Strategic Management Journal
Strat. Mgmt. J., **25**: 1063–1074 (2004)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/smj.406

ARE MANAGERS FROM MARS AND ACADEMICIANS FROM VENUS? TOWARD AN UNDERSTANDING OF THE RELATIONSHIP BETWEEN ACADEMIC QUALITY AND PRACTICAL RELEVANCE

DAVID C. BALDRIDGE,[1]* STEVEN W. FLOYD[2] and LÍVIA MARKÓCZY[3]
[1] *College of Business, Oregon State University, Corvallis, Oregon, U.S.A.*
[2] *School of Business, University of Connecticut, Storrs, Connecticut, U.S.A.*
[3] *The A. Gary Anderson Graduate School of Management, University of California, Riverside, California, U.S.A.*

In this paper, we propose a positive relationship between the academic quality and practical relevance of management research. The basis for this is the idea that academicians and practitioners both value research that is interesting and justified—meaning research that challenges and extends existing beliefs and research that offers compelling evidence for its conclusions. We acknowledge that there are likely to be many cases where academicians and practitioners disagree on what is interesting and justified. We argue, however, that there are also likely to be cases where the judgments of the two groups converge. Results from a stratified, random sample of 120 publications are consistent with this argument—showing a positive correlation between an objective measure of an article's academic quality and expert panel ratings of its practical relevance. The analysis also shows positive associations between panel members' global assessment of relevance and ratings of an article's interestingness and justification. These results lend support to the hypothesized overlap, but leave room for considerable difference in the way practitioners and academicians evaluate management research. Copyright © 2004 John Wiley & Sons, Ltd.

Few readers will be unfamiliar with the controversy regarding the practical relevance of academic research. The topic has been the subject of a number of articles, books, and special issues in academic journals (i.e., *Administrative Science Quarterly* 1982, Vol. 27, No. 4; *Academy of Management Journal,* 2001, Vol. 44, No. 2; Lawler *et al*., 1985; Murphy and Saal, 1990; Larwood and Gattiker, 1999) and the focus of three recent presidential addresses at the Annual Meetings of the Academy of Management (Mowday, 1997; Hitt, 1998; Huff, 2000a). While this debate has many elements, one of the core questions is whether academic quality and practical relevance are mutually exclusive (e.g., Tranfield and Starkey, 1998) or mutually reinforcing (e.g., Schendel, 1991; Mowday, 1997). Tranfield and Starkey (1998: 352), for example, argue that 'pure' academic research has its own 'style,' 'methods,' and 'academic agenda' that 'undermine the very notion of management research'. Shrivastava and Mitroff (1984) argue that academicians and

Keywords: academic quality; practical relevance
*Correspondence to: David C. Baldridge, College of Business, Oregon State University, Corvallis, OR 97331-2603, U.S.A. E-mail: david_baldridge@hotmail.com

Received 24 May 2002
Final revision received 22 December 2003

practitioners hold irreconcilably different views about what constitutes academic quality and relevant research. Schendel (1991), on the other hand, provides an example of one who views the relationship between relevance and academic quality as mutually reinforcing—arguing that if academic theories are not sound or well tested they have little to offer practice (Schendel, 1991: 2).

In the current paper, we offer support for this latter view. We begin with a review of previous research, focusing on how differences in research design may explain inconsistent results. We then argue that academicians and practitioners both value research that is *interesting*—in the sense that it questions basic assumptions and beliefs (Daft, Griffin, and Yates, 1987; Davis, 1971)—and *justified*—in the sense that its claims are well supported by evidence (Deutsch, 1997). While we do not claim that these common criteria lead to agreement on the value of all, or even most research, we do maintain that they provide a basis for expecting a positive association between judgments of academic quality and practical relevance.

THEORETICAL BACKGROUND

Empirical studies investigating the relationship between academic quality and practical relevance have produced conflicting findings, ranging from a positive relationship to a negative relationship. We believe variations in research design are a source of systematic differences in these findings. Past studies differ in several important ways. First, in some studies subjects were asked to read actual research articles (or abstracts of research articles) and then rate the relevance of the article. In other studies, researchers simply asked subjects about the criteria they generally use to assess the relevance of research. Past studies also differ as to whether rating criteria were elicited first and then used to rate research relevance, or whether raters were first asked to rate the relevance of an article and only then asked to reveal their criteria. Finally, some studies used practitioners to rate relevance while others relied solely on academicians.

