

REPLICATION IN STRATEGIC MANAGEMENT: SCIENTIFIC TESTING FOR VALIDITY, GENERALIZABILITY, AND USEFULNESS

RAYMOND HUBBARD¹*, DANIEL E. VETTER² AND ELDON L. LITTLE³

¹College of Business and Public Administration, Drake University, Des Moines, Iowa, U.S.A.

²School of Business, Central Michigan University, Mount Pleasant, Michigan, U.S.A.

³Division of Business and Economics, Indiana University Southeast, New Albany, Indiana, U.S.A.

A number of studies have shown that little replication and extension research is published in the business disciplines. This has deleterious consequences for the development of a cumulative body of knowledge in these same areas. It has been speculated, but never formally tested, that replication research is more likely to be published in lower tiers of the journal hierarchy. The present paper indicates very low levels of replication in management and strategic management journals, regardless of their prestige. Moreover, even those replications that are published tend not to be critical—odd in applied social sciences that are largely preparadigmatic and where extensibility, generalizability and utility of scientific constructs tend to be low. The goal of science is empirical generalization, or knowledge development. Systematically conducted replications with extensions facilitate this goal. It is clear, however, that many editors, reviewers, and researchers hold attitudes toward replication research that betray a lack of understanding about its role. Long-run strategies to dispel these misconceptions must involve changes in graduate training aimed at making the conduct of such vital work second nature. It is further suggested that journals in all tiers create a section specifically for the publication of replication research, and that top-tier journals take the lead in this regard. © 1998 John Wiley & Sons, Ltd.

Strat. Mgmt. J., Vol. 19, 243–254, (1998)

INTRODUCTION

A number of authors from a variety of disciplines have indicated that the integrity of their respective empirical literatures is likely to be compromised (Dewald, Thursby, and Anderson, 1986; Feige, 1975; Greenwald, 1975; Hubbard and Armstrong, 1992; Hubbard and Lindsay, 1995; Lindsay, 1994; Stewart and Feder, 1987). In particular, they are said to be filled with Type I errors (erroneous rejections of the null hypothesis) and other dubious results. These authors attribute much of the blame for this state of affairs on editorial and

reviewer biases favoring the publication of statistically significant ($p < 0.05$) results. This, in turn, encourages data mining among researchers anxious to produce such results. The upshot is published Type I error rates well in excess of those prescribed by nominal alpha levels such as 0.05.¹ Rosenthal (1979) refers to this as the 'file drawer problem,' which taken to the extreme implies that journals are filled with the 5 percent of studies that are Type I errors, while the remaining 95 percent with null outcomes gather dust in the file drawers of researchers.

Worse yet, some erroneous results, Type I or otherwise, filter into the textbooks to be taught as established truths semester after semester. Indeed,

Key words: empirical generalization; replication; replication with extension; trickle-down effect

*Correspondence to: Raymond Hubbard, College of Business and Public Administration, Drake University, Des Moines, IA 50311, U.S.A.

¹ Lovell (1983) provides a 'rule of thumb' to deflate exaggerated claims of significance arising from data mining in a regression analysis context.

many textbook results may be no more than examples of management folklore, i.e., stories, customs, and beliefs that lack empirical confirmation, but whose probity is nevertheless accepted at face value. 'Classic' examples from the management discipline are Frederick W. Taylor's account of the loading of pig iron (Wrege and Perroni, 1974) and the Hawthorne effect (Franke, 1980; Franke and Kaul, 1978). More recent empirical studies refuting management folklore include the demonstration that use of the Boston Consulting Group's growth-share matrix actually misleads decision-makers and is harmful to company performance (Armstrong, 1996), as are management science techniques oriented toward maximizing market share at the expense of profit maximization (Armstrong and Collopy, 1996).

But is the problem concerning the integrity of research findings as dire as some authors claim? Don't the usual scientific checks on the veracity of published empirical works—peer review, critical commentary on existing papers and, most important of all, replication research—ensure that specious results will be kept to a minimum?

Unfortunately, the answer to the latter question appears to be no. Peer review, while admittedly unreliable as a quality control device (Armstrong, 1997; Cicchetti, 1991; Gans and Shepherd, 1994), nevertheless cannot be expected to check calculations in papers. Nor can peer review determine how much data mining (Lovell, 1983) may have occurred, since reporting of such activity is scrupulously avoided in most articles. Moreover, although comments and rejoinders play very valuable roles in promoting a reliable literature, and as such deserve encouragement, they are not as powerful as replication research in this regard.

The principle of replicability plays a fundamental role in the research process. After all, replications serve to protect against the uncritical assimilation of erroneous empirical results into the literature (Hubbard and Vetter, 1996). Replications with extensions also help to accomplish this, but more importantly go further by determining the scope and limits of initial findings by seeing if they can be *generalized* to other populations, time periods, organizations, geographical areas, measurement instruments, contexts, and so on. Thus replications with extensions are basic to empirical generalization or knowledge development (Galtung, 1967; Hubbard and Vetter, 1996;

Lindsay and Ehrenberg, 1993). Unreplicated research findings, even those that are highly statistically significant, are only speculative in nature (Hubbard and Armstrong, 1994) and 'virtually meaningless and useless' in themselves (Lindsay and Ehrenberg, 1993: 219). It is not for nothing that the principle of replicability has been hailed as the hallmark of science (Blaug, 1992) and is almost universally accepted as the most important criterion of genuine scientific knowledge (Rosenthal and Rosnow, 1984).

