

1
2
3
4 AN ASSESSMENT OF RESEARCH
5
6 DESIGNS IN STRATEGIC
7
8 MANAGEMENT RESEARCH:
9
10 THE FREQUENCY OF THREATS
11
12 TO INTERNAL VALIDITY
13
14

15 Donald D. Bergh, Ralph Hanke, Prasad Balkundi,
16
17 Michael Brown and Xianghong Chen
18
19

20
21 **ABSTRACT**

22
23 *The authors content analyze 76 empirical Strategic Management Journal arti-*
24 *cles to determine how studies control for threats to internal validity, a common*
25 *source of flaws in research designs. Results indicate that most studies fail*
26 *to control for one or more threats to internal validity. In particular, selection*
27 *effects were the most frequently appearing threat, followed by history effects,*
28 *ambiguity about the direction of causal inference, changing data sources and*
29 *subject mortality. In general, the results suggest that strategy researchers*
30 *need to more carefully account for threats to the internal validity of their*
31 *research designs. Suggestions for addressing these problems are provided.*

32
33 A flaw is a crack or a defect (*Webster's Expanded Dictionary*, 1991). In empirical
34 research, flaws have serious implications, especially when they reside within
35 research designs. For example, if a research design contains a flaw, then other

36
37 **Research Methodology in Strategy and Management**
38 **Research Methodology in Strategy and Management, Volume 1, 349–371**
39 **Copyright © 2004 by Elsevier Ltd.**
40 **All rights of reproduction in any form reserved**
ISSN: 1479-8387/doi:10.1016/S1479-8387(04)01112-9

1 aspects of an empirical study – such as measures, data and analytical results –
2 will likely reflect the flaw (Campbell & Stanley, 1963; Cook & Campbell, 1979;
3 Pedhazur & Schmelkin, 1991). In addition, flaws in research designs can lead
4 to false inferences, obscure true relationships and even cancel out actual results
5 (Cook & Campbell, 1979). Flawed research designs are important, and there is a
6 need to understand how they arise, whether they are random or recurrent and what
7 they might mean for empirical findings and theory development. While researchers
8 generally observe a value that “no study is perfect,” nonetheless, failure to consider
9 flaws in research designs is to allow them – and their effects – to perpetuate.

10 Prior research has documented the types and trends of research designs
11 (Grunow, 1995; Stone-Romero, Weaver & Glenor, 1995). No studies, however,
12 have identified and considered flaws in research designs, an omission our study
13 attempts to address. Our study focuses on a common source of flaws in research
14 designs, threats to internal validity (Campbell & Stanley, 1963; Cook & Campbell,
15 1979), and examines whether researchers account for these sources of flaws. This
16 is an important subject to study because if such threats occur non-randomly, and
17 we can identify the most commonly arising ones, then we can offer assessments
18 of what these threats might mean for theory development while also provide
19 guidance for how research can be improved.

20 We focus on the research designs of strategic management research. We chose
21 strategic management research because this area has come under significant critical
22 evaluation, with assessments of such methodological subjects as measurement,
23 analysis and integration (Bergh, 1995; Bowen & Wiersema, 1999; Ketchen &
24 Shook, 1996). Our review would help contribute to a comprehensive evaluation of
25 methodological practices in the strategic management field.

26 We begin by reviewing the common threats to internal validity and how they
27 can lead to flaws. We then examine a sample of 76 empirical studies published in
28 the *Strategic Management Journal* to see whether threats to internal validity are
29 controlled and what, if any, flaws might correspondingly exist. We then present our
30 main findings: that threats to internal validity are relatively common in empirical re-
31 search and that some tend to recur across studies. We find that some threats are par-
32 ticular to whether a research design is cross-sectional or longitudinal, and we find
33 that one threat tends to exist across all design types. Overall, the results suggest that
34 flaws are nonrandom. The flaws and their implications are discussed, and we offer
35 suggestions as to how researchers can eliminate them from their research designs.

37 38 **THREATS TO INTERNAL VALIDITY**

39
40 Flaws in research designs are associated most with internal validity (Campbell &
Stanley, 1963; Cook & Campbell, 1979). Internal validity is the “extent to which

1 the results of a study can be attributed to treatments [variables] rather than flaws
2 in the research design” (Vogt, 1993, p. 114). Studies with high internal validity
3 provide results that are not subject to flaws whereas designs with low internal
4 validity produce results that are subject to third-variable effects and confounds
5 (Campbell & Stanley, 1963; Cook & Campbell, 1979; Spector, 1991). Internal
6 validity is most affected by how research designs account for several factors,
7 including history, maturation, testing, instrumentation, regression, selection,
8 mortality, selection-interaction effects and ambiguity about the direction of causal
9 inference (Campbell, 1957; Campbell & Stanley, 1963; Cook & Campbell, 1979;
10 Spector, 1981).

11 The failure to control for any one of these factors, known as “threats to internal
12 validity,” can produce flaws.

13 *History.* A history effect can arise when specific events occur between the first
14 and second measurement that are in addition to the experimental or independent
15 variable (Campbell & Stanley, 1963). History poses a threat to internal validity
16 “when an observed effect might be due to an event which takes place between the
17 pretest and the posttest, when this event is not the treatment of research interest”
18 (Cook & Campbell, 1979, p. 51). Research designs that employ a “one-shot”
19 (cross-sectional) and/or a simple change design (pretest, posttest), are especially
20 vulnerable to history effects. Campbell and Stanley observe “[I]f the pretest (O_1)
21 and the posttest (O_2) are made on different days, then the events in-between may
22 have caused the difference” (1963, p. 7). In order to rule out such threats, the
23 researcher has to make the case that history is implausible, or, alternatively, that
24 history is plausible but did not occur. Theoretical arguments are needed for the first
25 case while data are usually required for making the second case. If history effects
26 cannot be controlled for, “then the researcher has to admit that he or she cannot
27 draw confident causal conclusions because a frequently plausible threat cannot be
28 ruled out” (Cook & Campbell, 1979, p. 100).