Inconsistent methods and conflicting results

As shown in Table 1, studies that had practitioners first read and evaluate the relevance of actual research articles (or abstracts of articles) and then reveal their evaluation criteria found a positive relationship. In contrast, studies that began by identifying the evaluation criteria and stopped here, or applied predefined criteria to evaluate the relevance of research reports, found a negative relationship. For example, since Shrivastava's (1987) study used this latter method, the negative relationship found may be partly a result of his study's design.

When studies elicit relevance criteria first and then use these to rate the value of research, they risk biasing the resulting evaluations with a 'folk theory' of what is relevant. Behavioral decision-making research (Goldberg, 1968; Kelly and Fiske, 1951) shows that people generally are not aware of the procedures that drive their decisions. Nevertheless, they hold 'folk theories' about the basis for their judgments (Goldberg, 1968). Similarly, studies on individual and organizational learning have demonstrated that there are important differences between the meaning created when people express their views (an espoused theory) and the rules and preconceptions that actually guide people's action (theory in use). 'Moreover, most [individuals are] unaware of the gap between their espoused theory and their actions' (Argyris, 1992: 25).

For this reason, when soliciting assessments from academicians and practitioners, it is important to ask for judgment first and then seek supporting criteria. Otherwise the criteria espoused for academic quality or practical relevance are not likely to reflect the actual criteria used in making such judgments. Needless to say, any conclusions based purely on espoused criteria are not likely to be valid. In the current study we seek to avoid this pitfall.

Hypotheses

Our working hypothesis is that both academic and practitioner communities value research that is (1) interesting and (2) justified by evidence and that these common criteria provide the basis for an empirical association between judgments of academic quality and practical relevance. Further, we would argue that neither one of these criteria in either community is sufficient but that favorable judgments on both are necessary to establish academic quality and practical relevance. As shown by the hatched area in Figure 1, we expect that a small but significant portion of management research will be viewed as interesting and justified

Table 1. Summary of past empirical research

Study characteristics	Weiss and Bucuvalas (1977)	Dunn (1980)	Duncan (1974)[a]	Shrivastava (1987)
# of reports used	50 abstracts	100 articles	None	23 abstracts
Academic quality rated by	155 practitioners (senior federal, state and local officials in mental health agencies)	Group of Researchers[b]	240 researchers and 159 practitioners (managers and consultants)	2 research assistants
Practical relevance rated by	Same practitioners	Same group of researchers[b]	Same researchers and same practitioners	Same research assistants
Timing for eliciting rating criteria	Elicited after rating articles	Elicited after rating articles	Elicited without rating articles	Predefined by researchers before rating articles
Rating criteria for relevance	• Action orientation • Conformity with users' expectations • Challenge of status quo	• Likelihood of knowledge use	• Practicality and usefulness • Profitability • Applicability to special situations • Ease of implementation	• Goal relevance • Operational validity • Innovativeness • Cost of implementation
Rating criteria for academic quality	• Technical quality • Statistical sophistication • Objectivity • Quantitative data • Internal consistency • Data supports recommendations • Comprehensive set of explanatory variables • Generalizibility • Validity • Addition to descriptive or casual knowledge	• Reliability of information • Validity of information	• Logical preciseness • Empirical validity	• Conceptual adequacy • Methodological rigor • Accumulated empirical evidence
Findings	Positive relationship	Positive relationship	Negative relationship	Negative relationship

[a] In this study 240 researchers and 159 practitioners ranked concepts describing academic quality and practical relevance based on the perceived importance of these items to the knowledge flow process between the academic community and practitioners.
[b] The article does not say how many coders were included in the group.

by both academicians and practitioners. The darkened areas of the figure also acknowledge, however, that some research may by judged high quality by academicians but not relevant by practitioners and vice versa. Thus, we do not suggest that the magnitude of the relationship between academic quality and practical relevance is necessarily large.

