



1042-2587  
© 2007 by  
Baylor University

# The Past, Present, and Future of Entrepreneurship Research: Data Analytic Trends and Training<sup>1</sup>

Michelle A. Dean  
Christopher L. Shook  
G. Tyge Payne

**Competent data analysis is essential to entrepreneurship research and to the discipline's progression. A three-study design was used to evaluate quantitative analytic trends and the adequacy of entrepreneurship doctoral training. First, trends were identified by assessing hypothesis-testing techniques in *Entrepreneurship Theory and Practice* and the *Journal of Business Venturing*. Second, top entrepreneurship scholars were surveyed regarding the importance of various quantitative analytical techniques to future research and their expectations regarding doctoral training. Third, newly minted entrepreneurship PhDs were queried regarding their perceived competence with these same techniques. These studies provide a past, present, and future perspective on data analysis techniques and competencies in entrepreneurship.**

## Introduction

Recent discussions among entrepreneurship scholars have focused on the domain of entrepreneurship and its place in the general management field (e.g., Brush et al., 2003; Ketchen, 2003; Sharma & Chrisman, 1999). As with any relatively young field of study, entrepreneurship is constantly evaluating and reevaluating its place among related fields of study in an attempt to establish legitimacy (Bruyat & Julien, 2001; Busenitz et al., 2003). In a recent assessment of the field, Busenitz et al. (2003) argued that in establishing legitimacy, theory development and method are inextricably intertwined. Whereas it is theory that determines the discipline's boundaries, it is method that facilitates the testing of such theories and enables communication within the discipline and

---

Please send correspondence to Michelle A. Dean, tel.: (619) 594-6839; e-mail: michelle.dean@sdsu.edu.

1. Paper resubmitted to *Entrepreneurship Theory and Practice* as a research note. A previous version of this paper was presented at the 2005 Academy of Management Conference, Honolulu, Hawaii. The authors would like to thank Garry Bruton, Wendy J. Casper, Jeffrey Conte, Justin L. Davis, Editor James J. Chrisman, and two anonymous reviewers for their insights and suggestions.

across related disciplines. Accordingly, both theoretical development and method are important keys to legitimizing the field. Whereas the bulk of the work assessing the progress of entrepreneurship research has been on theoretical development (e.g., Brush et al., 2003; Busenitz et al., 2003; Gartner, 1988; Shane & Venkataraman, 2000; Smith, Gannon, & Sapienza, 1989; Venkataraman, 1997; Vesper, 1983), the focus in this article is on an important aspect of method: quantitative data analytic techniques. In focusing this study on quantitative data analytic techniques, we do not wish to downplay the importance of theory or theory's relationship to analytic techniques and trends. Further, our focus on quantitative data analysis is not meant to imply that qualitative methods are not valuable. Instead, we have focused the study on quantitative data analytic techniques because the appropriate use of and sufficient sophistication in the use of analytical techniques is a critical component to the advancement and legitimacy of the entrepreneurship field, and to date, scant attention has been paid to the use of data analytic techniques in entrepreneurship.

In 2003, Brush and colleagues assessed the current state of doctoral education and developed seven specific recommendations for improvement. Among these recommendations were to "increase the availability of Ph.D. programs and concentrations . . . that provide rigorous theory, research and methodological training" (p. 316) and to design doctoral student education "to overcome the limitations of past research, and at the same time, prepare students to rigorously study determinants of creation, opportunity recognition, exploration and the exploitation of these opportunities, as well as the outcomes of these processes" (p. 321). Adequate training in data analysis provides a foundation for students to understand extant entrepreneurship research and prepares them to engage in high-quality research (Brush et al., 2003).

Thus far, analyses of methods in entrepreneurship have been quite limited. An early analysis (Low & MacMillan, 1988) suggested that entrepreneurship research was in need of more sophisticated multivariate and longitudinal research designs that were more theoretically driven. More recently, Chandler and Lyon (2001) conducted a study in an effort to determine if Low and MacMillan's call for improved research methods in entrepreneurship research was being realized. In their study, Chandler and Lyon contrasted the first and last years of their data and found that univariate and descriptive techniques used for hypotheses testing declined over the decade. However, to date, a *comprehensive* examination of what research techniques are currently being utilized in entrepreneurship studies, what entrepreneurship doctoral students know about various data analytic techniques, and what they need to know, has not been undertaken.

The purpose of this article is to assess the past and present state of research techniques used in entrepreneurship studies. Specifically in this manuscript, our goals are to: (1) determine which quantitative data analytic techniques are considered to be important in entrepreneurship research; and (2) assess the adequacy of data analytic training of recent entrepreneurship PhD students. These goals are accomplished using three related studies. The first study tracks data analytic trends in the two premier journals dedicated to entrepreneurship research, *Entrepreneurship Theory and Practice (ETP)* and the *Journal of Business Venturing (JBV)*, from inception. The second and third studies surveyed established and new entrepreneurship scholars, respectively, to assess student training needs and recent graduate competencies with various data analytic techniques. Using this three-study design, a comparison between present training and expectations for future research can be determined. Our attention now turns toward evaluating historical trends in data analysis as a means of developing a baseline for establishing competence with data analytic methods.

## Study 1: Data Analytic Trends

### Sample and Data

The first study tracks trends in the use of data analytic techniques in the entrepreneurship field over time. The sample consisted of a random sample of approximately 50% of all articles published in *ETP* (formerly *American Journal of Small Business*) from 1976 to 2004 and in *JBV* from 1985 to 2004. These two journals were chosen because they are widely recognized as the strongest journals whose mission is limited to entrepreneurship research (Chandler & Lyon, 2001; Shane, 1997). According to citation analysis, these two journals have more impact on the field than other journals devoted solely to entrepreneurship (Romano & Ratnatunga, 1997). Although other quality journals publish entrepreneurship articles, we deemed it important to limit the analysis to journals whose mission is to publish only entrepreneurship-related manuscripts to isolate data analytic trends in entrepreneurship research (cf. Shook, Ketchen, Cycyota, & Crockett, 2003). Further, in Chandler and Lyon's study, over 83% of the sampled manuscripts came from *ETP* and *JBV*, and Busenitz et al. (2003) found that general management journals include only a very small amount of space for entrepreneurship (approximately 1.8%). In fact, when we examined three top management journals, *Administrative Science Quarterly* (*ASQ*), *Academy of Management Journal* (*AMJ*), and *Strategic Management Journal* (*SMJ*), over the period of 1996–2004, we found only 29 published entrepreneurship articles. Of these articles, 25 involved quantitative data analysis, 2 were qualitative studies, and 2 were solely theoretical. Additionally, a large portion of these studies was the result of special calls for papers; *AMJ* devoted a special issue to international entrepreneurship in 2000, while *SMJ* published a special issue on strategic entrepreneurship in 2001. Although there were 10 quantitative studies published in 2001, 5 of the 9 years we examined had one or no studies published in these journals. Thus, meaningful trends could not be reasonably derived.