29 *Maturation.* Alternative explanations due to maturation can arise when an
30 observed effect might be due to the research subjects’ changing over the study
31 period and when those changes are not attributed to the study setting. Campbell
32 and Stanley refer to maturation as “those biological or psychological processes
33 which systematically vary with the passage of time, independent of specific
34 external events” (1963, pp. 7–8). For example, if the respondents become “older,
35 wiser, stronger, more experienced, and the like between the pretest and posttest
36 and when this maturation is a not the treatment of research interest” (Cook &
37 Campbell, 1979, p. 52). Maturation effects are most evident in pretest-posttest
38 research designs, such as in the case of a simple change design. In these situations,
39 the posttest might be artificially higher or lower because of subject maturation.
40 Hence, if not controlled for, maturation can lead to changes in posttest measures
that may not be associated with the hypothesized effects. Like history effects,

1 maturation creates an alternative explanation that must be ruled out with study
2 variables, study setting, or data.

3 *Testing.* Familiarity with a test can sometimes affect responses to subsequent
4 administration of the test. Testing is a threat to internal validity when “an effect
5 might be due to the number of times particular responses are measured” (Cook &
6 Campbell, 1979, p. 52). These authors further note that “exposure to an outcome
7 measure at one time can lead to shifts in performance at another” (1979, p. 102).
8 Testing effects are of particular relevance to pretest-posttest designs, as the
9 posttest might contain a learning or experience artifact that may fall outside the
10 hypothesized relationships. In such cases, the researchers need to control for
11 the gain or loss associated with testing. An alternative approach to controlling for
12 testing effects is to utilize control groups. With designs that employ experimental
13 and control groups, the effects of testing (and maturation) should be evident
14 in both groups (Campbell & Stanley, 1963, p. 14). If left unchecked, however,
15 testing may provide an alternative explanation for why studies find an increase or
16 decrease in their subjects between a pretest and posttest.

17 *Instrumentation.* An instrumentation effect is created when either a measuring
18 instrument is changed (such as the scales or variables) or when the observers or
19 scorers change over the course of the study (Campbell & Stanley, 1963). Also
20 referred to as “instrument decay” (Campbell, 1957), instrumentation refers to
21 autonomous changes in the measuring instrument which might account for a
22 change between pretest-posttest. Instrumentation is involved when a test metric
23 changes or observers become more experienced between a pretest and posttest.
24 Note that instrumentation differs from maturation, in that in the former case, the
25 focus is on the administrators of the study whereas in the latter case, attention is
26 directed on the subjects of the investigation. For example, as Campbell and Stanley
27 observe, “[i]f essays are being graded, then grading standards may shift between
28 O_1 and O_2 If parents are being interviewed, the interviewer’s familiarity with
29 the interview schedule and with the particular parents may produce shifts. A
30 change in observers between O_1 and O_2 could cause a difference” (1963, p. 9).
31 Thus, failure to recognize changes in the measures, instruments or observers could
32 result in a rival explanation of the study findings. The effects of instrumentation
33 can be controlled for, namely by adapting for measure differences or by assigning
34 observers randomly.

35 *Regression.* Statistical regression is a threat to internal validity when changes in
36 values are associated with a regression toward-the-mean effect. More specifically,
37 statistical regression effects are created when study subjects are chosen on the
38 basis of “extreme scores.” For example, if subjects are selected for study because
39 they performed poorly on a particular test, which becomes O_1 for them, then on
40 subsequent testing, O_2 for this group will almost surely average higher. This “result

1 is not due to any genuine effect of X [the independent variable or treatment] . . .
2 It is rather a tautological aspect of the imperfect correlation between O_1 and O_2
3 “(Campbell & Stanley, 1963, p. 5). Cook and Campbell go one step further, noting
4 that statistical regression: “(1) operates to increase obtained pretest-posttest gain
5 scores among low pretest scores, since this group’s pretest scores are more likely
6 to have been depressed by error; (2) operates to decrease obtained change scores
7 among persons with high pretest scores, since their pretest scores are likely to
8 have been inflated by error; and (3) does not affect obtained change scores among
9 scorers at the center of the pretest distribution, since the group is likely to contain
10 as many units whose pretest scores are inflated by error as units whose pretest
11 scores are deflated by it” (1979, pp. 52–53). This regression-toward-the-mean
12 effect can therefore represent an alternative explanation of study findings and is
13 especially common when subjects are selected because of their extreme scores.
14 It is remedied most effectively by selecting study subjects randomly, with no
15 regard to scores on a pretest measure, and by adding a third posttest measure
16 (Rogosa, 1988).

17 *Selection.* A selection effect can arise when study subjects are selected because
18 they possess a characteristic that is related to the independent or dependent
19 variables. Frequently referred to as “self-selection,” “selecting on the dependent
20 variable,” or “bias in the assignment of subjects to treatments,” selection occurs
21 when the “members of the groups being studied are in the groups, in part, because
22 they differentially possess traits or characteristics . . . that possibly influence or
23 are otherwise related to the variables of the research problem” (Kerlinger, 1986,
24 p. 349). Selection is a “problem that may arise in the comparison of groups
25 when the groups are formed by individuals who choose to join them and thus
26 are not formed by a researcher assigning subjects to . . . groups” Vogt (1993,
27 p. 207). By selecting particular subjects into a study, therefore, the relationships
28 and effects observed may reflect those subjects only. As such, selection is a
29 threat to internal validity “when an effect may be due to the difference between
30 the kinds of people in one . . . group as opposed to another” (Cook & Campbell,
31 1979, p. 53).

32 Selection is most evident in cross-sectional and group comparison designs.
33 For example, it occurs when differences between two groups can be attributed to
34 the composition of one of the groups. For example, “[i]f O_1 and O_2 differ, this
35 difference could well have come about through the differential recruitment of
36 persons making up the groups: the groups might have differed anyway, without
37 the occurrence of X” (Campbell & Stanley, 1963, p. 12). Selection can be ruled
38 out as a rival explanation when randomization has assured group equality.