Given its importance, it might be expected that the management and social science literatures would be replete with examples of published replication research. Yet studies that have estimated the publication incidence of such work in these fields typically have produced disquieting results. For example, less than 10 percent, and frequently less than 5 percent, of empirical works published in the management and social sciences are replications with extensions (Bozarth and Roberts, 1972; Hubbard and Armstrong, 1994; Hubbard and Vetter, 1996; Reid, Soley, and Wimmer, 1981; Zinkhan *et al.*, 1990). Thus, the vast majority of empirical articles in these areas consist of uncorroborated 'single-shot' studies. More disturbingly, in the few instances where extensions are undertaken, they often conflict with, or at best partially support, the original findings (Hubbard and Vetter, 1996; Zinkhan *et al.*, 1990). This suggests that the knowledge bases in these fields rest on shaky ground.

The previously cited studies have invariably based their estimates of published replication research on content analyses of the *leading journals* in their fields, and as a result may have *underestimated* the amount of such work contained in their literatures. The argument is a simple one: The finite amount of space in the leading journals is largely reserved for original work, so that replication research may be consigned for publication in journals outside the top tier in any given discipline.

The above argument, suggested but not condoned by Hubbard and Armstrong (1994), that a *trickle-down* effect possibly occurs with the publication of replication research, ultimately is an empirically testable proposition. It is, however, one that has remained free from scrutiny. The first objective of the present paper, therefore, is to evaluate this trickle-down hypothesis. In particular, the publication frequency of replication

research is examined across three tiers of journals in the strategic management area for the 20-year period 1976–95. Second, we determine whether the amount of replication research has been increasing or decreasing over time. Third, we investigate to what extent replication studies support, partially support, and conflict with the results of their predecessors. Fourth, we report on the timeliness of replication research undertaken in the three tiers of the journal hierarchy. Finally, we estimate how much discussion concerning the reliability of previous studies is published in the strategic management area.

HYPOTHESES

The trickle-down hypothesis states that because of the need to reserve limited space in the top journals for original work, replication research is more likely to be published in lower-tier journals. This hypothesis could be expressed in a strict hierarchical manner, that is, top-tier journals publish fewer replications than second-tier journals, which in turn publish less of this work than third-tier journals, and so on down the line. However, the hypothesis is stated here in a more general way:

Hypothesis 1: Top-tier journals will publish less replication research than their second- and third-level counterparts.

Earlier calls for establishing a replication tradition in the business disciplines (e.g., Armstrong, 1982; Brown and Coney, 1976; Reid *et al.*, 1981) have increased in intensity over the years (e.g., Hubbard, Brodie, and Armstrong, 1992; Hubbard and Armstrong, 1994; Hubbard and Vetter, 1996; Lindsay and Ehrenberg, 1993; Mittelstaedt and Zorn, 1984; Zinkhan *et al.*, 1990). Such calls have also mustered editorial support (Ashenfelter, 1986; Kane, 1984; Monroe, 1992a, 1992b; Stout and Heck, 1995). Given this growing awareness of the importance of replications and extensions, their publication incidence should be higher for more recent years than for earlier ones. Consequently, the following hypothesis is offered:

Hypothesis 2: The publication frequency of replication research will be higher for the period 1986–95 than it was during 1976–85 across all three tiers of the journal hierarchy.

The outcomes from a replication study either support, partially support, or conflict with the original findings. Clearly, the higher the incidence of contradictory results, the greater the difficulty of developing a cumulative knowledge base. There is, however, no pressing *a priori* reason to expect differences in the proportion of these three outcomes across the three tiers in the journal hierarchy. Hence, the next hypothesis is advanced:

Hypothesis 3: There is no difference in the frequency of outcomes—support, partial support, or conflict—of replication research across the three tiers of the journal hierarchy.

With respect to the timeliness of replication research, Hubbard and Vetter (1996) note that rapid attempts to replicate the findings of original studies are preferable to those conducted at later dates. Other things being equal, replications that either confirm or contradict the findings of the original work in a timely manner are especially important and merit rapid dissemination. Successful early replication attempts, such as Oswald and Jahera's (1991) corroboration of Kim, Lee, and Francis' (1988) results on the positive relationship between ownership structure and the financial performance of the firm, help foster the development of a cumulative knowledge base. In contrast, early disconfirmations, as in Robinson and Pearce's (1983) contradiction of Wood and LaForge's (1979) findings that formal strategic planning is positively related to organizational performance, correctly impede agreement. These observations notwithstanding, there is no peremptory rationale for assuming that the timeliness of published replications with extensions should be expected to vary across the three tiers of journals in the strategic management hierarchy. Accordingly, the following hypothesis is formulated:

Hypothesis 4: There is no difference in the timeliness of replication articles published in the first, second, and third tiers of the journal hierarchy.

The conduct of science is meant to be a *public* enterprise. As such, independent replication constitutes the primary means for determining the authenticity and generalizability of a discipline's empirical findings. But other important research activities warrant attention in this context. Chief

among these is the valuable role of the research comment and rejoinder. The comment and rejoinder provide a forum for publicly debating the integrity of a discipline's literature. Consequently, it is instructive to monitor their publication frequency. Again, however, there is no compelling reason to anticipate differences in these frequencies throughout the journal hierarchy, as the next hypothesis asserts:

Hypothesis 5: There is no difference in the publication incidence of comments and rejoinders across the three tiers of the journal hierarchy.

RESEARCH DESIGN

Researchers wishing to empirically investigate the replicability of their results can do so using 'internal' and 'external' analyses. Although better than no 'replication' at all, internal analyses—such as (double) cross-validation, the jackknife, and the bootstrap—involve the use of data from the sample in hand, and always yield exaggerated claims concerning the replicability of findings (Thompson, 1994). In contrast, external analyses, which employ a new and independent sample of subjects, invoke true replication and are inherently superior to internal evaluations (Thompson, 1994). Our study is devoted to an examination of external approaches to replication.