The current sample consists of 582 articles, 276 from *JBV* and 316 from *ETP*. Of these randomly chosen articles, a total of 354 articles (196 from *JBV* and 158 from *ETP*) analyzed data quantitatively. The average and median sample size of these articles was 1,090 and 395, respectively. Of the 354 articles that included quantitative analyses, 166 reported response rates. The median response rate of this subset of studies was 32.3%.

### Analyses and Results

Following procedures in similar studies (Shook et al., 2003; Stone-Romero, Weaver, & Glenar, 1995), the data analytic procedures used to test each study's hypotheses or research questions were coded. For example, if a study reported using factor analysis to examine the reliability of the survey items administered, coefficient alpha to examine the internal consistency of the items on the survey, and hierarchical regression to test the study's hypotheses, hierarchical regression was coded as the analytic method used for hypothesis testing. For studies testing more than one hypothesis, each primary method used to test each hypothesis was coded. One coauthor coded all 354 empirical articles and made the final decision in cases of discrepancies, and a second individual, a graduate student in management, coded a random set of 206 articles. The interrater reliability was 88%, which is higher than those in similar studies (e.g., 83%, Ford, MacCallum, & Tait, 1986).

Periodic percentage use indices (PUIs; Stone-Romero et al., 1995) were computed to assess trends in the data over time. Specifically, the frequency of use of each data analytic

Table 1

Empirical Study Design Characteristics and Data Analytic Method Percentage Use Indices

	Year article published													
	76	77	78	79	80	81	82	83	84	85	86	87	88	89
Number of articles	3	5	9	6	8	7	2	2	4	13	12	13	10	13
Number from <i>ETP</i>	3	5	9	6	8	7	2	2	4	8	7	4	4	1
Number from <i>JBV</i>	—	—	—	—	—	—	—	—	—	5	5	9	6	12
Number of hypotheses	11	5	19	20	8	7	4	5	4	27	36	58	28	39
Cross-sectional	1.0	1.0	1.0	1.0	.88	.71	1.0	1.0	1.0	.92	.92	.77	.80	.85
Longitudinal	0	0	0	0	.13	.29	0	0	0	.08	.08	.23	.10	.18
Single industry	.33	.20	.67	.50	.25	.14	.50	0	.25	.31	.42	.54	.30	.38
Multi-industry	.67	.60	.33	.50	.75	.86	.50	1.0	.25	.69	.58	.46	.60	.62
Nonindustry sample	0	.20	0	0	0	0	0	0	.50	0	0	0	0	0
Domestic sample	1.0	1.0	.89	.83	.75	1.0	.50	1.0	1.0	1.0	.83	.92	.90	.92
International sample	0	0	.11	.17	0	0	.50	0	0	0	.08	0	0	0
Both domestic & INT	0	0	0	0	.25	0	0	0	0	0	.08	.08	0	.08
Single country	1.0	1.0	1.0	1.0	.75	1.0	1.0	1.0	1.0	.92	.92	1.0	.92	.92
Multicountry	0	0	0	0	.25	0	0	0	0	0	.08	.08	0	.08
Individual LOA	0	.80	.67	.17	.13	.29	.50	.50	.50	.15	0	.15	.40	.15
Group LOA	0	0	0	0	0	0	0	0	0	0	0	0	.10	.08
Organization LOA	1.0	.20	.33	.83	.88	.71	.50	.50	.50	.77	1.0	.85	.40	.77
Multiple LOA	0	0	0	0	0	0	0	0	0	.08	0	0	0	0
Industry LOA	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Country LOA	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Descriptive stats	.36	.60	.89	.15	.88	0	.75	.80	.50	.67	.25	.14	.36	.18
Nonparametric stats	0	.40	.05	.65	0	0	.25	0	0	.15	.06	.52	.14	.10
Correlation	.64	0	.05	0	0	0	0	.20	0	0	.53	.19	.04	.03
Test of means	0	0	0	0	0	.43	0	0	.50	0	.06	.09	.18	.10
ANOVA	0	0	0	0	0	0	0	0	0	.04	0	0	0	.13
ANCOVA	0	0	0	0	0	0	0	0	0	0	0	0	0	.05
MANOVA	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MANCOVA	0	0	0	0	0	0	0	0	0	0	0	0	0	.03
Simple regression	0	0	0	0	0	.29	0	0	0	.04	0	0	.18	.15
Multiple regression	0	0	0	0	.13	0	0	0	0	.04	0	0	0	.10
Hierarchical regression	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Stepwise regression	0	0	0	0	0	.14	0	0	0	.04	0	0	0	0
Discriminant analysis	0	0	0	.15	0	0	0	0	0	0	0	.02	0	.08
Logistic regression	0	0	0	0	0	0	0	0	0	0	0	0	0	.05
Network analysis	0	0	0	0	0	0	0	0	0	0	0	0	.04	0
Cluster analysis	0	0	0	0	0	0	0	0	0	0	.03	0	.07	0
Path analysis	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SEM	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Causal mapping	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Canonical correlation	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Curvilinear regression	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CFA	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EFA	0	0	0	0	0	.14	0	0	0	0	.06	.02	0	0
Reliability analysis	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hybrid conjoint analysis	0	0	0	0	0	0	0	0	0	0	0	0	0	0

*ETP*, Entrepreneurship Theory and Practice; *JBV*, Journal of Business Venturing; INT, international; LOA, level of analysis; ANOVA, analysis of variance; ANCOVA, analysis of covariance; MANOVA, multivariate analysis of variance; MANCOVA, multivariate analysis of covariance; SEM, structural equations modeling; CFA, confirmatory factor analysis; EFA, exploratory factor analysis.