39 *Mortality.* Also known as attrition, or losing subjects in the course of a study,
40 a mortality effect can arise when a result may be due to the subjects that leave or

1 drop out of a study. This effect results in a “selection artifact,” since the study is
2 then composed of different kinds of subjects at the posttest (Cook & Campbell,
3 1979). Mortality is a threat to internal validity when findings are due to differential
4 attrition of study subjects (Campbell & Stanley, 1963). The effects of mortality
5 are difficult to estimate. Nevertheless, in some research designs, especially
6 those involving experiments with control groups and pretest-posttest structures,
7 mortality can be modeled in the analysis and its effects empirically measured.
8 Another solution is to manipulate the observation periods and determine whether
9 effects vary over time, and whether those effects are related to attrition and
10 retention. However, in most cases, the mortality effects must be estimated by com-
11 paring the characteristics of the lost subjects with those that remain. Differences
12 can be identified and their implications considered as alternative explanations
13 of the study findings.

14 *Interactions with selection.* Some of the foregoing threats to internal validity,
15 especially history, maturation and instrumentation, can interact with selection
16 to produce additional threats to internal validity. For example, “[S]election-
17 maturation results when experimental groups are maturing at different speeds . . .
18 Selection history . . . results from the various treatment groups coming from
19 different settings so that each group could experience a unique local history
20 that might affect outcome variables . . . Selection-instrumentation occurs when
21 different groups score at different mean positions on a test whose intervals are not
22 equal” (Cook & Campbell, 1979, p. 53). More specifically, selecting subjects on
23 the basis of an historical event (selection-history), on the basis of age, experience,
24 or life-cycle (selection-maturation), or on the availability of a data source
25 (selection-instrumentation) can produce settings whereby three or more threats to
26 internal validity are simultaneously present. In the case of selection-history, we
27 may find a design that contains selection effects, history effects, and the unique
28 interaction between the two. As a consequence, the potential for alternative
29 explanations rises significantly in such circumstances.

30 *Ambiguity about direction of causal inference.* This threat to internal validity
31 can arise when the temporal precedence or antecedence among relationships is
32 unclear. It prevents researchers from knowing which variable causes what effects
33 and can lead to tautologies in the interpretation of the findings. These problems
34 are most likely to arise in one time-period study designs (i.e. the “one-shot”
35 snapshot design). Controls for the effects of ambiguity about the direction of
36 causal inference are easiest to obtain in experimental settings, when one direction
37 of the causal inference is implausible, or when data are collected at multiple time
38 periods (Cook & Campbell, 1979). However, in many designs, it is impossible
39 to determine an ordering of the relationships, thus raising the likelihood that
40 alternative explanations may account for the effects.

THREATS TO INTERNAL VALIDITY IN STRATEGY RESEARCH

To identify the presence of flaws in research designs in strategic management research, we conducted a content analysis of empirical studies appearing in regular and special issues of the *Strategic Management Journal* from 1994 through 1998. The *SMJ* was used because it contains articles on only strategic management topics, thus minimizing guesswork in defining what should and should not be considered a strategic management study (Bergh & Holbein, 1997; Ferguson & Ketchen, 1999). The *Journal* published 211 empirical articles during that period, of which 203 were of the non-experimental design type; that is, subjects were not assigned randomly nor were the independent variables manipulated. Of the remaining studies, 4 employed an experimental design (subjects were assigned randomly and independent variables were manipulated) and 4 used a quasi-experimental design (no random assignment of subjects, but independent variables were manipulated) (Spector, 1981).

We chose this time period because it is relatively recent, giving us the opportunity to see how research designs are applied in current settings. The time period spans five years, providing a long enough interval for representing research within the field of strategic management.

Sample

We randomly selected 90 out of the 203 non-experimental design types, as these types of designs constitute the vast majority used in strategic management research. It was important to ensure consistency in design selection because experiments, quasi-experiments and non-experiments can have different threats to internal validity (Campbell & Stanley, 1963; Cook & Campbell, 1979). The selection of 90 studies would provide a reasonable representation of the 203 empirical studies, and because assessing research designs is a time-intensive venture, only a limited number of articles could be effectively evaluated. The research designs of those 90 empirical non-experimental studies were the unit of analysis for our assessment.

Coding

Each of the 90 articles was reviewed to determine whether researchers accounted for threats to internal validity. The relevant threats to the internal validity of

1 cross-sectional designs were history, maturation, testing, selection, selection
2 interactions with history, maturation, instrumentation, and ambiguity about causal
3 inference. The relevant threats to the internal validity of longitudinal designs were
4 history, maturation, testing, instrumentation, regression, selection, mortality, and
5 selection interactions with history, mortality, and testing (Campbell & Stanley,
6 1963; Cook & Campbell, 1979).

7 The coding process involved several stages. (1) The 90 articles were divided
8 among four doctoral students, each of whom had completed a one-year course on
9 research methods. (2) The students identified the structures of the research designs
10 of their assigned studies. Following Campbell and Stanley (1963) and Cook and
11 Campbell (1979), they designated independent variables as “X’s” and dependent
12 variables as “O’s.” They assessed the sampling and measurement and in some cases,
13 analysis and results sections of their articles to determine how many times each
14 variable had been measured in the studies. During this process, 2 studies were non-
15 empirical and were dropped from further consideration, as our focus was on em-
16 pirical studies only. Also, 11 studies were not described sufficiently for evaluation
17 and we were forced to drop them from the study. (3) Each research design of the 77
18 remaining studies was evaluated to determine whether the designs had controlled
19 for the relevant and applicable threats to internal validity. Threats are addressed
20 when control variables are used to reduce their applicability or when the design is
21 structured so as to reduce the threat. For example, the threat of history is addressed
22 when researchers account for factors occurring between two periods and/or when
23 they use a design feature (such as a short time interval between two observation
24 points). Threats were coded as “0” if not applicable to the study design, as “1” if
25 applicable, but not addressed, and “2” if applicable and addressed. The definitions,
26 criteria and assessments of threats to internal validity are reported in Table 1. The
27 Appendix reports how the coding process was performed. The content analysis of
28 four studies is shown so to provide insights into researchers tend to address research
29 design issues.