While considered useful, no attempt was made in the present study to determine whether an author had replicated his or her own results *within* the context of the *original* article. Because of our concern with the willingness of journals to publish replications as worthy contributions in their own terms, however, we did include in our counts instances where authors published replications of their own work as *separate* articles.

It is worth noting that by including these latter works we are being more liberal in our estimates of published replication research than some other authors and editors. The argument has been made that when authors replicate their own research they perhaps transfer biases that may have been associated with the original study to the replication. Rosenthal and Rosnow (1984), for example, refer to this as the problem of 'correlated replicators,' and call into question the value and impartiality of 10 replications performed by

the same researcher vs. 10 replications each conducted by a different investigator. Similarly, in a bold initiative begun in 1984, the editorial board of the *Quarterly Journal of Business and Economics (QJBE)* agreed to give *priority* to the publication of replication research. As a corollary, however, the decision was also made to reject papers where an author replicated his/her own work, or where a student replicated the work of his/her professor. As the editor of the *QJBE* (McCabe, 1984: 79) stated: 'Clearly, it is the independence of the two works that makes replication valuable.'²

Definitions

The definitions of replication and replication with extension follow those used by other researchers who sought to estimate their publication occurrence (Brown and Coney, 1976; Hubbard and Armstrong, 1994; Hubbard and Vetter, 1996; Reid *et al.*, 1981). Thus, a *replication* is a substantial duplication of a previously published empirical research project that is chiefly concerned with increasing the internal validity of the research design. A replication focuses on determining whether the initial results are *reproducible*. Accordingly, replications attempt to duplicate all facets of the original work. An example would be repeating the study with another sample drawn from the same population.

A *replication with extension* is a duplication of a previously published empirical research project that is primarily concerned with increasing the external validity, or *generalizability*, of previous research findings. Consequently, an extension modifies various aspects of the research design used in the original study, such as the manipulated (independent) or measured (dependent) variables. This is because the major goal of extensions is to ascertain whether earlier results are capable of being generalized to other populations, product categories, organizations, time periods, measurement instruments, geographical areas, investigators, and so on, rather than being idiosyncratic or localized in nature. Systematically conducted replications with extensions, or 'differ-

² The *QJBE* policy change seems to have been successful. In a census of all articles published in the *QJBE* from 1984 to 1994, Fuess (1996) discovered that some 23 percent (62/275) of papers were replications or extensions.

entiated' replications (Lindsay and Ehrenberg, 1993), serve this purpose. They are central to empirical generalization, or knowledge development, and the route for determining whether research results are useful and can be applied to practical problems (Hubbard and Lindsay, 1995).

Samples

To be eligible for inclusion in the three-tier journal hierarchy, certain criteria had to be met. First, the journals had to regularly feature empirical work. Because it does not, the prestigious *Academy of Management Review* was eliminated from consideration. Second, the chosen journals had to have been published for enough years to cover all, or a good part, of the 20-year (1976–95) time period encompassed in this study. With the exception of the *Strategic Management Journal* (1980–95), all other selected journals spanned the entire 20 years. Third, the journals had to be focused primarily on the management discipline. This precluded using the journal rankings for the management area developed by Sharplin and Mabry (1985) and Salancik (1986), 50 percent of whose lists of 20 and 24 journals, respectively, were mainstream or specialized psychology and sociology periodicals. Coe and Weinstock's (1984) ranking of journals, while more clearly management-oriented than the above two, were felt to be less suitable for our purposes than the more recent one conducted by Franke, Edlund, and Oster (1990).

Franke *et al.*'s (1990) Social Science Citation Index analysis is based on a list of 17 management journals identified by MacMillan and Stern (1987) as being influential in business policy or strategic management. The three-level hierarchy in the present study is comprised of nine of these 17 journals that meet the criteria described above. Thus, the first tier consists of the number one, two and three journals listed in the Franke *et al.* (1990) study, namely, the *Strategic Management Journal* (SMJ), *Administrative Science Quarterly* (ASQ), and the *Academy of Management Journal* (AMJ). The second tier includes their number seven, eight and nine listings, the *Sloan Management Review* (SMR), *Journal of Management Studies* (JMS), and the *California Management Review* (CMR). Finally, their number 13, 16, and 17 entries, the *Journal of Management* (JM), *Journal of General Man-*

agement (JGM), and *Human Resource Management* (HRM), constitute the third tier.

The publication incidence of replications and extensions was examined by content-analyzing a randomly selected issue from each of these nine journals for every year from 1976 through 1995. This investigation of 176 journal issues produced a total of some 1373 research articles, notes, and commentaries, of which 701 (51.1%) were empirical.

Two of the authors independently classified the 701 empirical papers to ascertain whether they were replications or extensions. To be accorded replication or extension status, an article had to contain a manifest citation of the original work. However, the replicating authors did not have to identify their work as being of a replication nature, as this was the present authors' responsibility. If uncertainty about an article's classification arose, it was reconciled by the present authors. Usually they were included in the count, thus perhaps overestimating the publication frequency of replication research.

Cohen's (1960) kappa (κ) was used as a measure of rater agreement; κ values of 0.75 or more are generally considered to represent excellent agreement among judges. Results for the first ($\kappa = 0.81$, $z = 6.7$, $p < 0.001$), second ($\kappa = 0.78$, $z = 6.8$, $p < 0.001$), and third ($\kappa = 0.83$, $z = 8.2$, $p < 0.001$) levels in the journal hierarchy suggest a strong concordance of opinions among the reviewers.