Year article published

90	91	92	93	94	95	96	97	98	99	00	01	02	03	04
16	15	17	19	14	19	12	18	21	13	22	14	19	12	16
3	7	2	9	3	7	2	7	7	3	6	7	16	3	6
13	8	15	13	11	12	10	11	14	10	16	7	3	9	10
48	44	83	87	55	65	61	68	122	79	125	73	70	45	53
.94	.93	.88	.94	.86	.63	.92	.78	1.0	.69	.91	.76	.84	.67	.75
0	0	.12	.05	.14	.37	.08	.22	0	.31	.09	.21	.16	.33	.25
.19	.13	.18	.32	.21	.21	.33	.28	.24	.38	.32	.21	.05	.08	.44
.75	.60	.71	.58	.79	.63	.58	.67	.67	.62	.64	.71	.84	1.0	.50
0	.27	0	.05	0	.16	.08	.06	.09	0	.05	.07	.11	0	.06
.81	.73	.47	.74	.71	.74	.66	.39	.62	.69	.64	.64	.42	.25	.38
.13	.20	.35	.16	.29	.26	.17	.61	.33	.23	.32	.29	.42	.67	.56
0	.07	.18	.11	0	0	.17	0	.05	.08	.05	.07	.16	.08	.06
1.0	.80	.76	.84	.86	1.0	.75	.83	.90	.92	.95	.93	.79	.83	.81
0	.20	.24	.16	.14	0	.25	.17	.10	.08	.05	.07	.21	.17	.19
.38	.33	.35	.37	.43	.53	.17	.28	.38	.23	.50	.14	.26	.25	.13
0	0	0	0	.07	.11	.17	0	.05	.08	.14	.07	0	0	0
.50	.60	.53	.53	.43	.37	.42	.72	.43	.69	.36	.71	.68	.67	.69
0	.07	.06	.05	.07	0	0	0	10	0	0	.07	.05	.08	.13
.06	0	0	0	0	0	.25	0	0	0	0	0	0	0	0
0	0	.06	.05	0	0	0	0	.05	0	0	0	0	0	.06
.27	.07	.02	.07	.09	0	.16	.03	.01	.01	.01	.03	.06	0	0
.13	.14	.45	.05	.05	.05	0	.01	.06	0	.02	0	.07	0	.02
.10	.07	.01	.14	.04	.05	.11	.04	.09	.14	.06	.05	0	0	0
.13	.23	.12	.08	.02	.08	0	.03	.09	.03	.13	0	.03	0	0
0	.02	.02	.05	16	0	0	0	.11	0	.05	.11	.13	.02	0
.06	0	0	0	.04	0	0	0	.04	0	0	0	0	0	0
0	.05	0	.07	0	0	0	.15	0	6	0	.04	.10	0	.04
0	0	.03	.01	0	0	0	0	0	0	0	.08	0	0	0
0	.14	0	.07	.05	.05	.18	.03	.12	.08	.37	.10	.07	.04	.02
.06	.02	.14	.38	0	.20	.49	.15	.03	.06	.01	.01	.24	.36	.19
0	0	0	0	.07	.02	.15	.09	.11	.51	.03	.15	.10	.07	.38
0	0	.02	.02	0	0	0	.13	.06	0	.05	0	.01	.02	0
.13	.02	.10	0	.04	.11	0	0	.05	0	.01	0	0	0	0
0	0	0	.02	.29	.15	.03	.12	.03	.09	.03	.18	.09	.36	.28
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
.04	0	0	0	0	0	0	0	.01	.01	.01	0	.01	.02	0
0	0	0	0	0	.12	0	0	.14	0	0	0	0	0	0
0	0	0	0	0	.12	0	0	0	0	.20	.21	0	.09	0
0	0	0	0	0	0	0	0	0	.01	0	0	0	0	0
.04	.18	.01	.03	0	0	0	0	0	0	0	.01	0	0	0
0	0	0	0	0	0	0	0	.02	0	0	0	0	0	0
0	0	0	0	0	0	0	.03	0	0	.01	0	0	0	0
.04	.05	.02	.01	.02	0	0	.03	0	0	.02	.01	0	0	0
0	.02	0	0	0	0	0	.01	0	0	0	0	0	0	0
0	0	.02	0	0	0	0	0	0	0	0	0	0	0	0

method was divided by the total number of hypotheses that were tested per period. A PUI of .30, for example, would indicate that 30% of the hypotheses tested during a specified time period relied on a particular method. Additionally, study design issues such as time frame (cross-sectional or longitudinal), industry sample (single, multi, or nonindustry), country sample (U.S.-only, international, both U.S. and international, single country, or multiple country), and level of analysis (individual, group, organization, industry, country, or multilevel) were examined; PUIs were calculated for these issues as well. Here, the level of analysis is the study rather than the hypothesis level, so PUIs were calculated by the frequency of use of a particular study design divided by the number of empirical studies published in a given time period.

Following previous studies, we assessed trends by computing correlations between years and annual PUI values for each method (Shook et al., 2003; Stone-Romero et al., 1995). Then, for description purposes only, we highlight trends over time by presenting aggregated PUIs arranged by approximately 10-year intervals (Shook et al., 2003). Starting from the inception of *ETP*, these time periods include 1976–1985, 1986–1995, and 1996–2004; these are labeled as 1, 2, and 3, respectively. These three time periods demonstrate that average sample sizes have continued to increase over time, whereas the average response rate has remained steady. The average sample size (and median sample size reported in parentheses) for time periods 1, 2, and 3 were 169 (107), 375 (141), and 510 (175), respectively. For the 166 studies reporting response rates, the median response rate for each time period was 31, 33, and 32%, respectively.