30 Reliability of the coding values was determined using a retest and percent
31 agreement evaluation (Carmines & Zeller, 1979; Jones, Johnson, Butler & Main,
32 1983). Half of each student’s studies were identified randomly and exchanged
33 with other students for recoding. The students independently repeated the coding
34 process described above. The codes for the exchanged studies were compared.
35 The percent of agreement between the original coding and the retest was 94%.
36 The discrepant cases were discussed and agreement was reached on all but one.
37 This process for computing inter-rater reliability is appropriate for assessing the
38 coding of dichotomous and nominal variables, such as those in our study, and
39 offers reliability estimates that are consistent with alternative approaches, such as
40 Kappa (see Jones et al., 1983, p. 514). Note that we subsequently discarded the

Table 1. Threats to Internal Validity.

Threat	Definition	Operationalization
History	When events occur between measurement periods	Control variables, features in research design
Maturation	When effect may be due to variation in age or experience	Control variables, random sampling
Testing	Familiarity with test	Control for gain/loss associated with testing
Instrumentation	When data source, metrics or coders change	Examination of data sources over study period
Regression	When subjects are selected on the basis of extreme scores	Three or more observation points, random sampling
Selection	Subjects are selected because they possess a trait related to study variables	Comparisons between respondents and non-respondents
Mortality	Differential loss of study subjects	Comparisons between retained and lost subjects
Ambiguity about causal inference	When temporal precedence among relationships is unclear	Inspection of temporal precedence of data periods
Selection – maturation	Selection of subjects on the basis of life cycle, size	Evaluation of sampling criteria, sample dimensions
Selection – history	Selection of subjects on the basis of an event of interest	Evaluation of sampling criteria, sample dimensions
Selection – mortality	Selection of subjects on the basis of retention or loss	Evaluation of sampling criteria, sample dimensions
Selection – testing	Selection of subjects on basis of test results	Evaluation of sampling criteria, sample dimensions
Selection – instrumentation	Selection of subjects on basis of data source	Evaluation of sampling criteria, sample dimensions

study in which agreement could not be reached from our sample, reducing our final sample size to 76 studies.

Analysis

For each of the 76 studies, each threat to internal validity was coded using the value described above. Counts of these values were summed, by threat, for the cross-sectional and longitudinal studies separately. Percentages of each code value were determined. This process allowed us to determine the percentage of study designs (e.g. cross-sectional or longitudinal) that addressed or did not address each particular threat to internal validity. In addition, correlation analysis was used to

1 identify patterns of code values across the study designs, and whether some threats
2 to internal validity were evident in multiple studies simultaneously.

3 4 5 *Findings*

6
7 The results of the content analysis reveal that two primary types of research
8 designs were present in the 76 studies. First, there were 35 cross-sectional studies,
9 which took either of two forms, where researchers collected the independent (X)
10 and dependent variable (O) at the same time (X/O) or with a temporal lag (XO).
11 The distinguishing feature of the cross-sectional (or “one-shot”) design, is the
12 independent and dependent variables are each observed on one occasion only.
13 Second, there were 41 time series and longitudinal studies. These studies appeared
14 in three general types: 0X0, X00, or X/O X/O X/O (note that more X and O’s may
15 have been used, but the design types did not change). These variations within the
16 cross-sectional and time series and longitudinal designs are important, as the appli-
17 cability of threats to internal validity vary relative to the design types. For example,
18 within the time series and longitudinal designs, the threats that apply to a panel
19 design (X/O X/O X/O, etc.) are different than those that apply to a before and after
20 design (00X00).

21 *Cross-sectional study designs.* We found that most pertinent threats to internal
22 validity were not controlled for in these research designs (Table 2). The most
23 frequently occurring threat to internal validity that was not controlled was the
24 ambiguity about causal inference, as 27 of 35 (77.1%) cross-sectional studies were
25 unable to rule this threat out of their research designs. This threat arose when study
26 designs consisted of data collected at the exact same time. Some cross-sectional
27 studies (8 of 35, 22.9%) included a time lag between the independent and
28 dependent variables, and these kinds of designs provide a guard against the threat
29

30 **Table 2.** Threats to the Internal Validity of Cross-Sectional Research Designs.

31 Threat	32 Controlled (%)	32 Not Controlled (%)
33 History	48.6	51.4
34 Maturation	57.1	42.9
35 Selection	25.7	74.3
36 Testing	0.0	0.0
37 Selection – history	9.0	91.0
38 Selection – maturation	15.0	85.0
39 Selection – testing	8.0	92.0
Ambiguity about causal inference	22.9	77.1

40 *Note:* Percentages based on design types where the threats were applicable ($N = 35$).

1 of ambiguities of causal inference. It therefore appears that most cross-sectional
2 designs cannot discern the direction of causal inference.

3 The second most frequent threat to internal validity that was not controlled for
4 was selection, as 26 of 35 studies (74.3%) contained selection threats. This finding
5 means that the subjects of most cross-sectional studies were selected because
6 they possessed a characteristic or trait that was related to the study variables. For
7 example, some subjects (companies) were selected on the basis of a dependent
8 variable in the study design, or whether they possessed some strategic attribute
9 that was important in the study. In addition, most of the 26 studies with selection
10 threats are vulnerable to interactions with other threats (92% include testing,
11 91% history, 85% maturation). These results indicate that the majority of subjects
12 in cross-sectional studies were selected because of an event occurring to the
13 subject, the subject possessed some test result (e.g. high performing firms), or had
14 some changing attribute. Such practices could create flaws whereby results are
15 vulnerable to characteristics and differences associated with the selected subjects
16 in addition to any effects of the independent variables.

17 Further, just over half of the studies (51.4%) did not control for threats due
18 to history effects. This means that events occurring at or between the periods
19 of observation were not recognized, and results may have been obscured due to
20 those events.

21 Correlation analysis indicates that some threats to internal validity were related
22 across the studies. In particular, history and maturation were correlated highly
23 ($r = 0.66$, $p < 0.001$), indicating that studies that failed to control for history
24 tended to also fail to control for maturation.