RESULTS

No strict replications were found in the present sample. Over all three levels of the strategic management journal hierarchy, only 37 of the 701 empirical articles, or 5.3 percent, were replications with extensions.

The first, or trickle-down, hypothesis stated that lower-tier journals would publish more replication research than the top-tier journals. Coincidentally, Table 1 shows some 5.3 percent of empirical papers published in first-tier journals were replications with extensions.³ Only 2.8 percent of

³ Compounding the coincidences, Hubbard and Vetter (1996) also found that 5.3 percent of empirical research published in leading management journals were extensions. It is therefore important to emphasize that they employed a different time period (1970–91), and composition of journals (*Academy of*

Table 1. Replications with extensions in the strategic management journal hierarchy: 1976–95

Journals/ hierarchy ^a	Number of empirical studies	Replications with extensions	Percentage of replications with extensions ^b	95% confidence interval
<i>AMJ</i>	262	11	4.2 (4.0)	1.8–6.6
<i>ASQ</i>	86	3	3.5 (1.9)	–0.4–7.4
<i>SMJ</i>	71	8	11.3 (5.6)	3.9–18.7
First tier	419	22	5.3 (3.6)	3.2–7.4
<i>CMR</i>	40	0	0	
<i>JMS</i>	46	2	4.3 (1.9)	–1.6–10.2
<i>SMR</i>	23	1	4.3 (1.1)	–4.0–12.6
Second tier	109	3	2.8 (0.9)	–0.3–5.9
<i>HRM</i>	33	1	3.0 (0.8)	–2.8–8.8
<i>JGM</i>	24	3	12.5 (2.4)	–0.7–25.7
<i>JM</i>	116	8	6.9 (4.4)	2.3–11.5
Third tier	173	12	6.9 (2.9)	3.1–10.7
Totals	701	37	5.3 (2.6)	3.6–7.0

^aFirst-tier journals: *Academy of Management Journal* (AMJ), *Administrative Science Quarterly* (ASQ), *Strategic Management Journal* (SMJ). Second-tier journals: *California Management Review* (CMR), *Journal of Management Studies* (JMS), *Sloan Management Review* (SMR). Third-tier journals: *Human Resource Management* (HRM), *Journal of General Management* (JGM), *Journal of Management* (JM).

^bValues in parentheses refer to the percentage of journal space, in pages, allocated to replications.

empirical studies in second-tier journals were extensions, while this figure is 6.9 percent for third-tier counterparts. These differences across the three tiers are not statistically significant at the 0.05 level ($\chi^2_{(2)} = 2.40$). Hypothesis 1, therefore, is not supported.

In response to mounting pleas to foster a replication tradition in the business disciplines, the second hypothesis postulated that there should be

an increase in the frequency of published replication research for the period 1986–95 over the period 1976–85. Table 2 shows that whereas this total frequency did increase, from 4.8 percent to 5.7 percent, it was not statistically significant at the 0.05 level ($z = 0.53$). Thus, Hypothesis 2 also is not supported.

It is evident that little replication research is published in the strategic management literature. Why is this the case? Hubbard and Armstrong (1994) presented a number of explanations for this state of affairs in the business (and social science) fields. These include the following: (1) misinterpreting statistical significance levels, especially $p < 0.05$, as a measure of the replicabi-

Management Journal, *Administrative Science Quarterly*, *Journal of Applied Psychology*, and *Organizational Behavior and Human Decision Processes*—formerly *Organizational Behavior and Human Performance*), from those used in this study.

Table 2. Replications with extensions: 1976–85 vs. 1986–95

Journal hierarchy	1976–85				1986–95			
	Number of empirical studies	Replications with extensions	Percentage ^a	95% confidence interval	Number of empirical studies	Replications with extensions	Percentage ^a	95% confidence interval
First tier	218	10	4.6 (3.3)	1.8–7.4	201	12	6.0 (3.9)	2.7–9.3
Second tier	55	1	1.8 (0.6)	–1.7–5.3	54	2	3.7 (1.2)	–1.3–8.7
Third tier	80	6	7.5 (3.4)	1.7–13.3	93	6	6.5 (2.6)	1.5–11.5
Totals	353	17	4.8 (2.5)	2.6–7.0	348	20	5.7 (2.7)	3.3–8.1

^aValues in parentheses refer to the percentage of journal space, in pages, allocated to replications.

lity of a result; (2) assuming that the statistical power of replication studies is low, such that they are less deserving of publication; (3) assuming that information needed to perform a replication with extension is difficult to obtain; (4) considering extensions to be inconsequential; (5) concluding that original works are not worthy of replication; (6) perceiving an editorial-reviewer bias against publishing such work; and, as noted earlier, (7) presuming that replication research is more likely to be published in lower-tier journals.

Hypothesis 3 asserted no differences in the outcomes of replications across the three tiers in the hierarchy. Classification of the outcomes themselves is based on the conclusions reached by the replicating author(s). First-tier journals, in fact, are more likely to publish supportive (40.9%) than conflicting (13.6%) results. Third-tier journals, on the other hand, are more apt to publish conflicting (50%) than supportive (8.3%) findings (Table 3). Perhaps this is because the appearance of too many contradictory results in first-tier journals damages their prestige, whereas this is of lesser concern among third-tier periodicals. In any event, these outcomes are significantly different at the 0.05 level ($\chi^2_{(2)} = 6.58$). All told, some 27 percent of replications published throughout the hierarchy conflict with the results of earlier work. Thus, substantial disagreement exists among empirical findings in the published literature.