PUIs for the research designs and data analytic strategies employed by year are presented in Table 1. In Table 2, we summarized the aggregated PUIs for the data analytic methods using a framework adapted from Hitt, Gimeno, and Hoskisson (1998). In terms of data analytic techniques employed in entrepreneurship studies, some statistically significant trends emerged from the data. The use of simple descriptive statistics (e.g., means and standard deviations [SDs]) and nonparametric statistics (e.g., chi-square, Mann–Whitney U, Spearman correlation, Kruskal–Wallis, Wilcoxon-matched pairs signed rank) to test hypotheses has declined significantly over time ( $r = -.72$ ;  $p < .001$  and  $r = -.35$ ;  $p < .10$ , respectively). Other data analytic procedures have been increasing in their use. The most significant increasing trend is the use of general linear models to test hypotheses ( $r = .82$ ;  $p < .001$ ). Within this category of data analytic techniques, the method that has seen the largest increase in use over time is hierarchical regression ( $r = .59$ ;  $p < .001$ ). Other significant positive trends in general linear model techniques include multiple regression ( $r = .51$ ;  $p < .01$ ), multivariate analysis of variance (MANOVA) ( $r = .46$ ;  $p < .05$ ), analysis of variance (ANOVA) ( $r = .42$ ;  $p < .05$ ), and structural equations modeling (SEM) ( $r = .44$ ;  $p < .05$ ). For each period, the most popular methods based on PUIs were:

- Period 1 (1976–1985)—descriptive statistics, nonparametric statistics, and correlation analysis (56, 19, and 8% of all hypotheses tested, respectively);
- Period 2 (1986–1995)—nonparametric statistics, descriptive statistics, and multiple regression (18, 12, and 12% of all hypotheses tested, respectively);
- Period 3 (1996–2004)—hierarchical regression, simple regression, and multiple regression (16, 14, and 13% of all hypotheses tested, respectively).

For the same time periods, we aggregated PUIs for study characteristics employed by the researchers. As shown in Table 3, trends were also found in terms of the study design characteristics. The use of longitudinal studies has significantly increased over time ( $r = .53$ ;  $p < .01$ ), and the use of cross-sectional study designs has significantly declined over time ( $r = -.56$ ;  $p < .01$ ). The majority of quantitative studies has been published at the

Table 2

## Data Analytic Technique Percentage Use Indices and Correlations with Year

	Overall	1976–1985	1986–1995	1996–2004	Correlation with year
Data analytic technique					
Descriptive statistics	.11	.56	.12	.04	-.72***
Nonparametric tests	.10	.19	.18	.03	-.35†
Correlations	.08	.08	.11	.06	-.17
Tests of mean differences	.07	.05	.10	.05	-.13
General linear models	.42	.07	.29	.58	.82***
ANOVA	.05	.01	.04	.05	.42*
ANCOVA	.00	.00	.01	.01	.11
MANOVA	.03	.00	.02	.04	.46*
MANCOVA	.00	.00	.01	.01	.24
Simple regression	.10	.03	.06	.14	.31
Multiple regression	.12	.02	.12	.13	.51**
Hierarchical regression	.09	.00	.01	.16	.59***
Stepwise regression	.02	.02	.01	.03	.11
Canonical correlation	.01	.00	.03	.00	.05
Explicitly dynamic methods					
Event history/hazard studies	.00	.00	.00	.00	
Discrete events methods	.11	.03	.11	.12	.50**
Discriminant analysis	.03	.03	.05	.01	.03
Financial event study	—	—	—	—	—
Logistic regression	.08	.00	.06	.11	.67***
Longitudinal data methods	—	—	—	—	—
Methods for analysis of interdependence					
Network analysis	.00	.00	.00	.00	-.05
Methods explicitly accounting for heterogeneity					
Cluster analysis	.01	.00	.01	.01	.11
Causal structure methods	.06	.0	.03	.09	.45*
Path analysis	.02	.00	.02	.03	.22
Structural equation modeling	.04	.00	.02	.06	.44*
Methods to analyze decision making	.00	.00	.00	.00	.13
Hybrid conjoint analysis	.00	.00	.01	.00	.09
Causal mapping	.00	.00	.00	.00	.20
Methods to account for imperfect measurement of constructs	.02	.01	.02	.01	-.04
Confirmatory factor analysis	.00	.00	.00	.00	.22
Exploratory factor analysis	.01	.01	.02	.01	-.12
Reliability analysis	.00	.00	.00	.00	.05

†  $p < .10$ ; \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$

ANOVA, analysis of variance; ANCOVA, analysis of covariance; MANOVA, multivariate analysis of variance; MANCOVA, multivariate analysis of covariance.

organization (60%) or individual (31%) level of analysis, although studies examining the group level have shown the strongest upward trend over time ( $r = .41$ ;  $p < .01$ ). There was also a slight positive trend in analyzing data at the country level of analysis ( $r = .33$ ;  $p < .10$ ). Furthermore, significantly more research is being performed on international samples ( $r = .71$ ;  $p < .001$ ) as compared to research using U.S.-only samples. Although the majority of studies (90%) report data from a single country, the number of studies reporting data from only one country has significantly declined over time ( $r = -.53$ ;  $p < .01$ ).

Table 3

## Empirical Study Characteristics Percentage Use Indices

	Overall	1976–1985	1986–1995	1996–2004	Correlation with year
Study design					
Cross-sectional	.86	.93	.87	.83	–.56**
Longitudinal	.14	.07	.13	.17	.53**
Level of analysis					
Individual	.31	.34	.32	.28	–.17
Group	.04	0	.03	.05	.41*
Organization	.60	.64	.59	.58	.17
Multilevel	.03	.02	.03	.05	–.21
Industry	.01	0	.01	.02	.13
Country	.01	0	.01	.01	.33†
Sample					
U.S.-based	.69	.92	.77	.53	–.75***
International	.24	.05	.17	.40	.71***
Both U.S. and international	.06	.03	.06	.08	.30
Single country	.90	.97	.89	.87	–.53**
Multiple countries	.10	.03	.11	.13	.53**
Number of industries					
Single industry	.28	.34	.28	.26	–.27
Multiple industry	.66	.61	.66	.68	.26
Nonindustry sample	.06	.05	.06	.06	.03

†  $p < .10$ ; \*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$ 

## Study 2: Expert Rankings of Data Analytic Methods

### Sample and Data

The second study surveyed established entrepreneurship scholars to assess doctoral student training needs with quantitative data analytic techniques. The participants for this study included entrepreneurship researchers who: (1) were serving on the editorial boards of *ETP* and *JBV* during the summer of 2003; and/or (2) had demonstrated success in publishing multiple times in *ETP* and *JBV* between 1999 and 2003. We identified a total of 193 such scholars, but we were not able to locate contact information for 34. Of the 147 remaining scholars, we received 32 usable responses for a response rate of 21%. Based on Armstrong and Overton's (1977) suggestion for assessing potential response bias by comparing early and late responders, we compared the first and last third of the respondents on their current position (i.e., assistant, associate, or full professor) and number of articles published in *ETP* and *JBV*. No significant differences were found; thus, there was no evidence of a response bias.