25 *Time-series and longitudinal study designs.* These study designs tended not to
26 control for several threats to internal validity (Table 3). More specifically, 38 of
27 41 (95%) of the time series studies did not control for selection. In addition, of
28 those 38 studies, most contained interaction effects with mortality (100%), testing
29 (88%), maturation (69%), and history (75%). As a consequence, these studies
30 could have results that are vulnerable to characteristics and traits of the subjects.
31 Namely, studies tended not to control for differences between subjects that were
32 retained or lost. They also tended not to account for selection on particular scores,
33 changes, or events occurring during the study period.

34 Further, most study designs (62%) did not control for changes in measurement
35 instruments, coders or variable metrics over time, suggesting that changes in
36 results may be due to instruments or metrics rather than just changes in subjects
37 or study conditions.

38 On the positive side, most time series and longitudinal study designs did control
39 for threats to history, maturation, and regression. In particular, regression effects
40 were minimized through the use of measures of subjects at three or more periods,

Table 3. Threats to the Internal Validity of Time Series and Longitudinal Research Designs.

Threat	Controlled (%)	Not Controlled (%)
History	56.1	43.9
Maturation	56.1	43.9
Testing	0.0	100.0
Instrumentation	19.0	81.0
Regression	100.0	0
Mortality	45.0	55.0
Selection	5.0	95.0
Selection – history	25.0	75.0
Selection – maturation	31.0	69.0
Selection – testing	12.0	88.0
Selection – mortality	0	100.0

Note: Percentages based on design types where the threats were applicable ($N = 41$).

thus ruling out any regression effects associated with an extremely high or low initial score.

Correlation analysis indicates that some threats to internal validity were related in longitudinal study designs. For example, as with the cross-sectional designs, the failure to control for history and maturation threats to internal validity were correlated highly ($r = 0.51, p < 0.001$).

DISCUSSION

Flaws are an important concern for researchers, as they have the potential to produce erroneous conclusions, may prevent fair testing of hypotheses and could thereby inhibit cumulative theory development. Consequently, investigation into the sources and presence of flaws in research designs is needed, as well as a consideration of their potential implications. However, no prior research has yet documented the existence of flaws in research designs. We do not know if research designs contain flaws and whether results are therefore vulnerable to rival explanations, confounds or third-variable effects. While we might assume that the peer review process of publication identifies and eliminates studies with the most serious of flaws, how confident can we be that flaws are actually ruled out when there has not yet been a conversation about them? Accordingly, our study attempts to identify the presence of flaws in research designs.

Our study results indicate that researchers have addressed some threats to the internal validity of research designs. In particular, maturation is not a threat to

1 cross-sectional or longitudinal designs, and history and regression are generally
2 controlled for in these latter types of designs as well. However, although a few
3 studies addressed each threat to internal validity, most did not, and some threats
4 were pervasive across the reviewed studies. Indeed, while strategy researchers tend
5 to control for alternative theoretical explanations of their empirical results (e.g.
6 effects attributable to debt, owners, slack, industry effects, etc.), they appear
7 to be much less vigilant about threats to internal that may reside within their
8 study designs.

9 Our data indicate that the most prevalent, non-controlled threat to internal
10 validity is selection. This threat exists in most studies, irrespective of design type.
11 It occurs when researchers select samples on the basis of particular characteristics
12 that are related to the independent and/or dependent variables. In other words,
13 subjects may be selected for study because they are most likely to reflect the
14 hypothesized relationships, thereby providing a higher likelihood that support
15 for the expected associations will be found. In many circumstances, researchers
16 select samples on the basis of a variable they are trying to predict. In other cases,
17 companies and organizations are selected for study because they possess traits or
18 characteristics of interest, traits that may be related to the variables or research
19 problem, thus resulting in a selection effect (Kerlinger, 1986). In addition, selec-
20 tion threats often occur on the basis of another threat to internal validity. Study
21 subjects may be selected on history (event of interest), maturation (particular
22 age, change, size or life cycle), data source (testing results, such as high or low
23 performer), or mortality (survival). In these cases, researchers are creating multiple
24 selection threats.

25 The implication for flaws is straightforward; by “selecting particular subjects
26 into a study, the relationships and effects may reflect those subjects only . . . [and
27 results] may be due to the difference between the kinds of people in one . . . group
28 as opposed to another” (Cook & Campbell, 1979, p. 53). This suggests that
29 results may be due to the characteristics of the subjects in the study rather than
30 predicted relationships between the variables that are measuring the subjects.
31 Hence, selection creates a “bias in the assignment of subjects to treatment groups”
32 (Spector, 1981, p. 18) and the results may be more attributable to that bias than
33 the hypothesized relationships. Results may therefore reflect attributes of the
34 subjects rather than the theoretical relationships.

35 Alternatively, with a sample that is not derived from the independent or de-
36 pendent variable, then the research design controls for the alternative explanation
37 that an empirical result is due to the sampled firms possessing the hypothesized
38 criteria. However, such designs are infrequently used in strategy research. By
39 selecting samples on the basis of their values with respect to independent or
40 dependent variables, many researchers may be inadvertently “stuffing the ballot

1 box” with results likely to support their expectations. In these cases, they do
2 not know why the hypothesized linkages may have been supported – was it
3 because the theory was correct or because the data were likely to conform to the
4 predictions? And, as our results indicate, flaws due to selection threats take many
5 different forms.

6 In addition to selection effects, history may affect empirical results, especially
7 for cross-sectional designs. This means that, for cross-sectional study designs,
8 events not accounted for may also influence the empirical relationships. For
9 example, when attempting to explain the effects of strategy on performance,
10 researchers using cross-sectional research designs do not generally account for
11 confounding historical events, such as restructuring charges and changes in
12 accounting conventions. By not recognizing such events, then results become
13 vulnerable to alternative explanations, and results may be difficult to compare
14 across studies (e.g. Bowen & Wiersema, 1999).