Consistent with Davis (1971), we define research as interesting based on the extent to which it challenges assumptions or extends knowledge that is based on these assumptions, whether such assumptions are (1) expressed as part of a formal theory or (2) represented in the way people behave on a daily basis. A new theory is thus interesting from an academic perspective when it contradicts the premises of a more established theory or extends existing theories. From a practical perspective, a theory is interesting 'when it denies the significance of some part of ... present on-going practical activity and insists

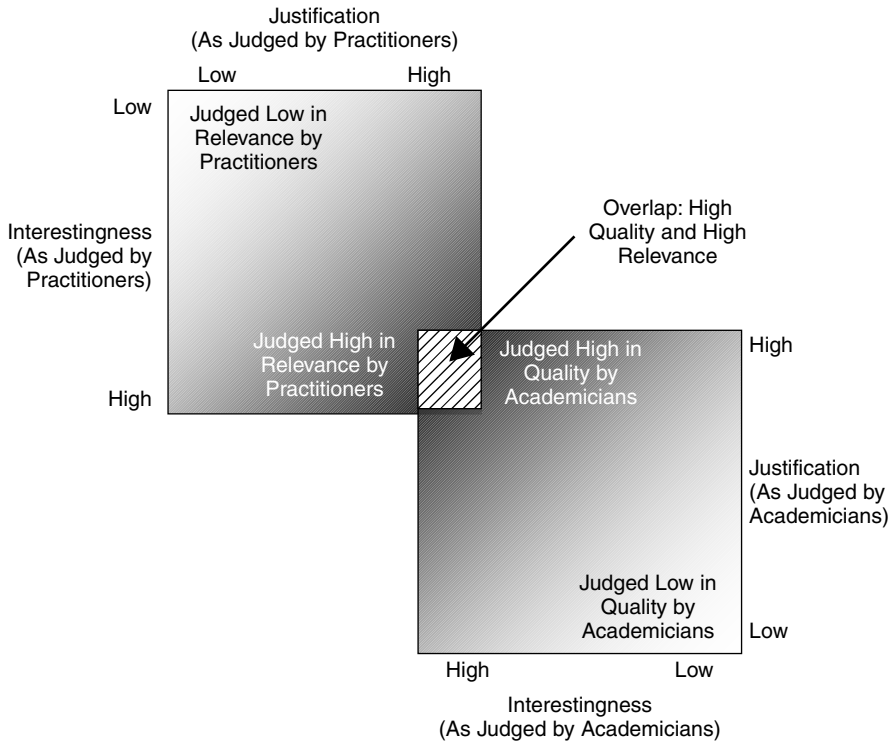

Figure 1. Conceptual relationship between academic quality and practical relevance

that [people] should be engaged in some new on-going practical activity instead' (Davis, 1971: 311). These two perspectives of what makes a theory interesting are likely to align with one another to the extent that formal academic theories and practical experience are grounded in common assumptions or beliefs.

There are reasons for thinking such common ground exists. First, academic research is focused explicitly on improving the effectiveness of organizations, an interest obviously shared by practitioners. According to the mission of the U.S. Academy of Management, for example: 'Our professional goal [is] to ... improve the effectiveness of organizations through our teaching, research, and practice of management' (Academy of Management Code of Ethical Conduct, 2000: 1296). Thus, many academicians have thought about the needs of managers and the practical problems of management, and this should increase the common ground between theory and practice. In addition, there are several opportunities for interactions between practitioners and academicians—in classrooms, consulting, media discussions, etc. To the extent that an awareness of practitioner needs and concerns influences how research questions are framed and/or how studies are designed, it is likely that academic research will resonate, at least to some degree, with practitioners' perceptions of relevant issues. As evidence that academicians seek to link their work to practice, Daft *et al*. (1987: 783) found that scholars consider research interesting when it 'reaches out into the uncertain world of organizations and returns with something clear, tangible and well understood. Significant research takes an issue that is not clear, is in dispute, or is out of focus and brings it into resolution.'

John Reed, formerly CEO of Citicorp, offers further testimony:

> academics stand farther back and draw from the disciplines to get a better sense of what is going on. My own experience over the years has been that [this type of research] tends to be more useful than the [type that depicts best practices] because it creates a framework that allows practitioners to understand how to locate the specific business problems that we may be dealing with within a broader space. (Huff, 2000b: 58)

In sum, although practitioners and academic researchers may often disagree about what is considered interesting, there is also a potential for overlap in their evaluations. We believe that the significance of this overlap has been downplayed in past research and underestimated by practitioners (Priem and Rosenstein, 2000).

Interesting research will not be seen as high quality by academicians or relevant to practitioners, however, unless it is also supported by evidence. Studies show that the more a theory challenges taken-for-granted assumptions (i.e., the more interesting it is), the more people are likely to care about the evidence or justification supporting it (Kunda, 1990). Moreover, academicians and practitioners are not likely to differ in their need for justification. Indeed, practitioners may be especially circumspect, since changing their practices may affect organizational success and their personal reputation (Weiss and Bucuvalas, 1980).