An initial request to complete a survey and a follow-up request were e-mailed, along with a link to a webpage. On the webpage, respondents were presented a list of 31 data analytic techniques and were asked to indicate "how important are each of these data analytic techniques to the future of entrepreneurship research?" The list of techniques was developed using the results of Study 1 and suggestions made by Hitt et al. (1998) and Stone-Romero et al. (1995). Conceivably, experts could deem that a technique is important to the field but not believe that every new scholar should be competent with such a technique. Hence, we were also interested in the experts' opinion on whether or not new



scholars need to be competent with every method that experts believed to be important to the field. Respondents were also asked to “indicate the minimum level of competence that every new Ph.D. should have” with the same list of data analytic techniques using a 5-point Likert scale. Anchors for the scales were 1 (“not at all”) and 5 (“to a great extent”).

## **Analysis and Results**

As shown in Table 4, established entrepreneurship scholars indicated that a wide variety of techniques were important for the future of the field; this seems to be consistent with the plurality of the entrepreneurship discipline (Smith et al., 1989). There were nine techniques with a mean above 4.0 and 29 techniques with a mean above the midpoint of the scale (i.e., 3.0). The nine techniques that experts indicated as being particularly important for the future of entrepreneurship research were: (1) correlation, (2) ANOVA/analysis of covariance (ANCOVA), (3) multiple regression, (4) hierarchical regression, (5) logistic regression, (6) event history, (7) exploratory factor analysis (EFA), (8) confirmatory factor analysis (CFA), and (9) SEM. Only two techniques (repertory grid and canonical correlation) were not rated as particularly important for the future of the field.

Established scholars also reported that every new scholar should be competent with a wide variety of quantitative data analytic techniques. More specifically, established scholars rated 24 techniques above the midpoint of 3.0 and nine techniques above 4.0. The nine techniques with a mean rating above 4.0 were: (1) correlation, (2) *t*-tests, (3) ANOVA/ANCOVA, (4) MANOVA/multivariate analysis of covariance (MANCOVA), (5) simple regression, (6) multiple regression, (7) hierarchical regression, (8) logistic regression, and (9) EFA. According to the experts, there were seven techniques with which every student does not need to be competent. These techniques were financial event study, diffusion models, seemingly unrelated regression, repertory grid, cognitive mapping, policy capturing, and canonical correlation. Looking across the mean scores for the future of the field and the minimum level of competence for PhD students, we found six techniques that had an average importance rating of above 4.0. The SDs for these six techniques (correlation, ANOVA/ANCOVA, multiple regression, hierarchical regression, logistic regression, and EFA) were below 1.0; this seems to indicate agreement across the respondents.

## **Study 3: PhD Training in Data Analytic Methods**

### **Sample and Data**

The third study surveyed new entrepreneurship scholars to assess recent PhD graduate competencies with quantitative data analytic techniques. The data for this study were drawn from entrepreneurship researchers who attended the Academy of Management’s Entrepreneurship Division Doctoral Professional Development Workshop (PDW) between 1998 and 2003. A total of 114 scholars attended the PDW between 1999 and 2003. We were not able to locate contact information for eight scholars. Usable responses were collected from 40 scholars for a response rate of 35%, which compares to a 31% response rate achieved in a similar study of new strategic management scholars (Shook et al., 2003). Responses were received from two scholars who attended the PDW in 1998, 6 in 1999, 4 in 2000, 11 in 2001, 7 in 2002, 9 in 2003, and 1 respondent who did not answer when s/he attended the PDW. Again, the extrapolation procedure suggested by Armstrong and Overton (1977) was used to assess potential response bias according to gender and year of PDW attendance; no evidence of a response bias was found.

Table 4

## Expert Rankings and New Scholar Perceived Competence Levels of Analytic Techniques

	Results of study 2: Expert rankings				Results of study 3: New scholar perceived competence			
	For future of field		Students should be trained		Upon completion of PhD		At present	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Correlations	4.13	1.12	4.81	.48	4.15	1.15	4.55	.84
Tests of mean differences								
<i>T</i> -tests	3.94	1.09	4.68	.65	4.25	.95	4.52	.91
General linear models								
ANOVA and ANCOVA	4.19	.91	4.57	.86	3.54	1.35	3.93	1.14
MANOVA and MANCOVA	3.84	1.13	4.23	1.01	3.28	1.32	3.43	1.24
Simple regression	3.90	1.30	4.80	.55	4.32	1.23	4.55	1.06
Multiple regression	4.55	.62	4.87	.43	4.47	1.15	4.47	1.18
Hierarchical regression	4.48	.79	4.65	.71	3.47	1.60	3.62	1.69
Canonical correlation	2.93	1.31	2.87	1.09	2.13	1.27	1.83	1.11
Explicitly dynamic methods								
Event history	4.03	1.02	3.52	1.15	1.80	1.14	1.83	1.22
Discrete events methods								
Financial event study	3.26	1.39	2.73	1.11	1.45	1.04	1.55	1.18
Logistic regression	4.42	.72	4.30	.79	3.20	1.62	3.17	1.48
Multinomial logistic regression	3.87	1.12	3.55	1.21	2.12	1.27	2.20	2.00
Discriminant analysis	3.52	1.00	3.65	.95	2.51	1.39	2.43	1.48
Longitudinal data methods								
Panel data analysis	3.93	1.11	3.55	1.24	2.16	1.44	1.87	1.45
Repeated measures analysis	3.55	1.21	3.19	1.11	1.90	1.19	1.73	1.09



An initial request to complete a survey and a follow-up request were e-mailed to the participants, along with a link to a webpage. On the webpage, respondents were asked to indicate their proficiency level with the same list of methods assessed in Study 2. Specifically, we asked the respondents, “To what extent did your doctoral training enable you to use these methods?” as well as “To what extent are you now capable of using these methods?” Anchors for the scales were 1 (“not at all”) and 5 (“to a great extent”). Shook et al. (2003) utilized similar measures and techniques in a previous study of strategic management scholars.