Hypothesis 4 predicted no differences in the timeliness of published replication research across the three tiers in the journal hierarchy. This pre-

diction was borne out. There was no statistically significant difference ($F = 1.10$, $p > 0.05$) in the time lags between the original and replication studies in the first (5.9 years), second (9.7 years), and third (7.3 years) tiers.⁴ It is readily apparent, however, that replications are hardly published in a timely manner.⁵

Hypothesis 5 posited no difference in the publication frequency of comments and rejoinders throughout the journal hierarchy. Table 4 reveals that 3.0 percent of all studies published in first-tier journals are devoted to commentaries and rejoinders. Corresponding figures for second- and third-tier journals are 1.3 percent and 4.0 percent, respectively. These differences in results for the three tiers are marginally significant at the 0.05 level ($\chi^2_{(2)} = 6.15$). Hypothesis 5 is not supported. On average across all journal levels, only 2.8 percent of the total number of studies are commentaries. The amount of critical debate regarding the validity of published work in the strategic management area is minimal.⁶

⁴ Because there are only three observations for the second tier, a t -test was also performed on the mean publication time lags for the first and third tiers, but was not significantly different ($t = 0.93$, $p > 0.05$).

⁵ In fairness, it must be acknowledged that replication research can be time-consuming to conduct and write up. Time delays may also be exacerbated because such research frequently encounters editorial-reviewer resistance to publication. As a recent example, consider the difficulties attending the eventual publication of Armstrong, Coviello, and Safranek's (1993) replication paper on escalation bias, reported in Armstrong (1996).

⁶ The statistical power (the probability of rejecting a false null hypothesis) of the five significance tests associated with

Table 3. Outcomes of replications with extensions by journal hierarchy: 1976–95

Journal hierarchy	Support ^{a,b}	Partial support ^{a,b}	Conflict ^{a,b}	Total
First tier	9 (40.9)	10 (45.5)	3 (13.6)	22
Second tier	1 (33.3)	1 (33.3)	1 (33.3)	3
Third tier	1 (8.3)	5 (41.7)	6 (50.0)	12
Totals	11 (29.7)	16 (43.2)	10 (27.0)	37

^aValues in parentheses are percentages.

^bThe classification of replication outcomes—support, partial support, conflict—was based on the conclusions reached by the replicating author(s). The present authors examined the replication studies, weighed the evidence based on the replicating author's findings (and often those of the original paper), and exercised judgment in interpreting these.

Table 4. Published research commentaries by journal hierarchy: 1976–95

Journal hierarchy	Total number of studies	Number of comments/rejoinders	Percentage ^a	95% confidence interval
First tier	501	15	3.0 (0.8)	1.5–4.5
Second tier	452	6	1.3 (0.4)	0.3–2.3
Third tier	420	17	4.0 (2.7)	2.1–5.9
Totals	1373	38	2.8 (1.2)	1.9–3.7

^aValues in parentheses refer to the percentage of journal space, in pages, allocated to commentary.

DISCUSSION

For those scholars concerned about the importance of published replication research in the business fields, the present paper offers little solace. No evidence was found to support the exist-

ence of a trickle-down effect for a sample of journals from the strategic management literature. What was discovered is that only low levels of replication research occur throughout the journal hierarchy. But the results of those replications that do exist often conflict with those of the originals, casting doubt on the reliability of published reports that may not have been subjected to critical examination.

the five hypotheses in this study was also calculated. Standard procedures outlined by Cohen (1988) were followed. That is, the power assessments involved (1) nondirectional tests, (2) $\alpha = 0.05$, and (3) Cohen's conventional definitions of small, medium, and large effect sizes in the population. The average statistical power of the five tests to detect small, medium, and large effects are 0.48, 0.71, and 0.85, respectively. These figures exhibit reasonable power, and compare favorably with those found by Mazen, Hemmasi, and Lewis (1987)—0.23, 0.59, and 0.83—who also used Cohen's procedures in their power analysis of 44 empirical studies published in the strategic management literature. It should also be noted, however, that the main thrust of this paper—that there is a dearth of replication in management—does not depend on the results of the significance tests.

The publication of replication research is critical for establishing the reliability, validity, and generalizability of empirical findings. Yet more empirical findings in the management literature seem to be uncorroborated, fragmented, and isolated. This is not a foundation on which a science can be built or important policy decisions can be made.

Ideally, management practice would be guided by well-developed theories that have been sub-

jected to rigorous and extensive empirical investigation. Instead, Gill and Whittle (1993) contend, a situation of 'management by panacea' prevails. In their opinion, consultant-driven approaches to organizational improvement—such as management by objectives, organization development, and total quality management—enjoy an initial phase of high enthusiasm which is, in turn, followed by a period of disillusionment and replacement by the next stage panacea. Gill and Whittle (1993) maintain that such packaged programs tend to be accepted as an act of faith, in addition to suffering from weak theoretical foundations and a paucity of repeated empirical testing.

In contrast, the physical sciences value replication. For example, Madden, Easley, and Dunn's (1995) survey of natural science editors concluded that they typically endorse replication as a necessary part of research. Moreover, Lindsay and Ehrenberg (1993: 217) observe that in the physical sciences important findings are replicated often, 'first deliberately and then as a built-in part of subsequent work.' A prominent example is rejection of the 'discovery' of cold fusion through the failure to replicate it (Close, 1991; Taubes, 1993). Because of the relative youthfulness of management and the underlying social sciences, it might be expected that replication and extension of previous results would be accorded critical priority (Hubbard and Vetter, 1996; Lindsay and Ehrenberg, 1993). The evidence presented here indicates this is not the case.