While both scholars and practitioners demand justification when they are presented with interesting findings, the basis for these assessments may differ. Scholars tend to consider formal theory, quality of argument, conceptual adequacy, measurement issues, and strength of inference (Staw, 1995). Practitioners, on the other hand, often lack training in the scientific method and are more likely to apply rougher assessments of justification. For example, they may consider the credibility of the source (e.g., journal quality, author reputation) or plausibility of the arguments. Despite their lack of sophisticated concepts, non-scientists are still likely to be careful in assessing the quality of research evidence when their assumptions are challenged (Kunda, 1990).

The above arguments lead to the following hypotheses:

Hypothesis 1: There is a positive relationship between judgments of academic quality and practical relevance.

Hypothesis 2: There is a positive relationship between practitioners' judgments of practical relevance and their judgments of interestingness and justification.

Hypothesis 3: There is a positive relationship between academicians' judgments of academic quality and practitioners' judgments of interestingness and justification.

METHODOLOGY

Data were collected for 120 articles published in six leading management journals over a 2-year period. We tracked the number of citations and prominence of the citing journal for a 72-month time period, used this as a proxy for academic quality, and compared to an expert panel's ratings of practical relevance. We begin by discussing the sample of articles used and then explain our measures of academic quality and practical relevance.

Research article sample

To create our sample, we drew from the population of articles published in the most frequently cited general management journals in 1994 and 1995. We chose 1994 and 1995 to balance a trade-off between the currency of ideas and the window of opportunity for an article to be recognized by the academic community. (Our citation count measure of academic quality used a 72-month window from the date of publication.) According to the Social Sciences Citation Index, the top six general management journals in these years were the *Administrative Science Quarterly, Academy of Management Journal, Academy of Management Review, Strategic Management Journal, Journal of Management*, and *Journal of Management Studies*. We excluded highly ranked specialized journals, however, such as *Psychological Bulletin*, because articles in such journals do not address general management problems. While the *Strategic Management Journal* is also specialized, it was included because the strategy field focuses on problems faced by general managers (*Academy of Management Newsletter*, 2002).

Across the six journals there were a total of 428 articles published. In order to create a sample of sufficient size[1] that would be representative of the mix of strategy and non-strategy articles in the population, the articles were categorized independently by two of the authors into strategy and all others. We used Rumelt, Schendel and Teece's definition of strategy:

> 'strategy' is about the direction of organizations, and most often, business firms. It includes those

[1] Using the conventions identified in Cohen (1988), a random sample of 120 observations is needed to detect with 80 percent confidence, a significant (alpha = 0.05) correlation of 0.25.

> subjects which are of primary concerns to senior management, or to anyone seeking reasons for the success and failure among organizations. . . . Firms have choices to make if they are to survive. Those which are *strategic* include: the selection of goals, the choice of products and services to offer: the design and configuration of policies determining how the firm positions itself to compete in product markets (e.g., competitive strategy); the choice of the appropriate level of scope and diversity; and the design of organization structure, administrative systems and policies used to define and coordinate work. (Rumelt, Schendel, and Teece, 1991: 6)

The main factors considered in the classification process were the research question and the set of variables in the study. This proved to be a relatively simple way for distinguishing strategy articles from the rest. In the initial classifications we disagreed in only two cases, and these were easily resolved with discussion. The random sampling process was then stratified in order to maintain equal representation across the six journals and to reflect the proportion of strategy articles in each journal. (For example, 26 (34%) of the 75 articles published in *AMJ* were in strategy, so that we drew seven strategy studies out of the total of 20 randomly selected from *AMJ*.) The final sample consisted of 57 articles in strategic management and another 63 on other management topics.

Relevance measurement

Expert panel

Consistent with past research on the reputation of academic journals (e.g., see Coe and Weinstock, 1984; Gomez-Mejia and Balkin, 1992), we chose an expert panel design as a way to achieve a direct measure of practical relevance for a large sample of articles. Expert panels are designed to measure constructs whose assessment requires a particular knowledge base. In the present case, we conceptualized relevance as a construct that applies to a particular study to the extent that an appropriately defined community of practitioners judges it to be interesting and justified. Thus, we needed to define a panel of judges who qualified as an appropriate group.[2]

After discussions with colleagues, we decided to ask members of the executive advisory board of *The Academy of Management Executive* (*AME*), a practitioner journal published by the Academy of Management, to serve as an expert panel. Unlike the academicians on the editorial board, the advisory board is comprised of approximately 82 executives, consultants, and human resource professionals. This group appeared to meet the key design criteria: namely, (1) knowledge on the practice of management and (2) sufficient diversity of backgrounds to represent a range of the managerial perspective.

