the events and their coding. This matrix identifies which of the constructs are highly correlated with one another and the level of significance. Correlations between categories indicate the extent of the convergent validity (Lee, 1999). Next, I created a concept-indicator model (Lee, 1999), which is a map of constructs representing their relative correlations. Each construct was represented by a box with lines drawn between constructs to indicate significantly positive or negative correlations (solid and dashed line respectively, see Figure 2). The value of positive correlations is depicted adjacent to the line connecting the constructs. I repositioned the constructs so that highly correlated constructs are closer together and negatively correlated constructs are separated. This analysis creates groups that emerge from the data that may have been missed using content or historical analysis alone.

The construct analysis indicated two clusters of variables. Within each group the variables were significantly correlated, but between groups they were negatively correlated. This indicated that the two groups were distinct and important for further analysis. Specifically, one cluster included technological innovation and incumbent firms but excluded entrepreneurial activities and universities. Additionally, two other variables were outliers in that they did not have a significant positive correlation with any other variable and in one case, was negatively correlated with two variables (entrepreneurial activity and government and resource endowments).

FIGURE 2.
Example of Construct Correlation Mapping



Exploratory factor analysis

To detect any underlying relationships between variables, I performed an exploratory factor analysis, a data reduction method that identifies the number of factors (latent variables) that effectively represent the data (Kline, 1998). A factor or latent variable is an unobserved variable that is not measured directly by observed variables (Kelloway, 1998). Observed variables are considered to be linear combinations of latent variables (Suhr, 2003)[4]. Exploratory factor analysis finds the small number of factors that linearly reconstruct the observed variables (STATA Press, 2001). Each observed variable is related to or "loads onto" each factor and the factor loadings are the correlation between a variable and a factor. (Kline (1998) has a more thorough discussion of exploratory factor analysis.) The benefit of exploratory factor analysis is that it allows the researcher to statistically examine relationships between variables that often indicate structures that would have otherwise been missed by the researcher. These relationships can help lead to hypothesis building and testing using regression analysis or other appropriate techniques.

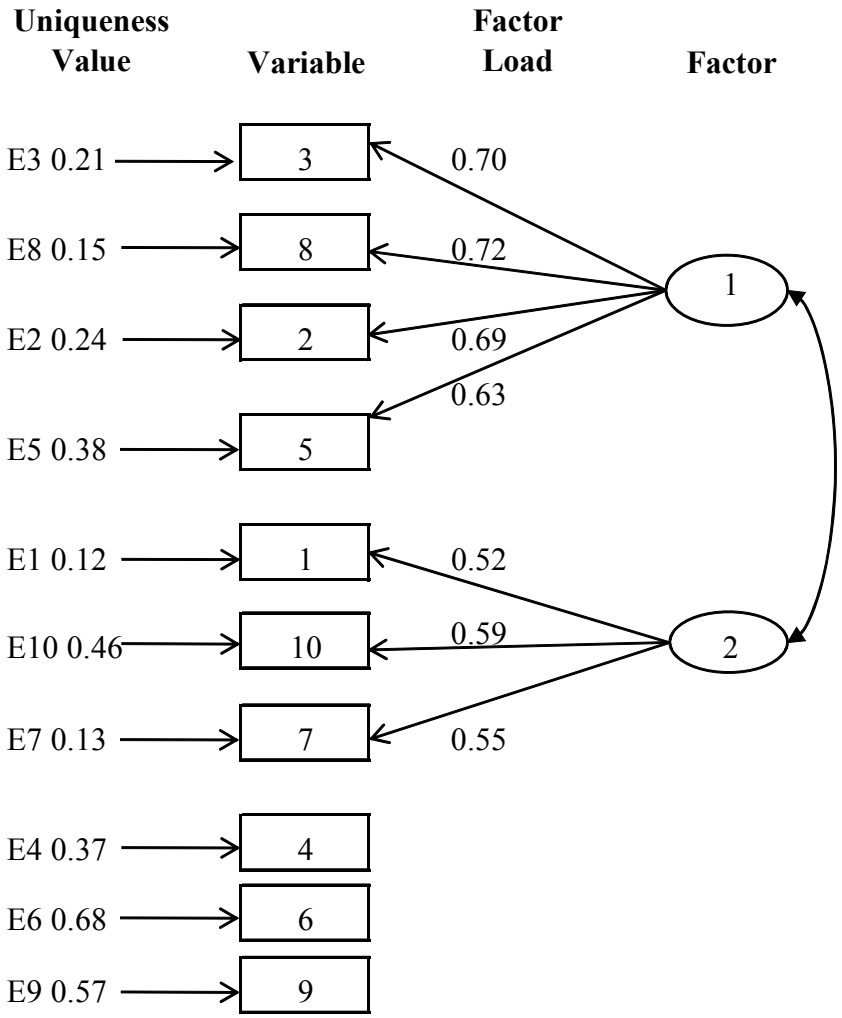
[4] For example, in an interview if the observed variables (questions) are measuring a similar construct, they would all be related to (or load onto) the same latent variable.