## Analysis and Results

Table 4 also presents the means and SDs for perceived competence with the various quantitative data analytic techniques both upon completion of the new scholar’s PhD and at present. To assess whether or not respondents’ reported competence with the various techniques changed between graduation and at the time of the survey, we used paired-sample *t*-tests with the Bonferroni adjustment; no significant change was discovered. Upon completion of their PhD and at the time of the survey, respondents rated their competence at or above the midpoint (i.e., 3.0 on the 5-point scale) with 13 of the 31 techniques, and above 4.0 with only four of the techniques (i.e., correlations, *t*-tests, simple regression, and multiple regression).

Looking across perceived competence levels, we noticed that new scholars appeared to be confident with a relatively small number of techniques. In most cases, the SDs across all techniques were above 1.0, which suggests that although the typical researcher was not confident with many of the data analytic techniques, some researchers may have been. One possible explanation for this finding is that a small group of scholars was reportedly proficient with many techniques. However, we instead found that most new scholars reported competence in a few techniques, perhaps those techniques that they had the opportunity to use in prior research projects. The number of techniques with which the new researchers reported competence (i.e., rated as 4 or 5) ranged from 0 to 19, and the average was 8.4 (SD = 5.2) at graduation.

We were also interested in whether or not training with specific techniques appears to be related to time. We first computed correlations between method competence at graduation and the number of years since the respondent’s PhD was earned. Of the 31 methods, none were significantly correlated. Next, we examined whether or not the breadth of techniques with which the new graduates reported being competent (i.e., rated as 4 and 5) had increased over time. The correlation between breadth of techniques and years since graduation was not significant. Thus, we found no evidence that the training of new scholars in terms of perceived competence with specific techniques or breadth of techniques had changed in the recent past.

Lastly, we analyzed whether or not new scholars’ reported competence with techniques at graduation met the experts’ expectations. More specifically, we tested whether or not the mean competence reported by the new scholars was significantly different from the mean level that the experts reported should be exhibited by every new graduate. Across the 31 methods, we found that the experts’ expectations were significantly higher for all but five techniques (*t*-tests, simple regression, multiple regression, EFA, and canonical correlation). Of the nine techniques rated above 4.0 by the experts, new scholars rated themselves above 4.0 with only four (correlations, *t*-tests, simple regression, and multiple regression). The five techniques rated above 4.0 by the experts but rated below 4.0 by new scholars were ANOVA/ANCOVA, MANOVA/MANCOVA, hierarchical regression, logistic regression, and EFA.

## Discussion

The results of our studies should be viewed in light of their limitations. In the first study, we limited our examination to a random half of all studies published in *ETP* and *JBV*. This sample appears appropriate because those journals are widely recognized as the strongest journals whose mission is limited to entrepreneurship research (Chandler & Lyon, 2001; Romano & Ratnatunga, 1997; Shane, 1997). It is possible, however, that had we examined all studies in the selected journals, or studies in different journals, a different mix of methods may have been identified. Additionally, we did not account for theoretical issues associated with choice of data analysis technique. Because research methods and various statistical techniques are simply tools used to test theory and address specific research questions, interpretation of these trends should be made with caution.

In the second study, the raw number of expert entrepreneurship scholars limited our analysis to a relatively small sample size. As the field matures and as scholarly entrepreneurship programs continue to increase in number and quality, the quantity of experts in this field will grow. Despite this limitation, the surveyed scholars, as a whole, demonstrated consistently high expectations for new scholars in the field. Similarly, the third study may be limited by the sample. We surveyed doctoral students and new graduates who had attended recent Academy of Management Entrepreneurship PDWs; other entrepreneurship graduates might report different levels of competence with data analytic techniques than our respondents. Because admission to the PDW is competitive, our respondents should be among the better trained students (Shook et al., 2003). Thus, our findings that new scholars lack confidence in their competence with many techniques may underestimate the lack of competence of new graduates in general. Further, our measures of new scholar competence were self-reported perceptions, which raises a concern about potential self-report bias. Finally, we assessed competence both at graduation and at present and did not find significant differences. This lack of change in competence may be due to the relatively recent graduation of the individuals in the sample. Assessing the competencies of the same sample in the future may yield interesting insights into the development of analytical competencies after graduation.

Despite these limitations, our results reveal several interesting findings and offer some important implications for the entrepreneurship field. First and foremost, our analysis of published research (Study 1) offers some encouragement for the future state of entrepreneurship. In general, data analysis in entrepreneurship is becoming more sophisticated. Over time, there has been a shift away from using descriptive statistics and nonparametric tests for hypothesis testing to more rigorous techniques such as general linear models, logistic regression, and SEM. Further, entrepreneurship researchers have increasingly utilized longitudinal designs, which are more effective in establishing causality (Low & MacMillan, 1988). These results parallel those reported by Shook et al. (2003), who used similar methods to analyze the techniques utilized by researchers published in the *SMJ* from 1980 through 2001.

Given Busenitz et al.'s (2003) finding that the prevalence of entrepreneurship studies does not appear to be increasing in major management journals, we were curious about how the methods for entrepreneurship studies published in top scholarly management journals compared to the methods in the entrepreneurship journals we examined. In an exploratory analysis, we compared the empirical entrepreneurship studies published in *AMJ*, *ASQ*, and *SMJ* during the 1996–2004 time period ( $N = 25$ ) to the *ETP* and *JBV* studies during that same period ( $N = 147$ ). We found that 68% of the techniques used in the general (i.e., nonentrepreneurship specific) management journals were general linear

modeling techniques, with another 16% of the studies using logistic regression. More specifically, we found that for the entrepreneurship studies published in the general management journals, there was a greater emphasis on the use of hierarchical regression (PUI of .43 compared to .16 in entrepreneurship journals,  $p < .001$ ), logistic regression (PUI of .16 compared to .11 in entrepreneurship journals,  $p < .01$ ), and event history studies (PUI of .05 compared to .00 in entrepreneurship journals,  $p < .05$ ). Given these findings, we were curious about whether or not the entrepreneurship studies in the general management journals were indicative of their journals as a whole; thus, we examined a sample of 25 nonentrepreneurship studies from the same journals and found no significant differences. Although the differences between entrepreneurship studies in general management journals and those in *ETP* and *JBV* are interesting and may reflect a preference on the part of the general management journals for the use of more sophisticated data analytic techniques and longitudinal study design, until entrepreneurship studies appear more frequently in the general management journals, a more robust and meaningful analysis may not be completed.