We approached the editors of *AME* and received their support for the study. The editors wrote a cover letter endorsing the research, and this letter was enclosed with a letter from us explaining the study's purposes and asking for participation from the advisory board members. In the letter, the task of evaluating articles was described as requiring between 2 and 3 hours. Of the 82 requests, 41 board members agreed to participate and, of these, 31 returned completed evaluation packets. We were grateful for this level of participation, and given the significant time required we considered it a sign of interest. Based on their job titles and self-described area of expertise, the 31 respondents were classified into three groups: nine senior managers (e.g., Chief Executive Officer, President, Vice President), eight consultants (e.g., Managing Partner, Senior Project Director, Management Consultant), and 14 human resource professionals (Vice President Human Resources Management, Manager of Human Resources, Senior Human Resources Advisor). The industries represented included financial services, automotive, telecommunications, defense, health care, oil and gas, computers, transportation, aerospace, and consulting. The average age of panel members was 48 years. As an aside, a one-way ANOVA (relevance responses across consultants, HR professionals, and line executives) showed no significant differences between groups in how they rated the relevance of articles.

[2] One way to create such a group would be to sample randomly from the broader population. Such a sample would need to be quite large, however. For example, the U.S. Bureau of Labor Statistics (BLS) tracks 26 different categories of managers (http://stats.bls.gov). If the 547,000 people identified in 1999 as top U.S. executives were considered the population of interest (i.e., general managers), a sample of 384 would be required, assuming a 5 percent confidence interval and an alpha of 0.05. Given the nature of the survey (rating academic articles), a conservative estimate of the response rate would be between 10 and 20 percent, requiring that 1920–3840 surveys be mailed. Not only would such an approach be prohibitively expensive and time consuming, it would also leave the other 2,398,000 'general and operations managers' identified by the BLS unrepresented.

Research summaries

Panel members were asked to evaluate a one- to two-page summary of each article. The use of summaries was necessary to reduce the time demand placed on the panel and to allow screening of technical jargon. In our estimate, reading, digesting, and evaluating the necessary number of full academic articles would have required days—an unrealistic expectation given the high stature of the panel. Use of summaries not only reduced the time demand but also made it possible for us to remove the potential negative bias created by practitioners' unfamiliarity with the tone and jargon of academic research. Although they were experts in the application of theory to practice, they were not research professionals. That is, they were more likely to draw ideas from journals such as the *Academy of Management Executive* or *Harvard Business Review* rather than journals like the *Academy of Management Journal.* In short, the hypotheses call for an assessment of ideas. Asking judges to interpret jargon and methodological details reaches beyond their area of expertise. Thus, we wanted to make the 'translation' for them. It should be noted that this practice is widely used in studies that rely on practitioners for ratings (see, for example, Weiss and Bucuvalas, 1980).

The summaries were written independently by two PhD students, compared for accuracy against the original article separately by two of us, and edited to correct any initial errors. Although technical jargon was avoided, the summaries stayed true to the terminology used in the papers, and for empirical papers included a description of the research design. The care taken in writing, editing, and checking the summaries provides considerable assurance of their face validity as representative of article content. Since relevance is a study variable, however, it is also important to provide objective evidence of the validity of its measurement. According to Carmines and Zeller, a strong test of validity is criterion validity, in which one seeks 'correspondence between the test and the criterion' (Carmines and Zeller, 1979: 17), with correspondence usually estimated by the size of their correlation. In this case, the 'test' may be seen as the ratings of the summaries and the criterion as the ratings of full-length articles. To further validate the summaries, therefore, we assigned judges in a pretest the task of rating one full article and then asked a *different* judge to rate a summary of the same article. The correlation between the full article and summary ratings was 0.51 ($p = 0.006$; $n = 17$). As a benchmark, the magnitude of this correlation is comparable to criterion validities reported in other recent work (e.g., McGrath, 2001). The lack of perfect correspondence in the ratings is likely due to the difficulty judges experienced in interpreting the academic jargon contained in the full articles as well as differences in individual opinion. Copies of the summaries are available upon request.