Using the nanotechnology data, I ran an exploratory factor analysis to determine underlying relationships between variables. The analysis of all four decades of data indicated no underlying structure as each variable loaded onto different factors. Reflecting on the findings from the historical and construct analyses, I ran the exploratory factor analysis using only the first phase of nanotechnology emergence (from the earliest data through the founding of the first five firms). This analysis indicated an underlying structure similar to the clusters of variables found in the construct analysis (see Table 4 and Figure 3). In the exploratory factor analysis, one factor had the highest loading for four variables, another factor had the highest loading for three variables, and three variables highly loaded on individual factors. Thus, two factors were retained. By integrating the findings from both analyses, patterns of activity over time emerged such as the integral role of incumbent firms in early technological innovations.

Table 4.
Example of Exploratory Factor Analysis Results with Factor Loadings for Each Variable

	Factors						**Uniqueness**
Variable	**1**	**2**	**3**	**4**	**5**	**6**	
1	-0.76	0.52	-0.02	0.06	0.02	0.16	0.11
2	0.69	0.20	0.41	0.21	0.17	0.06	0.24
3	0.70	0.50	-0.12	-0.13	-0.02	0.10	0.21
4	0.02	-0.25	0.73	-0.11	0.07	0.13	0.37
5	0.63	-0.02	-0.27	0.37	-0.13	0.04	0.38
6	0.24	0.06	0.31	0.32	-0.25	-0.04	0.68
7	-0.70	0.55	0.09	0.10	-0.23	0.04	0.12
8	0.72	0.45	-0.14	-0.31	-0.10	0.04	0.15
9	-0.06	-0.42	-0.41	0.16	0.14	0.21	0.57
10	-0.09	0.59	-0.03	0.18	0.38	-0.11	0.46

FIGURE 3.
Example of Results from Exploratory Factor Analysis



Regression analysis

There are many types of regression analysis techniques and the topic goes beyond the purview of this article. Regression analysis is typically used for hypotheses testing to examine very specific questions. In the nanotechnology study, I developed hypotheses using earlier findings in the content, historical, construct, and exploratory factor analyses. I tested these hypotheses using regression

analyses. To model the relationships between actors in the nanotechnology community, particularly government agencies, industries, and nanotechnology firms, I used longitudinal, pooled cross-sectional data for each year from 1984 (before the founding of the first nanotechnology firm) through 2005. These data were collected from institutional-, industry-, and firm-level sources and gathered for each year so that no data are missing, which creates a "balanced panel" of data (Yaffee, 2003). In one such regression, I found that a higher level of governmental nanotechnology research funding in a state led to a higher level of nanotechnology entrepreneurship in that state.

Integration and synthesis

Until now I have discussed data and methods in distinct categories. While each contributes to the understanding of entrepreneurship, the greater contribution comes from their synthesis and integration. Yauch and Steudel (2003) summarize the common argument that qualitative and quantitative methods can be used together in organizational research to triangulate, elaborate, or guide further research. It is not only qualitative and quantitative methods that are useful in tandem, but also multi-source and multi-level data. For example, by collecting data at different levels of the same setting, one can examine questions regarding not only the influence of each level on entrepreneurship, but also the interaction between levels. This leads to a deeper, more holistic understanding of entrepreneurship.