15 History was less of a problem with longitudinal studies, as it was generally
16 controlled for in these types of designs. This finding is different than McWilliams
17 and Siegel’s (1997) claim that time series designs (event studies in their instance)
18 do not usually control for possible confounding effects. Although we agree with
19 the general suggestion by McWilliams and Siegel that efforts should be made to
20 control for extenuating historical events, we do not find evidence consistent with
21 their claim that longitudinal study designs do not address these effects. To the
22 contrary, our results show that most – though not all – studies make an effort to
23 minimize the confounding effects of historical events. This difference in results
24 may be due to the types of studies that they, and we, considered. They focused
25 on event studies and window lengths, whereas we considered those designs and
26 other types of longitudinal ones. The dissimilarities in findings is probably due to
27 these differences in the domain of design types.

28 We also find evidence of several threats in time series designs. In particular,
29 over half of the time series studies were vulnerable to instrumentation threats,
30 threats that are generally not controlled for in the designs. This means that
31 changes in instruments, data sources or observers may produce changes in the
32 independent variables, resulting in additional alternative explanations. Spector
33 notes that “[v]ariables that are related to the independent variables may account
34 for observed differences in the dependent variable, giving the false appearance of
35 a valid answer to a research question” (1981, p. 24). Therefore, to the extent that
36 instruments influence independent variables, and those instruments change, then
37 time series studies are vulnerable to another threat to internal validity.

38 Moreover, we find that most time series designs did not control for the effects
39 of mortality and as a result, tested subjects that survived the time period. To the
40

1 extent that subjects that survive during a study period are different than those that
2 do not, then empirical results would be biased and theoretical explanations would
3 correspondingly reflect the companies that stay in the study only.
4

6 *Re-direction*

7
8 Given such findings, it seems that researchers need to revamp the use of the cross-
9 sectional study design. While the discontinued use of these designs – advocated by
10 [Campbell and Stanley \(1963\)](#) and [Cook and Campbell \(1979\)](#) – may be impractical,
11 we suggest that researchers consider using the lagged version (X0) as opposed
12 to the more common simultaneous design type (X/0). The lagged version helps
13 control for the threat of ambiguity about causal inference, a persistent and stubborn
14 problem plaguing cross-sectional designs. We also believe that cross-sectional
15 designs may play an important role in strategic management, especially in cases
16 where researchers are documenting associations and correlation, so their careful
17 application in the field is warranted. Such care should also be exercised in ensuring
18 that history effects do not occur in the lagged cross-sectional design.

19 The results also suggest that researchers need to reconsider sample selection
20 methodologies. Researchers are urged to evaluate how sampling decisions affect
21 study data and hypothesis testing. It is critical that researchers randomize, provide
22 evidence of differences between respondents and non-respondents and survivors
23 and non-survivors, and demonstrate how the subjects provide a fair test of the
24 hypotheses. Such tests mean that the study data provide equal odds of not finding
25 a relationship as of finding one, and that the study subjects were identified so
26 to provide the equivalence of those odds. In other words, the subjects must have
27 equal probability of having – and not having – a hypothesized relationship.

28 Finally, a general increase in the level of rigor is needed. The threat of
29 instrumentation – such as changes in the measuring instrument, coders or variable
30 approaches – can be easily assessed and weighed. For example, the effects of
31 changing data sources during a longitudinal study can be evaluated by comparing
32 the data sources for some year of overlap, and evidence can be provided of the
33 similarities and differences, which can then be factored into the results. Along
34 these lines, researchers need to apply recent methodological developments. Most
35 prominent of the developments that were not applied was common methods
36 variance. Even though the effects of this problem have been documented over
37 the years, its persistence is puzzling. The rigor of empirical research, and indeed,
38 freedom from alternative explanations, can be improved by discontinuing the use
39 of common methods for data collection.
40

Limitations

As with all empirical studies, ours has limitations that influence the generalizability of the findings. First, we focused on non-experimental studies only, so our results are limited to these types of studies. We cannot offer inferences to other design types. Second, our study may also suffer from a selection and time period bias. We selected articles published in a top-tier strategy journal, during the mid-to-late 1990s. To the extent that our selection process and study time period might create bias, then flaws in our own design may influence the validity of our conclusions. Third, our study documents the presence of threats to internal validity. It does not predict or explain the attendant flaws associated with those threats. No inferences can be attributed to any study in particular. Finally, we recognize that our own assessment protocol is inherently subjective. To the degree that we applied our evaluation in manner inconsistent with its basis, our results may be duly affected.

CONCLUSION

Reducing flaws in research designs is essential for rigorous and meaningful empirical results. Flawed designs lead to confounds biases, and rival explanations. It is difficult to imagine a more relevant subject within research methodology. And while studies in the social sciences are especially vulnerable to threats to internal validity (no study is truly perfect) we need to know which threats to internal validity are likely to exist and what they mean for empirical results and theory development. Our results show that strategic management researchers tend to control for some threats, but that others appear to exist systematically. Identification and consideration of those threats, and their attendant flaws, can help researchers as they design and execute their studies. The implications for theory development are simply too important to ignore.

REFERENCES

- Bergh, D. D. (1995). Problems with repeated measures analysis: Demonstration with a study of the diversification and performance relationship. *Academy of Management Journal*, 38, 1692–1708.
- Bergh, D. D., & Holbein, G. F. (1997). Assessment and redirection of longitudinal analysis: Demonstration with a study of the diversification and divestiture relationship. *Strategic Management Journal*, 18, 557–572.