Global measure of relevance

Since our objective was to minimize the 'folk theory' bias in the evaluation of relevance, we used a direct measure of overall relevance—what we describe as the 'global rating.' In addition to this global rating, we wanted a sense of the degree to which this global rating was based on the practitioners' assessment of what was interesting and justified. So, *following* the global rating we asked practitioners to evaluate research studies based on two dimensions: (1) interestingness and (2) justification. The entire process was defined in a set of user-friendly instructions. Each page in a panelist's packet, for example, represented a specific step, and judges were asked not to vary from the procedure. To further reduce the risk of rater error, we carefully organized and numbered the material in advance.

The rating task of each judge was standardized very carefully to achieve a two-step rating procedure. After reading the instructions, judges were asked to read each summary and provide an overall evaluation of its relevance by circling a number along a single Likert-type scale. Scale anchors were (1) 'No contribution to the practice of management,' (4) 'neutral,' and (7) 'Very significant contribution to the practice of management.' When all the summaries had been assigned a global rating, the judges were asked to put the global rating form in the appropriate envelope and seal it. They were specifically instructed not to change their global ratings after they had finished this phase of the task. Then, judges completed separate forms that rated each summary on the criteria of interestingness and justification.

Dimensional measures of relevance (interestingness and justification)

We measured practitioners' judgments of interestingness with an additive scale comprised of two items: (1) 'I found the ideas in this paper novel and interesting' and (2) 'This paper extends or clarifies my knowledge of a management issue.' The first of these captures Davis's (1971) core idea that interesting things are different and new. The second item is important to include in the measure because extending or clarifying knowledge makes it more useful, and more useful knowledge is inherently more interesting to practitioners, even if it is not entirely new. As an estimate of the internal consistency of the scale, the Pearson correlation coefficient between these two items is 0.864 ($p < 0.001$). Justification was measured with a single item: 'The evidence or argumentation offered in support of this paper's conclusions was adequate.' All items were Likert-type and scale anchors were 'Completely Disagree' (1), 'Neutral' (4), and 'Completely Agree' (7). In addition, at the bottom of each form, there was an open-ended question asking judges to comment on 'What was particularly striking (good or bad) about this paper?' Several blank lines were left for a response.

Procedures

The pre-test mailing went to 20 judges. These packets included a cover letter, a detailed instruction sheet, a global rating form and accompanying envelope, a large envelope containing the dimensional rating sheets, a biographical data sheet, a self-addressed return envelope, and a set of randomly assigned summaries or articles. The cover letter explained the rationale for rating a full-article along with a group of summaries and asked judges to comment on the format, instructions, and other material. The judges found the materials adequate and easy to use, so no changes were required in the packet for the final mailing.

In both the pre-test and the final mailing, judges were assigned to articles and summaries on a random basis and, to the extent possible, we assigned summaries to more than one panelist. We had 31 panelists, 120 articles and an average of seven summaries per judge. (The number of summaries per judge was limited by the time estimate to complete the rating process.) Twenty-two summaries were rated by three raters, 81 were rated by two raters, and 120 were rated by one rater. Inter-rater reliability (IRR) for the three-rater subsample was 0.86 and 0.82 for the two-rater subsample. Because of the relatively high inter-rater agreement and because raters were equally well qualified, we used an average for both the global and dimensional relevance ratings where we had more than one rater.

Measurement of academic quality

Finding a measure of academic quality poses challenges. As scientists, however, management scholars are much more inclined to evaluate the quality of knowledge in a systematic way. The peer-reviewed journal provides a primary screen against poor-quality research. Beyond publication, the citation of an article by a peer within an article from a top journal is generally recognized as a measure of an article's impact. For example, studies of *journal* impact (Salancik, 1986; Sharplin and Mabry, 1985) have used citation counts as an objective measure of journal influence. As a measure of impact, then, the number of citations an *article* receives may be considered a measure of academic quality. Indeed, this measure of quality has been shown to be highly consistent with other measures of quality (Baird and Oppenheim, 1994). Specifically, Baird and Oppenheim (1994), who compared the relationship between citation measures and other measures of academic quality, found that citations tend to correlate strongly with all of these measures. In sum, while citation counting is not a perfect measure of academic quality, it is objective, readily observable, and highly correlated with other measures.