In building a research agenda, there is no single correct or linear order of analysis. However, certain tasks and tools are more effective at different stages in the research program. In the nanotechnology entrepreneurship example, historical analysis was useful because without it I would not have been able to categorize firms, technologies, or products. It also improved my ability to interpret the findings of other methods. Conducting content analysis before using other methods of analysis allowed me to explore the data without predetermined categories. The findings of these analyses enabled me to develop further questions and directions for research. By integrating these findings, I gained an overview of nanotechnology emergence (from the historical analysis), an understanding of specific patterns of activity (construct and exploratory factor analysis), and the ability to develop and test specific hypotheses (regression analysis). The integration of each of these pieces provided a much better understanding of the nanotechnology entrepreneurship puzzle than any one method or data source could have provided.

DISCUSSION, CONCLUSIONS, AND IMPLICATIONS

The tools and techniques recommended here have several benefits. By collecting data not at the time of but *before* the emergence of an organization, the researcher avoids the problems of left-truncation and success bias. Longitudinal data provide insights into the processes and mechanisms of emergence that are not detectable through cross-sectional data. Data about the social context of an organization across more than one level allow for an analysis of external factors influential in entrepreneurship and the interaction between levels. Using several data sources avoids single-source bias, and using multiple collection methods allows for triangulation to improve the validity of the findings. Overall, the tools and techniques advocated here can enable triangulation beyond scaling, reliability testing, and convergent validation. This achieves a more complex study design that captures a more holistic and contextual portrayal of the phenomenon (Jick 1979).

The multi-source, multi-level data technique is not without limitations. For one, these techniques may not be optimal when the research question is very narrowly defined or only one specific relationship is of interest. While multiple techniques could be useful, a more simple research design might be sufficient. Second, these techniques may be difficult to achieve if research resources are limited. All data collection is time consuming and requires a set of related skills. However, in new domains of study, the data collection net must be cast widely. Collecting an additional source or level of data may or may not require a new research study. In either case, the initial data collection familiarizes the researcher with the nuances and complexities of the setting. Overcoming this initial hurdle enables the researcher to more efficiently acquire additional data. This informs both the current research agenda as well as any future research program of similar theory or phenomena.

This paper contributes to entrepreneurship literature by providing scholars with tools to facilitate data collection for research related to new venture emergence and entrepreneurial processes. I encourage entrepreneurship scholars to expand their collection and analyses of data to address more contextual research questions in their studies. Many levels of contexts from the family to inter-population communities are largely lacking in our understanding. Using techniques and tools such as those proposed here can lead to a more holistic understanding of entrepreneurship, a better understanding of the external forces influential in new venture creation, and, most importantly, more valid results. These techniques also contribute to research methods in several fields since they can be applied to the creation of other social entities such as industries, populations, communities, fields, institutions, and social movements. Ultimately, these fields will benefit by a closer alignment of data and research questions.

We often work on a small portion of a large puzzle. And while we attempt to fit the pieces into our portion, let us not forget to look at how our area of focus integrates into the overall picture. I hope that this paper provides encouragement and assistance for scholars to link more pieces together. Whether we connect firms to their environment or findings from one field to another, we create a more holistic view of entrepreneurship. In terms of future research, opportunities abound for researchers to examine these links and complete more of the puzzle.

REFERENCES

Aldrich, H.E. 1990. Using an Ecological Perspective to Study Organizational Founding Rates. *Entrepreneurship Theory and Practice* Spring: 7-24

Aldrich, H.E., & Baker, T. 1997. Blinded by the Cites? Has There Been Progress in Entrepreneurship Research? In D.L. Sexton & R.W. Smilor (Eds) *Entrepreneurship 2000*, pp. 377-400. Chicago: Upstart Publishing Company.

Aldrich, H.E. & Martinez, M.A. 2000. Many are Called, but Few are Chosen: An Evolutionary Perspective for the Study of Entrepreneurship. *Entrepreneurship Theory and Practice* Summer: 41-56.

Aldrich, H.E. & Ruef, M. 2006. *Organizations Evolving*. Thousand Oaks, CA: Sage Publications.

Aldrich, H.E. & Waldinger, R. 1990. Ethnicity and Entrepreneurship. *Annual Review of Sociology* 16: 111-135.