The findings of the second and third studies are perhaps less encouraging than the findings from the first study. The survey of established entrepreneurship scholars (Study 2) supported the notion that new entrepreneurship students need a broad understanding of research methods in general and, more specifically, of data analytic techniques (Brush et al., 2003). Viewed alongside our findings regarding new scholar confidence in their ability to utilize various techniques (Study 3), the findings lead to interesting points of further discussion. New scholars reported being underprepared with most data analytic techniques. Further, when comparing methods that are growing over time with the perceived competence levels reported by new scholars, there appears to be a dearth of reported competence with methods that are increasingly needed in the future. More specifically, we noted that ANOVA/ANCOVA, MANOVA/MANCOVA, multiple regression, hierarchical regression, logistic regression, and SEM were being used more frequently in entrepreneurship research. However, the average new scholar reported feeling competent (i.e., rated above 4) with only one of these techniques (i.e., multiple regression). Such divergence may or may not be a cause for concern given the diverse nature of the entrepreneurship field. New scholars may not need to be highly competent in a large number of techniques; instead, it may be sufficient to have general knowledge of a range of techniques while developing special expertise in a limited number of methods specific to their respective areas of study. It may also be that these scholars' research interests only required expertise in multiple regression techniques, which resulted in the lack of reported competence in the other techniques. Although special expertise across a wide range of methods is not a realistic goal for any scholar, a general knowledge of a number of methods would serve new entrepreneurship scholars well in having the knowledge to move an area forward via interpreting data using appropriate, alternative approaches. New scholars identified nine data analytic methods in which they perceived they had average competence. This finding is encouraging as it may suggest a working knowledge of a technique's possibilities, which can increase the likelihood that such a technique will be applied in a scholars' research once the need arises and further training is sought.

As previously mentioned, the new scholars varied greatly from one individual to the next on how they reported their expertise with many of the techniques. For instance, the lack of confidence in using SEM among new scholars may be a cause for concern. However, taking into account the SD of the SEM response, we noticed that despite a lower average, several new scholars felt confidence in their abilities with SEM. This may indicate that what seem to be inadequacies in entrepreneurship training as a whole may actually be a reflection of the specialization of entrepreneurship scholars subject to the

trends emerging in subdisciplines. Nevertheless, the use of SEM is growing rapidly, and it is ideally suited for testing complex interrelationships among constructs while simultaneously accounting for measurement error. Indeed, Brush et al. (2003) named SEM as holding a particular promise for understanding key aspects of entrepreneurial creation, cognition, and opportunity recognition.

Another interesting finding involves the techniques that help establish the validity of measures. Low and MacMillan (1988) cited measure validation as critical to the advancement of entrepreneurship research. Although there was not an increase in the PUIs for CFA or EFA, this is not necessarily a cause for concern because researchers would be expected to use CFA or EFA to establish the properties of the measures *prior* to hypothesis testing. However, we do potentially see a cause for concern given the lower levels of competence for these two techniques. We would expect, as do the expert scholars surveyed in Study 2, that CFA and EFA are core techniques that every new scholar should be competent in. Other techniques with which new scholars reported low levels of competence are those used to assess longitudinal relationships (panel data analysis and repeated measures analysis). As longitudinal data become more common in entrepreneurship (e.g., Reynolds, 2000), expertise in analyzing such data will become much more important. As previously mentioned, the results of Study 1 suggest that longitudinal studies have increased in prevalence in *ETP* and *JBV*.

These findings suggest that new entrepreneurship scholars may not be adequately prepared to fully contribute to the growing demands of the field. Competence with a broad core of techniques leads to a common language and means of communication within a discipline (Pfeffer, 1993), as well as across disciplines. Further, such competencies establish a foundation for learning new, more specialized techniques if the need arises. This helps scholars address some of the more specific issues important to the field without detracting from its heterogeneity and breadth (Davidsson & Wiklund, 2001). As part of our ongoing comparison, we were curious about how new entrepreneurship scholars compared to new strategic management scholars. We compared the means of 77 strategy scholars' self-reported level of competence with various data analytic techniques upon completion of the PhD program (Shook et al., 2003) with the entrepreneurship students surveyed here. The results indicated that there were several techniques with which strategy scholars reported significantly higher competence levels than entrepreneurship scholars (i.e., simple regression—4.69 vs. 4.32;  $p < .05$ ; event history—2.50 vs. 1.80;  $p < .01$ ; financial event studies—2.23 vs. 1.45;  $p < .01$ ; multinomial logistic regression—2.80 vs. 2.12;  $p < .01$ ; panel data analysis—2.88 vs. 2.16;  $p < .01$ ; diffusion models—1.59 vs. 1.21;  $p < .05$ ; simultaneous equations—2.49 vs. 1.77;  $p < .01$ ).

New and more innovative ways to foster knowledge transfer are seemingly necessary if new scholars' competencies with data analytic techniques are to be improved. One option is to use collaboration across departments within a university or across universities (Brush et al., 2003; Shook et al., 2003). Another option might be to utilize the Academy of Management Entrepreneurship Division's list-serv and website to better facilitate methods dialogue and research collaboration across universities. Further, professional development programs that address research methods and statistical techniques of particular interest to entrepreneurship scholars might be initiated. Such programs would also aid young scholars in establishing a personal network of scholars upon which they can draw when the need arises.