The strategic management discipline (and other areas) would be better served by focusing on the search for significant *sameness* among research outcomes, that is, for outcomes that are reproducible under various conditions (Hubbard, 1995; Hubbard and Lindsay, 1995; Lindsay and Ehrenberg, 1993; Nelder, 1986). Imagine, for example, what might be accomplished if members of the discipline routinely replicated and extended studies they deemed to be *important* in their respective areas of expertise. The current reward structure in the field—insufficiently steeped in the tradition of disciplined, probing, structured scientific research—unfortunately stifles this kind of work in favor of 'original' novelty.

That journals emphasize the publication of original research is how it should be. After all, originality confers distinction (Hubbard and Vetter, 1996). But there needs to be a better balance between the publication of 'novel' results

on the one hand, and replication research on the other. A literature dominated by unreplicated findings is of marginal value. Or as Mittelstaedt and Zorn (1984: 14) observe: 'That which isn't worth replicating isn't worth knowing.' The goal of science is empirical generalization, or knowledge development leading to some degree of understanding. Systematically conducted replications and extensions help to achieve this goal, as Ehrenberg and Bound's (1993) work quintessentially reveals.

But some editors and reviewers discourage publication of replications in management and marketing (Kerr, Tolliver, and Petree, 1977; Madden, Easley, and Dunn, 1995). Consequently, conducting and attempting to publish replications may not be career-enhancing, particularly for untenured faculty. Researchers themselves frequently share these biases. (After all, researchers also are editors and reviewers.) Though it would be difficult to imagine many researchers arguing explicitly against the *principle* of replicability, unfortunately some hold a poor opinion of replication research and of those performing it. For example, Kane (1984) noted that choosing to undertake replications is likely to be construed as evidence of intellectual mediocrity, a lack of creativity, or even a bullying spirit in the social sciences (as opposed to a matter of responsibility in the physical sciences).

If the importance of replication as an integral part of the research process is to be established in the strategic management area, it must be inculcated in graduate school. Students currently take an impressive array of methodological courses in research design, multivariate statistics, and the like, but the topic of replication often is ignored. In otherwise laudable state-of-the-art discussions on research methodology (Aldag and Stearns, 1988) and doctoral education in business policy and strategy (Summer *et al.*, 1990), for example, no explicit attention is drawn to the crucial role of replication research.

Changes in the graduate education curriculum aimed at nurturing a deep understanding of the significance of replication would be highly productive. Students should be encouraged to replicate and extend published research papers as part of their training. This is, in fact, a requirement among graduate students in advanced econometrics courses at Ohio State University (Dewald, Thursby, and Anderson, 1986). Similarly, more

Master's theses and Ph.D. dissertations should be awarded for replication research. Such graduate training would not only instill a strong sense of the fundamental importance of replication, but would surely demonstrate that the derogatory term 'mere replication' is undeserved. Replication research need not be unimaginative or menial, but can be challenging and creative (Lindsay and Ehrenberg, 1993).

Systematic replication research, being the scientific vehicle for *deliberately* assessing the scope and limits of empirical findings, is also more basic, more understandable, and more useful than meta-analysis. Sohn (1996) believes that meta-analysis should not be credited as a means of knowledge discovery because its data base, the literature, is untrustworthy. Ironically, encouraging the publication of replications and extensions would alleviate many of the sample-size and construct validity problems that can plague meta-analyses. Replication and meta-analytic research are complementary, not competing, research practices (Allen and Preiss, 1993; Hubbard and Armstrong, 1994). To underscore this point, a meta-analysis may not be replicable (Bullock and Svyantek, 1985; Felson, 1992; Kilpatrick, 1992).

It is further recommended that journals in all levels of the hierarchy, but especially those in the top tier, establish a separate section for publishing replications. Top-tier journals must be actively involved in order to elevate the respect accorded to such work. The argument that space in the leading journals should be largely reserved for original research is, as noted earlier, a truism. What is informative, however, is that present levels of published replications in no way threaten the amount of journal space allocated to original work. Only 5.3 percent of all empirical articles were replications with extensions and, as shown in parentheses in Table 1, they utilized only 2.6 percent of journal pages. There seems to be little interference with the publication of original work, and this would be likely unaffected if attention to replications were doubled or quadrupled.

Top-tier journals must promote the publication of replications. Just as most of the work deemed worthy of replication is likely to be published in the leading journals, so it should be with extensions. Replications appearing in lower-tier journals are less likely to attract attention. Sterling, Rosenbaum, and Weinkam (1995), for example,

cited a study by Begg and Berlin (1988) in which a positive result appearing in a prestigious journal continued to influence medical practice even though subsequent publications in less prestigious ones questioned its reliability.

CONCLUSIONS

Replication and extension are vital to knowledge development. Replication allows us to demonstrate that something really is there, with extensions to see how broadly and in what instances it exists and can be used. Otherwise expressed, systematic replication replaces piecemeal, untested results with useful findings that address practical problems. But few replications appear in business journals.

Replication research should be applauded rather than denigrated. Two suggestions are offered to encourage scientific replications in the business literature. First, changes must be made in graduate training so that the value of replication is emphasized to the point that its practice becomes second nature. Second, journal editorial policies must be aimed at actively soliciting replications with extensions, and sufficient journal space must be reserved specifically for the publication of this necessary work. Without a replication tradition, hopes for developing a cumulative knowledge base in the strategic management area will remain illusory.

ACKNOWLEDGEMENTS

The authors wish to thank two anonymous reviewers for their helpful comments on earlier versions of this manuscript. Any remaining errors are our responsibility.