Baum, J.A.C. & Haveman, H.A. 1997. Love thy neighbor? Differentiation and agglomeration in the Manhattan hotel industry, 1898-1990. *Administrative Science Quarterly*, 42: 304-338.

Busenitz, L. W., West, G. P., Shepherd, D., Nelson, T., Chandler, G. N., & Zacharakis, A. 2003. Entrepreneurship research in emergence: Past trends and future directions. *Journal of Management*, 29(3): 285-308.

Carroll, G.R. & Mosakowski, E. 1987. The Career dynamics of self-employment. *Administrative Science Quarterly,* 32: 570-589.

Chandler, G.N. & Lyon, D.W. 2001. Issues of research design and construct measurement in entrepreneurship research: The past decade. *Entrepreneurship Theory and Practice,* 25: 101-113.

Crook, T. R., Shook, C. L., Morris, M. L., & Madden, T. M. 2010. Are We There Yet? An Assessment of Research Design and Construct Measurement Practices in Entrepreneurship Research. *Organizational Research Methods,* 13(1): 192-206.

Davidsson, P. 2003. The domain of entrepreneurship research: Some suggestions. In J. Katz, D. Shepherd (Eds.), *Advances in Entrepreneurship, Firm Emergence and Growth*, Vol. 6: 315-372. London: JAI.

Davidsson, P. & Wiklund, J. 2001. Levels of analysis in entrepreneurship research: Current research practice and suggestions for the future. *Entrepreneurship Theory & Practice,* 26: 81-99.

Dean, M. A., Shook, C. L., & Payne, G. T. 2007. The past, present, and future of entrepreneurship research: Data analytic trends and training. *Entrepreneurship Theory and Practice*, 31(4): 601-618.

Eisenhardt, K.M. 1989. Building theories from case study research. *Academy of Management Review,* 14: 532-550.

Gartner, W.B. 2001. Is there an elephant in entrepreneurship? Blind assumptions in theory development. *Entrepreneurship Theory and Practice*, 25: 27-39.

Granovetter, M. 1985. Economic action and social structure: the problem of embeddedness. *American Journal of Sociology*, 91: 481-510.

Greene, W.H. 2008. *Econometric Analysis* 5th ed. Upper Saddle River, N.J.: Prentice Hall.

Hannan, M. T. & Freeman, J. 1987. The Ecology of Organizational Founding - American Labor Unions, 1836-1985. *American Journal of Sociology,* 92: 910-943.

Hargadon, A. & Douglas, Y. 2001. When Innovations Meet Institutions: Edison and the Design of the Electric Light. *Administrative Science Quarterly*, 46: 476-501.

Ireland, R.D., Reutzel, C.R., & Webb, J.W. 2005. Entrepreneurship research in AMJ: What has been published, and what might the future hold? *Academy of Management Journal*, 48: 556-564.

Ireland, R.D., & Webb, J.W. 2007. A Cross-disciplinary exploration of entrepreneurship research. *Journal of Management*, 33: 891-927.

Ireland R.D., Webb J.W., Coombs J. 2005. Theory and methodology in entrepreneurship research. In DJ Ketchen, DD Bergh Eds., *Research Methodology in Strategy and Management.* London: JAI Publishing.

Jick, T.D. 1979. Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly,* 24: 602-611.

Kelloway, E.K. 1998. *Using Lisrel for Structural Equation Modeling.* Thousand Oaks, CA: Sage Publications.

Klein, K.J. & Kozlowski, S.W. 2000. From Micro to Meso: Critical Steps in Conceptualizing and Conducting Multilevel Research. *Organizational Research Methods*, 10: 211-236.

Kline, R.B. 1998. *Principles and Practice of Structural Equation Modeling*. New York: The Guilford Press.

Leblebici, H., Salancik, G., Copay, A., & King, T. 1991. Institutional change and the transformation of interorganizational fields: An organizational history of the United-States radio broadcasting industry. *Administrative Science Quarterly,* 36: 333-363.

Lee, T.W. 1999. *Using Qualitative Methods in Organizational Research.* Thousand Oaks, CA: Sage Publications.

Low, M.B. & MacMillan, I.C. 1988. Entrepreneurship: Past research and future challenges. *Journal of Management*, 14: 139-161.