In the current study, we measured academic quality as the number of citations an article received in the top 40 management journals (as determined in Johnson and Podsakoff's, 1994, study) over a 72-month period following the article's publication. We adjusted this figure first by deleting (i.e., not counting) self-citations and second by weighting citation counts by the average rating given to a journal in a large study of journal reputations (Glick *et al*., 1997). A visual inspection of the distribution of article citation count revealed a significant degree of skewness and kurtosis, suggesting deviation from normality. Accordingly, as suggested by Tabachnick and Fidell (1996), a log

transformation was performed on the academic quality variable prior to conducting the analysis.[3]

Following guidelines in Neter, Wasserman, and Kutner (1983: 400–407), we used scatter plots and 'box plots' of the data to identify potential outliers. Based on this analysis, five cases were initially considered. An analysis of the impact of each case on the correlations among study variables showed that only three of the five were influential. In order to be conservative, therefore, only these three were eliminated from subsequent analysis.

RESULTS

Table 2 shows the means, standard deviations, and correlations for the variables in the study, including the global (*a priori*) measure of relevance, the two-dimensional (*a posteri*) measures of interestingness and justification (for practitioners), and the adjusted citation count measure of academic quality. Three patterns of correlation are suggested by the prior theoretical discussion. First, if Hypothesis 1 is valid, then there should be a positive and significant correlation between the objective measure of academic quality and the global (*a priori*) measure of relevance. As shown in Table 2, the Pearson coefficient between the measure of academic quality and the global measure of relevance is positive and significant ($r = 0.203$, $p < 0.05$).

Second, if Hypothesis 2 is valid, then the expert panelists' global ratings of relevance should be significantly correlated with the two-dimensional ratings of interestingness and justification. As shown in Table 2, correlations between panelists' *a posteri* assessments of interestingness and justification and their *a priori* judgments of global relevance are positive and significant ($r = 0.764$, $p < 0.001$ and $r = 0.664$, $p < 0.001$, respectively). Although concerns about multicollinearity limit the viability of regression coefficients as evidence of the relative weights of the two dimensions, regressing the two dimensions on the global scale also showed positive and significant effects (adjusted $R^2 = 0.627$, $F = 98.587$, d.f. = 2,114, $p < 0.001$) with $\beta = 0.575$ ($p < 0.001$) for interestingness and $\beta = 0.292$ ($p < 0.001$) for justification.

Third, if Hypothesis 3 is valid, then there should be a correlation between the objective measure of academic quality and panelist's *a posteri* assessment of interestingness and justification. As shown in Table 2, a positive correlation was found between academic quality and both interestingness and justification. The relationship between academic quality and justification, however, was not statistically significant ($r = 0.105$) and the relationship between academic quality and interestingness was marginally significant ($r = 0.165$, $p < 0.100$). Although inconsistent with the hypothesis, in retrospect the weak relationship between academic and practitioner assessments of justification may not be surprising. Unlike the interesting criteria, common ground on this dimension is not cultivated by the professional values of academicians, and interactions between these groups rarely concern questions of justification.

Although the significant correlation between the global measure of relevance and the measure of academic quality provides support for the hypothesis that practitioners' and academicians' judgments overlap, the effect size also leaves significant room for cases where their judgments diverge or where there is no relationship at all. In order to get a better feel for the amount of overlap, we sought to determine how many of the articles in the sample

[3] As an aside, we actually collected six scores for each article: the number of citations in all journals, this number reduced by self-citations, the number of citations in the top 40 journals, this number adjusted by self-citations, a scale where each citation in the top 40 journals was weighted by the journal's rank in the reputation study, and this number adjusted by self-citations. The average correlation among these measures was 0.96 and using one over another does not materially affect the results.

Table 2. Descriptive statistics and correlation matrix for study variables

Variables	Mean	SD	1	2	3	4
1. Global relevance	3.86	1.12	1.0			
2. Interestingness	4.03	1.24	0.764***	1.0		
3. Justification	3.79	1.22	0.664***	0.646***	1.0	
4. Academic quality	42.06	41.52	0.203*	0.165†	0.105	1.0

†$p < 0.10$; *$p < 0.05$; ***$p < 0.001$
$N = 117$

received high scores on measures of both relevance and academic quality. These would represent the articles from the sample occupying the common area in Figure 1.