REFERENCES

- Aldag, R. J. and T. M. Stearns (1988). 'Issues in research methodology', *Journal of Management*, **14**, pp. 253–276.
- Allen, M. and R. Preiss (1993). 'Replication and meta-analysis: A necessary connection', *Journal of Social Behavior and Personality*, **8**, pp. 9–20.
- Armstrong, J. S. (1982). 'Research on scientific journals: Implications for editors and reviewers', *Journal of Forecasting*, **1**, pp. 83–104.

- Armstrong, J. S. (1996). 'Management folklore and management science—on portfolio planning, escalation bias, and such', *Interfaces*, **26**, pp. 25–55.
- Armstrong, J. S. (1997). 'Peer review for journals: Evidence on quality control, fairness, and innovation', *Science and Engineering Ethics*, **3**, pp. 63–84.
- Armstrong, J. S. and F. Collopy (1996). 'Competitor orientation: Effects of objectives and information on managerial decisions and profitability', *Journal of Marketing Research*, **33**, pp. 188–199.
- Armstrong, J. S., N. Coviello and B. Safranek (1993). 'Escalation bias: Does it extend to marketing decisions?', *Journal of the Academy of Marketing Science*, **21**, pp. 247–253.
- Ashenfelter, O. (1986). 'Editorial statement', *American Economic Review*, **76**, p. v.
- Begg, C. B. and J. A. Berlin (1988). 'Publication bias: A problem in interpreting medical data', *Journal of the Royal Statistical Society, B*, **151**, pp. 419–463.
- Blaug, M. (1992). *The Methodology of Economics: Or How Economists Explain*. Cambridge University Press, New York.
- Bozarth, J. D. and R. R. Roberts (1972). 'Signifying significant significance', *American Psychologist*, **27**, pp. 774–775.
- Brown, S. W. and K. A. Coney (1976). 'Building a replication tradition in marketing'. In K. L. Bernhardt (ed.), *Marketing 1776–1976 and Beyond*. American Marketing Association, Chicago, IL, pp. 622–625.
- Bullock, R. J. and D. J. Svyantek (1985). 'Analyzing marketing analysis: Potential problems, an unsuccessful replication, and evaluation criteria', *Journal of Applied Psychology*, **70**, pp. 108–115.
- Cicchetti, D. V. (1991). 'The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation', *Behavioral and Brain Sciences*, **14**, pp. 119–135.
- Close, F. (1991). *Too Hot to Handle: The Race for Cold Fusion*. Princeton University Press, Princeton, NJ.
- Coe, R. and I. Weinstock (1984). 'Evaluating the management journals: A second look', *Academy of Management Journal*, **27**, pp. 660–666.
- Cohen, J. (1960). 'A coefficient of agreement for nominal scales', *Educational and Psychological Measurement*, **20**, pp. 37–46.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Earlbaum, Hillsdale, NJ.
- Dewald, W. G., J. G. Thursby and R. G. Anderson (1986). 'Replication in empirical economics: The Journal of Money, Credit and Banking project', *American Economic Review*, **76**, pp. 587–603.
- Ehrenberg, A. S. C. and J. A. Bound (1993). 'Predictability and prediction (with comments)', *Journal of the Royal Statistical Society, A*, **156**, pp. 167–206.
- Feige, E. L. (1975). 'The consequences of journal editorial policies and a suggestion for revision', *Journal of Political Economy*, **83**, pp. 1291–1295.
- Felson, D. T. (1992). 'Bias in meta-analytic research', *Journal of Clinical Epidemiology*, **45**, pp. 885–892.
- Franke, R. H. (1980). 'Worker productivity at Hawthorne', *American Sociological Review*, **45**, pp. 1006–1027.
- Franke, R. H., T. W. Edlund and F. Oster (1990). 'The development of strategic management: Journal quality and article impact', *Strategic Management Journal*, **11**(3), pp. 243–253.
- Franke, R. H. and J. D. Kaul (1978). 'The Hawthorne experiments: First statistical interpretation', *American Sociological Review*, **43**, pp. 623–643.
- Fuess, S. M., Jr. (1996). 'On replication in business and economics research: The QJBE case', *Quarterly Journal of Business and Economics*, **35**, pp. 3–13.
- Galtung, J. (1967). *Theory and Methods of Social Research*. Columbia University Press, New York.
- Gans, J. S. and G. B. Shepherd (1994). 'How are the mighty fallen: Rejected classic articles by leading economists', *Journal of Economic Perspectives*, **8**, pp. 165–179.
- Gill, J. and S. Whittle (1993). 'Management by panacea: Accounting for transience', *Journal of Management Studies*, **30**, pp. 281–295.
- Greenwald, A. G. (1975). 'Consequences of prejudice against the null hypothesis', *Psychological Bulletin*, **82**, pp. 1–20.
- Hubbard, R. (1995). 'The earth is highly significantly round ($p < 0.0001$)', *American Psychologist*, **50**, p. 1098.
- Hubbard, R. and J. S. Armstrong (1992). 'Are null results becoming an endangered species in marketing?', *Marketing Letters*, **3**, pp. 127–136.
- Hubbard, R. and J. S. Armstrong (1994). 'Replications and extensions in marketing: Rarely published but quite contrary', *International Journal of Research in Marketing*, **11**, pp. 233–248.
- Hubbard, R., R. J. Brodie and J. S. Armstrong (1992). 'Knowledge development in marketing: The role of replication research', *New Zealand Journal of Business*, **14**, pp. 1–12.
- Hubbard, R. and R. M. Lindsay (1995). 'Caveat emptor applies to the consumption of published empirical research results, too', *Management Research News*, **18**, pp. 49–55.
- Hubbard, R. and D. E. Vetter (1996). 'An empirical comparison of published replication research in accounting, economics, finance, management, and marketing', *Journal of Business Research*, **35**, pp. 153–164.
- Kane, E. J. (1984). 'Why journal editors should encourage the replication of applied econometric research', *Quarterly Journal of Business and Economics*, **23**, pp. 3–8.
- Kerr, S., J. Tolliver, and D. Petree (1977). 'Manuscript characteristics which influence acceptance for management and social science journals', *Academy of Management Journal*, **20**, pp. 131–141.
- Kilpatrick, S. J. (1992). 'Environmental tobacco smoke and lung cancer: A critique of a meta-analysis', *The Statistician*, **41**, p. 331.
- Kim, W. S., J. W. Lee and J. C. Francis (1988). 'Investment performance of common stocks in relation to insider ownership', *The Financial Review*, **23**, pp. 53–64.
- Lindsay, R. M. (1994). 'Publication system biases