- 1 Bowen, H. P., & Wiersema, M. F. (1999). Matching method to paradigm in strategy research:
2 Limitations of cross-sectional analysis and some methodological alternatives. *Strategic*
3 *Management Journal*, 20, 625–636.
- 4 Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological*
5 *Bulletin*, 54, 297–312.
- 6 Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*.
7 Skokie, IL: Rand McNalley.
- 8 Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Newbury Park, CA: Sage.
- 9 Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis for field studies*.
10 Skokie, IL: Rand McNally.
- 11 Ferguson, T. D., & Ketchen, D. J., Jr. (1999). Organizational configurations and performance:
12 The role of statistical power in extant research. *Strategic Management Journal*, 20,
13 385–395.
- 14 Grunow, D. (1995). The research design in organization studies: Problems and prospects. *Organization*
15 *Science*, 6, 93–103.
- 16 Hitt, M. A., Gimeno, J., & Hoskisson, R. E. (1998). Current and future research methods in strategic
17 management. *Organizational Research Methods*, 1, 6–44.
- 18 Jones, A. P., Johnson, L. A., Butler, M. C., & Main, D. S. (1989). Apples and oranges: An empirical
19 comparison of commonly used indices of inter-rater agreement. *Academy of Management*
20 *Journal*, 26, 507–519.
- 21 Ketchen, D. J., Jr., & Shook, C. L. (1996). The application of cluster analysis in strategic management
22 research: An analysis and critique. *Strategic Management Journal*, 17, 441–458.
- 23 Kerlinger, F. N. (1986). *Foundations of behavioral research* (3rd ed.). Fort Worth, TX: Harcourt,
24 Brace & Jovanovich College Publishers.
- 25 Kochar, R., & David, P. (1996). Institutional investors and firm innovation: A test of competing
26 hypotheses. *Strategic Management Journal*, 17, 73–84.
- 27 McGrath, R., MacMillan, I., & Venkatraman, S. (1996). Defining and developing competence: A
28 strategic process paradigm. *Strategic Management Journal*, 16, 251–275.
- 29 McWilliams, A., & Siegel, D. (1997). Event studies in management research: Theoretical and
30 empirical issues. *Academy of Management Journal*, 40, 626–657.
- 31 Mohr, J., & Spekman, R. (1994). Characteristics of partnership success: Partnership attributes,
32 communication behavior, and conflict resolution techniques. *Strategic Management Journal*,
33 15, 135–152.
- 34 Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design and analysis*. Hillsdale, NJ: Lawrence
35 Erlbaum.
- 36 Rogosa, M. (1988). Myths about longitudinal research. In: K. W. Schaie, R. T. Campbell, W. Meredith
37 & S. C. Rawlings (Eds), *Methodological Issues in Aging Research* (pp. 171–209). New York:
38 Springer.
- 39 Rouse, M. J., & Daellenbach, U. S. (1999). Rethinking research methods for the resource-based
40 perspective: Isolating sources of sustainable competitive advantage. *Strategic Management*
Journal, 20, 487–494.
- Spector, P. E. (1981). *Research designs*. Newbury Park, CA: Sage.
- Stone-Romero, E., Weaver, A., & Glenor, J. (1995). Trends in research design and data analytic
strategies in organizational research. *Journal of Management*, 21, 141–157.
- Sundaramurthy, C., Mahoney, J., & Mahoney, J. (1997). Board structure, anti-takeover provisions and
stakeholder wealth. *Strategic Management Journal*, 18, 231–245.
- Vogt, W. P. (1993). *Dictionary of statistics and methodology*. Newbury Park, CA: Sage.

APPENDIX A EXAMPLES OF CONTENT ANALYSES

Content analyses of four studies.

(1) Analysis of [Sundaramurthy, Mahoney and Mahoney \(1997\)](#). Board structure, anti-takeover provisions and stakeholder wealth.

Overview: Tests whether stock market responses to adoption of anti-takeover provisions vary relative to directory board structure.

Sample: 261 firms from the S&P 500 that adopted anti-takeover provisions during the 1984–1988 period.

Sources: Data on anti-takeover provision data came from the Investor Responsibility Research Center (IRCC). The daily security market rates of return were obtained from the Center for Research in Security Prices (CRSP).

Dependent variable: Cumulative average normal returns (CARs) earned by shareholders accompanying the adoption of anti-takeover provisions. Event date is the day the proposal statement of anti-takeover provision is mailed to shareholders. The event window was –50 days before the mailing date to +5 days afterwards.

Independent variable: Independent monitoring of the directory board. Operationalized in three ways: (1) proportion of outsiders on board; (2) proportion of outsiders who were not appointed during the CEOs tenure; and (3) whether the CEO and Chairperson was the same person.

Design notation: $O_{-50} \dots O_0 \dots O_{+5}$

X

Design structure is the measurement the dependent variable (O) – stock market price at and between date –50 and date +5 days relative to the event date – and the measurement of the independent variables (X) at the event date.

History: Controls account for confounds due to takeover rumors and adoption year effects.

Maturation: Controlled for with variable measuring size of sampled firms.

Testing: Controlled for with sampling screens.

Instrumentation: Data sources, metrics or coders did not change; threat is controlled for within the design.

Regression: This threat is controlled for in a design with three or more observations.

- 1 Selection: Non-controlled threat to internal validity. The companies were selected
2 on the basis of an event. Any factors associated with companies receiving threats
3 might be potential confounds (i.e. debt, inept management, etc.).
- 4 Mortality: Controlled, as no subjects dropped out over the study period.
- 5
- 6 Selection-maturation: Non-controlled threat to internal validity, as subjects were
7 chosen on the basis of size (S&P 500).
- 8
- 9 Selection-mortality: Not a threat. Selection of subjects was not based on retention
10 or loss over time period.
- 11 Selection-history: Non-controlled threat to internal validity. Since selection of
12 subjects was on the basis of an event of interest (anti-takeover provision), this
13 threat exists.
- 14 Selection-instrumentation: Non-controlled threat to internal validity. Since the
15 selection of subjects was based on whether they were listed on particular data
16 sources (IRCC, CRSP), this threat exists.
- 17
- 18 Selection-testing: This threat does not exist, as the selection of subjects was not
19 based on test results.
- 20
- 21 Ambiguity about Causal Inference: Variables were ordered temporally, controlling
22 for this threat.
- 23 Common methods variance: Data for independent and dependent variables are
24 from different sources, indicating that this threat is controlled.
- 25
- 26 (2) Analysis of [Kochar and David \(1996\)](#). Institutional investors and firm innova-
27 tion: A test of competing hypotheses.
- 28 Overview: This study tests three competing hypotheses on the effects of
29 institutional investors on firm innovation.
- 30
- 31 Sample: 135 public firms traded on the New York Stock Exchange, American
32 Exchange or NASDAQ. These firms were manufacturers.
- 33 Data sources: Compact disclosure database, Money Market Directory Type,
34 Moody's Bank and Finance Manual, Nelson's Directory of Investment Managers,
35 Diaglo's PTS New Product Announcements database. Data for controls were also
36 gathered from COMPUSTAT tapes.
- 37
- 38 Dependent variable: Firm innovation, defined as the number of new products
39 developed by the focal firm over the period 1990–1992 period.
- 40 Independent variables: Proportions of ownership held by three categories of
institutional owners. Data for these variables were collected in 1989.