Miles, M.B. & Huberman, A.M. 1984. *Qualitative Data Analysis*. Beverly Hills, CA: Sage Publications.

Mullen, M.R, Budeva, D.G., & Doney, P.M. 2009. Research methods in the leading small business-entrepreneurship journals: A critical review with recommendations for future research. *Journal of Small Business Management*, 47: 287-307.

Padgett, J.F. & Ansell, C.K. 1993. Robust action and the rise of the Medici, 1400-1434. *American Journal of Sociology*, 98: 1259-1319.

Pedhazur, E. J. & Schmelkin, L.P. 1991. Measurement, Design, and Analysis. Hillsdale, NJ, Lawrence Erlbaum Associates, Publishers.

Rao, H., Morrill, C., & Zald, M.N. 2000. Power plays: How social movements and collective action create new organizational forms. *Research in Organizational Behavior,* 22: 237-281.

Ruef, M. 2000. "The emergence of organizational forms: A community ecology approach." *American Journal of Sociology*, 106: 658-714.

Shane, S. & Venkataraman, S. 2000. The promise of entrepreneurship as a field of research. *Academy of Management Review,* 25: 217-226.

Shook, C. L., Ketchen, D. J., Hult, G. T. M., & Kacmar, K. M. 2004. An assessment of the use of structural equation modeling in strategic management research. *Strategic Management Journal*, 25(4): 397-404.

Short, J. C., Ketchen, D. J., Combs, J. G., & Ireland, R. D. 2010. Research Methods in Entrepreneurship Opportunities and Challenges. *Organizational Research Methods,* 13(1): 6-15.

Singleton Jr., R.A., & Straits, B.C. 2005. *Approaches to Social Research* 4th ed. New York, NY: Oxford University Press.

STATA Press. 2001. *Stata Reference Manual*. College Station, Texas: StataCorp.

Stinchcombe, A.L. 1965. Social structure and organizations. In J.G. March ed., *Handbook of Organizations* pp. 142-193. Chicago, IL: Rand McNally.

Strauss, A. & Corbin, J. 1998. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* 2nd ed. Thousand Oaks, CA: Sage.

Stuart, T.E. & Ding, W.W. 2006. When do scientists become entrepreneurs? The social structural antecedents of commercial activity in the academic life sciences. *American Journal of Sociology*, 112: 97-144.

Suhr, D.D. 2003. Principal Component Analysis vs. Exploratory Factor Analysis, Working Paper. University of Northern Colorado.

Van de Ven, A. H. & Garud, R. 1994. The coevolution of technical and institutional events in the development of innovation. In J. Baum & J. V. Singh (Eds.), Evolutionary Dynamics of Organizations: 425-443. New York: Oxford University Press.

Ventresca, M.J. & Mohr, J.W. 2002. Archival Research Methods. In J. A. C. Baum Ed., *Blackwell Companion to Organizations* pp. 805-828. Malden, MA: Blackwell Publishers.

Woolley, J. L. 2007. Understanding Organizational Community Creation: The Nanotechnology Community. Management. Unpublished dissertation, University of California, Irvine.

Woolley, J. L. 2010. Technology Emergence through Entrepreneurship across Multiple Industries. *Strategic Entrepreneurship Journal.* 4: 1-21

Woolley, J. L. & Rottner, R.M. 2008. Innovation Policy and Nanotechnology Entrepreneurship. *Entrepreneurship Theory and Practice*, 32: 791-811.

Yaffee, R. 2003. A Primer for Panel Data Analysis: http://www.nyu.edu/its/pubs/connect/fall03/yaffee_primer.html. New York: New York University.

Yauch, C.A. & Steudel, H. 2003. Complementary Use of Qualitative and Quantitative Cultural Assessment Methods. *Organizational Research Methods,* 6: 465-481.

Youtie, J. & Shapira, P. 2008. Building an innovation hub: A case study of the transformation of university roles in regional technological and economic development. *Research Policy,* 37: 1188-1204.

Yin, R. 2008. *Case Study Research: Design and Methods* (4th ed.). Thousand Oaks, CA: Sage Publishing.

Zahra, S. A. 2007. Contextualizing theory building in entrepreneurship research. *Journal of Business Venturing*. 22: 443-452.