- associated with the statistical testing paradigm', *Contemporary Accounting Research*, **11**, pp. 33–57.
- Lindsay, R. M. and A. S. C. Ehrenberg (1993). 'The design of replicated studies', *American Statistician*, **47**, pp. 217–228.
- Lovell, M. C. (1983). 'Data mining', *Review of Economics and Statistics*, **65**, pp. 1–12.
- MacMillan, I. C. and I. Stern (1987). 'Delineating a forum for business policy scholars', *Strategic Management Journal*, **8**(2), pp. 183–186.
- Madden, C. S., R. W. Easley and M. G. Dunn (1995). 'How journal editors view replication research', *Journal of Advertising*, **24**, pp. 77–87.
- Mazen, A. M., M. Hemmasi and M. F. Lewis (1987). 'Assessment of statistical power in contemporary strategy research', *Strategic Management Journal*, **8**(4), pp. 403–410.
- McCabe, G. (1984). 'Editor's note', *Quarterly Journal of Business and Economics*, **23**, p. 79.
- Mittelstaedt, R. A. and T. S. Zorn (1984). 'Econometric replication: Lessons from the experimental sciences', *Quarterly Journal of Business and Economics*, **23**, pp. 9–15.
- Monroe, K. B. (1992a). 'Editorial. On replications in consumer research: Part I', *Journal of Consumer Research*, **19**, 2 pages.
- Monroe, K. B. (1992b). 'Editorial. On replication in consumer research: Part II', *Journal of Consumer Research*, **19**, 2 pages.
- Nelder, J. A. (1986). 'Statistics, science, and technology (with comments)', *Journal of the Royal Statistical Society, A*, **149**, pp. 109–121.
- Oswald, S. L. and J. S. Jahera (1991). 'The influence of ownership on performance: An empirical study', *Strategic Management Journal*, **12**(4), pp. 321–326.
- Reid, L. N., L. C. Soley and R. D. Wimmer (1981). 'Replication in advertising: 1977, 1978, 1979', *Journal of Advertising*, **10**, pp. 3–13.
- Robinson, R. B. and J. A. Pearce (1983). 'The impact of formalized strategic planning on financial performance in small organizations', *Strategic Management Journal*, **4**(3), pp. 197–207.
- Rosenthal, R. (1979). 'The "file drawer problem" and tolerance for null results', *Psychological Bulletin*, **86**, pp. 638–641.
- Rosenthal, R. and R. L. Rosnow (1984). *Essentials of Behavioral Research: Methods and Data Analysis*. McGraw-Hill, New York.
- Salancik, G. R. (1986). 'An index of subgroup influence in dependency networks', *Administrative Science Quarterly*, **31**, pp. 194–211.
- Sharplin, A. D. and R. H. Mabry (1985). 'The relative importance of journals used in management research: An alternative ranking', *Human Relations*, **38**, pp. 139–149.
- Sohn, D. (1996). 'Meta-analysis and science', *Theory & Psychology*, **6**, pp. 229–246.
- Sterling, T. D., W. L. Rosenbaum and J. J. Weinkam (1995). 'Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa', *American Statistician*, **49**, pp. 108–112.
- Stewart, W. W. and N. Feder (1987). 'The integrity of the scientific literature', *Nature*, **325**, pp. 207–214.
- Stout, D. E. and J. L. Heck (1995). 'The need for replications and extensions in financial education research: An editorial comment', *Journal of Financial Education*, **21**, pp. 9–11.
- Summer, C. E., R. A. Bettis, I. H. Duhaime, J. H. Grant, D. C. Hambrick, C. C. Snow and C. P. Zeithaml (1990). 'Doctoral education in the field of business policy and strategy', *Journal of Management*, **16**, pp. 361–398.
- Taubes, G. (1993). *Bad Science: The Short Life and Weird Times of Cold Fusion*. Random House, New York.
- Thompson, B. (1994). 'The pivotal role of replication in psychological research: Empirically evaluating the replicability of sample results', *Journal of Personality*, **62**, pp. 157–176.
- Wood, D. R. and R. L. LaForge (1979). 'The impact of comprehensive planning on financial performance', *Academy of Management Journal*, **22**, pp. 516–526.
- Wrege, C. D. and A. G. Perroni (1974). 'Taylor's pig-iron experiments', *Academy of Management Journal*, **17**, pp. 6–27.
- Zinkhan, G. M., M. Y. Jones, S. Gardial and K. K. Cox (1990). 'Methods of knowledge development in marketing and macromarketing', *Journal of Macromarketing*, **10**, pp. 3–17.