1 Design notation: The data for the independent variables were gathered one year
2 prior to the dependent variables, which were then gathered annually for three years
3 afterwards:

4 XOOO

5
6 History: Controlled for by use of control variables that accounted for change
7 occurring between observation periods.

8 Maturation: Controlled for, as sample subjects were not selected on the basis of
9 life cycle, size or age.

10
11 Testing: Controlled, as none of the subjects were obtained due to performance or
12 data source.

13 Instrumentation: Non-controlled threat to internal validity. No consideration to
14 possible changes in data sources is made.

15
16 Regression: Controlled for the dependent variable because the data were measured
17 on more than two occasions. Not controlled for with respect to the independent
18 variable.

19 Selection: Non-controlled threat to internal validity. Only manufacturers were
20 included. Discrepancies and tendencies in the propensity of manufacturers to
21 innovate may reflect selection bias.

22
23 Mortality: Controlled, as none of the firms were discarded over the study period.

24 Selection-history: Controlled, as subjects were not selected on the basis of an event
25 of interest.

26
27 Selection-maturation: Controlled, as the subjects were not selected on the basis of
28 life cycle stage or age.

29 Selection-mortality: Controlled, as the subjects were not selected on the basis of
30 retention or loss.

31 Selection-testing: Controlled, as the subjects were not selected on the basis of test
32 results.

33
34 Ambiguity about causal inference: Controlled for, as variables were temporally
35 ordered and dependent variable was measured on multiple occasions arrayed by
36 year.

37 Common methods variance: Controlled for, as data for independent and dependent
38 variables were drawn from different sources.

39
40 (3) Analysis of McGrath, MacMillan and Venkatraman (1995). Defining and
developing competence: A strategic process paradigm.

- 1 Overview: Study identifies antecedents to competence.
- 2
- 3 Sample: Convenience sample of 160 projects in 40 companies. Companies worked
- 4 in manufacturing and service industries.
- 5 Data source: From a survey of multiple respondents working on each project.
- 6 Senior executives selected projects for the study.
- 7
- 8 Dependent variable: Competence was measured using a multiple item scale.
- 9
- 10 Independent variables: Two variables, comprehension and competence, were
- 11 measured using multiple item scales.
- 12 Design notation: Data for the variables were collected simultaneously, producing
- 13 the traditional cross-sectional design. The notation is:
- 14
- 15
$$X/O$$
- 16 History: Controlled, as variables accounted for location and industry.
- 17
- 18 Maturation: Controlled, as variables measured firm size and duration of project.
- 19
- 20 Selection: Non-controlled threat to validity. Sample was based on a convenience
- 21 sample, so characteristics associated with those firms might pose alternative
- 22 explanations.
- 23
- 24 Testing: Non-controlled threat to validity. Executives selected projects for the
- 25 study, raising probability that nonrandom selection existed and that their best
- 26 performing projects were subsequently used. This was not controlled in the study.
- 27
- 28 Selection-history: Controlled, as sample was not based on historical event.
- 29
- 30 Selection-maturation: Controlled, as sample was not based on maturation of
- 31 subjects.
- 32
- 33 Selection-testing: Non-controlled threat to validity, as the executives hand-picked
- 34 (non-randomly) the projects for the study.
- 35
- 36 Ambiguity about causal inference: Non-controlled threat to validity, as the study
- 37 data were collected at the same point in time.
- 38
- 39 Common methods variance: Non-controlled threat to validity, since the data for
- 40 the dependent and independent variables both come from the same source, in this
- case, a survey.
- (4) Analysis of [Mohr and Spekman \(1994\)](#), Characteristics of partnership suc-
cess: Partnership attributes, communication behavior, and conflict resolution
techniques.

1 Overview: Study examines vertical relationships between manufacturers and
2 dealers in the personal computer industry.

3 Sample: 557 computer dealers were identified randomly. Each received a survey.
4 124 surveys were returned (1 for each dealer). Sample screens reduced final sample
5 to 102 usable surveys (representing 102 dealers).
6

7 Data sources: Survey. Dealers were identified using the Association of Better
8 Computer Dealers. One informant from each organization was used.

9 Dependent variables: Partnership success was the dependent variable and was
10 measured by sales volume and satisfaction with partnership between manufacturer
11 and dealer.
12

13 Independent variables: Attributes of partnership (commitment, coordination,
14 trust and interdependence), aspects of communication behavior (communication,
15 quality, information sharing, and participation), conflict resolution (joint problem
16 solving, persuasion, severe resolution, smoothing and arbitration) and closeness
17 in relationship.

18 Design notation: X/O, as this study is a traditional cross-sectional design whereby
19 the data for the independent and dependent variables were measured at the same
20 period in time.
21

22 History: Controlled, as variables measured the length and duration of partnership.

23 Maturation: Controlled, as variables measured the size of the dealership.

24 Selection: Controlled by random selection process and comparisons of early to
25 late respondents.
26

27 Selection-history: Controlled by sample screens (dealers owned by manufacturer
28 were discarded from study).
29

30 Selection-maturation: Controlled by sample screens (dealers who did not interact
31 with manufacturer).
32

33 Selection-testing: Controlled by sample screens (incomplete surveys were
34 discarded).
35

36 Ambiguity about causal inference: Non-controlled threat to validity, as the study
37 data were collected at the same point in time.

38 Common methods variance: Non-controlled threat to validity, since the data for
39 the dependent and independent variables both come from the same source, in this
40 case, a survey.

Uncited references

Hitt et al. (1998), Jones et al. (1989), McGrath et al. (1996) and Rouse and Daellenbach (1999).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40