Although such recommendations are easily made, they are much harder to implement because numerous issues contribute to the difficulty of training young entrepreneurship scholars. First, entrepreneurship research addresses a wide variety of issues and involves wide-ranging populations and samples (Davidsson & Wiklund, 2001). Mastery of an

extensive number of theories and content areas is necessary to be an “entrepreneurship” scholar; entrepreneurship draws from strategy, organization theory, and organization behavior, to name just a few relevant fields (Katz, 2003). Second, established theory is still forthcoming; thus, empirical studies of entrepreneurship must often be simultaneously theoretically driven and exploratory in nature. Such studies require more extensive background knowledge in related fields (e.g., strategy, leadership) in addition to the entrepreneurship literature. Further, data analytic tools are based on a number of issues such as levels of analysis, data limitations, and research questions. Therefore, the use of more sophisticated techniques is not necessarily indicative of better or more interesting research; instead, it may be a reflection of the current state of the field and its needs. Finally, time is limited and scholars may find it difficult to simultaneously handle the rigors of top research and the external outreach and teaching requirements of the entrepreneurship field.

## Conclusion

In their appraisal of entrepreneurship research, Low and MacMillan (1988, p. 1939) wrote, “As a body of literature develops, it is useful to stop occasionally, take inventory of the work that has been done, and identify new directions and challenges for the future. This reflective process is essential in order to derive the maximum benefits from future research.” Our focus in these three studies was on quantitative data analysis techniques. As we took inventory of data analytic technique usage across time, we identified a trend toward the use of more complex techniques. Our survey of established entrepreneurship scholars indicated that new scholars should be competent with a wide variety of techniques. However, our survey of new scholars indicated that they lack confidence in their competencies with a wide variety of techniques. Further, new entrepreneurship scholars reportedly feel confident in only a few core techniques, along with a variety of others on an individual basis. This suggests that the eclectic nature of entrepreneurship scholarship and the relatively quick growth of the field may play large roles in the usage of various techniques. Ideally, new scholars should have moderate levels of knowledge across a variety of areas while developing expertise in the methods that are most applicable for the research questions they want to explore. Looking toward the future and its challenges, our results indicate that adequate preparation of new scholars with more sophisticated data analytic techniques is a challenge that must be met, especially as competition for space in more general management journals continues to intensify.

## REFERENCES

- Armstrong, J.S. & Overton, T.S. (1977). Estimating nonresponse bias in mail surveys. *Journal of Marketing Research*, 14, 396–402.
- Brush, C.G., Duhaime, I.M., Gartner, W.B., Stewart, A., Katz, J.A., Hitt, M.A., et al. (2003). Doctoral education in the field of entrepreneurship. *Journal of Management*, 29, 309–331.
- Bruyat, C. & Julien, P.A. (2001). Defining the field of research in entrepreneurship. *Journal of Business Venturing*, 16, 165–180.
- Busenitz, L.W., West, G.P. III, Shepherd, D., Nelson, T., Chandler, G.N., & Zacharakis, A. (2003). Entrepreneurship research in emergence: Past trends and future directions. *Journal of Management*, 29, 285–308.



- 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.
- Davidsson, P. & Wiklund, J. (2001). Levels of analysis in entrepreneurship research: Current research practice and suggestions for the future. *Entrepreneurship Theory and Practice*, 25, 81–100.
- Ford, J.K., MacCallum, R.C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, 39, 291–314.
- Gartner, W.B. (1988). Who is the entrepreneur? Is the wrong question. *American Journal of Small Business*, 12, 11–32.
- Hitt, M.A., Gimeno, J., & Hoskisson, R.E. (1998). Current and future research methods in strategic management. *Organizational Research Methods*, 1, 6–44.
- Katz, J.A. (2003). The chronology and intellectual trajectory of American entrepreneurship education: 1876–1999. *Journal of Business Venturing*, 18, 283–301.
- Ketchen, D.J. (2003). Entrepreneurship: Past accomplishments and future challenges. *Journal of Management*, 29, 281–283.
- Low, M.B. & MacMillan, I.C. (1988). Entrepreneurship: Past research and future challenges. *Journal of Management*, 14, 139–162.
- Pfeffer, J. (1993). Barriers to the advance of organizational science: Paradigm development as a dependent variable. *Academy of Management Review*, 18, 599–620.
- Reynolds, P.D. (2000). National panel study of U.S. business startups: Background and methodology. In J.A. Katz (Ed.), *Advances in entrepreneurship, firm emergence and growth* (Vol. 4, pp. 153–228). New York: JAI Elsevier Science.
- Romano, C. & Ratnatunga, J. (1997). A citation classics analysis of articles in contemporary small enterprise research. *Journal of Business Venturing*, 12, 197–212.
- Shane, S. (1997). Who is publishing the entrepreneurship research? *Journal of Management*, 23, 83–95.
- Shane, S. & Venkataraman, S. (2000). The promise of entrepreneurship as a field of research. *Academy of Management Review*, 25, 217–226.
- Sharma, P. & Chrisman, J.J. (1999). Toward a reconciliation of the definitional issues in the field of corporate entrepreneurship. *Entrepreneurship Theory and Practice*, 23, 11–27.
- Shook, C.L., Ketchen, D.J., Cycyota, C.S., & Crockett, D. (2003). Data analytic trends in strategic management research. *Strategic Management Journal*, 24, 1231–1237.
- Smith, K.G., Gannon, M.J., & Sapienza, H.J. (1989). Selecting methodologies for entrepreneurial research: Trade-offs and guidelines. *Entrepreneurship Theory and Practice*, 13, 39–49.
- Stone-Romero, E.F., Weaver, A.E., & Glenar, J.L. (1995). Trends in research design and data analytic strategies in organizational research. *Journal of Management*, 21, 141–157.
- Venkataraman, S. (1997). The distinctive domain of entrepreneurship research: An editor's perspective. In J. Katz & R. Brockhaus (Eds.), *Advances in entrepreneurship, firm emergence, and growth* (Vol. 3, pp. 119–138). Greenwich, CT: JAI Press.
- Vesper, K.H. (1983). *Entrepreneurship and national policy*. Pittsburgh, PA: Carnegie-Mellon University.

---

Michelle A. Dean is an assistant professor in the Department of Management, College of Business Administration at San Diego State University.

Christopher L. Shook is an associate professor in the Department of Management at Auburn University.

G. Tyge Payne is an assistant professor in the Area of Management, Rawls College of Business at Texas Tech University.