To do this, we identified the 27 articles whose ratings on the global scale were equal to or exceeded 5.0. Rather than dividing the sample into quartiles, this cut-off was used because a rating of 4.0 was 'neutral' and therefore an article rated 5.0 or above was definitively judged to be relevant. Of these 27 articles, 10 (37%) were also in the set of 27 articles rated highest on academic quality. This proportion compares with 22.5 percent (27/120) or 6.1 articles that could be expected on the basis of chance ($\chi^2 = 13.112$, $p < 0.001$). Consistent with this, a split-sample analysis of variance shows that the high relevance group's mean citation count (standard deviation) of 53.86 (41.38) is significantly higher than the mean of 29.86 (25.94) for the 37 articles in the group judged to be less relevant, i.e., those rated 3.0 or less on the scale ($F = 8.112$; $p < 0.01$). In order to get another perspective on the overlap, however, it is worth noting that three of the articles in the high-relevance set have 72-month weighted citation scores of 0.0.[4]

DISCUSSION AND CONCLUSION

Are managers from Mars and academicians from Venus? Some of what has been written about the relationship between academic quality and practical relevance seems to suggest this is so—that these communities hold very different, if not irreconcilable, worldviews. The true nature of the relationship has profound implications for both groups. If relevance and academic quality are inversely related or unrelated then relaxing academic standards may be a rational path toward increased relevance. If, however, a positive relationship exists, then efforts to increase relevance by loosening academic standards are misguided, and academicians and practitioners alike need to pay attention to the academic quality of management research.

The results in this study show a positive correlation between an expert panel's assessment of practical relevance and an objective measure of academic quality. Moreover, a positive correlation was found between the expert panel's *a posteri* assessments of interestingness and justification and *a priori* judgments of global relevance. While these results lend support for our hypotheses, the magnitude of the overlap between academic quality and practical relevance was fairly small. Further, consistent with our hypotheses a positive correlation was found between academic quality and both interestingness and justification. The relationship between academic quality and justification, however, was not statistically significant and the relationship between academic quality and interestingness was only marginally significant.

Several limitations should also be mentioned. First, while the members of the expert panel were qualified to make the judgments we asked of them, their ratings may not be representative of those that would have been derived from a sample of practitioners. As noted earlier, however, obtaining a sufficiently large, representative sample was considered impractical. Second, we used citation counts as the measure of academic quality. This measure is indirect in the sense that it does not ask academicians for an assessment of an article's interestingness and justification. All articles in the sample had been published in top journals, however, and thus had received high marks for merit by academic reviewers. Further, method (justification) and contribution (interestingness) are part of the standard review criteria in these journals. In this context, citation counts appear to offer an unbiased indicator and a more precise discriminator of quality than seeking what would effectively be a second round of review. And as discussed previously, citation counts are highly consistent with other measures of academic quality. Third, we sampled from articles published only in top journals and, as a result, we significantly restricted the range on the citation count measure of academic quality.

[4] To examine how our conceptualization of relevance fit with panelists' unstructured reflections, we analyzed answers to the open-ended question 'What was particularly striking (good or bad) about this paper?' Responses were open-coded (Strauss and Corbin, 1990) independently by two of the authors, and the coders ultimately agreed on nine concepts capturing 87 percent of the initial data. These can be grouped semantically under the headings of novelty (i.e., interestingness, obviousness, novelty, usefulness, and importance) and justification (i.e., research design, logic of theory, fit to experience, and trustworthiness of conclusions or results) in a way that corresponds to the dimensions in Figure 1. This analysis must be viewed with caution because the comments had been influenced by the criteria used in the dimensional rating scales. The results may be seen as confirmatory but not definitive (Miles and Huberman, 1984).)

Future research is needed to better understand the relationship between academic quality and practical relevance. Perhaps most important, why are some articles that are considered highly relevant received so poorly in the academic community? Why are others considered high quality by academicians perceived to be irrelevant? Do research topics explain these differences? Are there patterns in the methodology that are important? Without answers to these questions it is difficult to make a blanket recommendation to pursue academic quality as a means for increasing practical relevance.

Instead, we offer two more limited implications about *how* to pursue academic quality and practical relevance. First, researchers should study questions that challenge both existing scientific theory and conventional management practice. Management practices, however, are not always based on the latest scientific theory. One way for researchers to cope with this problem is to ascertain early in the process whether current practice with respect to the research question conforms to current theory. When such alignment exists, pointing it out is likely to increase the interest of practitioners. When it is absent, researchers may want to reconsider the theoretical premises of the study or pursue a different question. In either case, the key point is that researchers should develop a thorough understanding of the theories in use by practitioners as an early step in the research process. Second, more research is needed that is considered justified by both academicians and practitioners. The problem here is that practitioners are unlikely to understand or appreciate many of the subtleties of the scientific method, thereby dismissing research on the basis that 'the sample is too small,' for example, without really knowing what size of sample is appropriate. To enhance practitioner assessments of methodology and evidence, researchers should heed Dubin's admonition that 'the applied scientist has to make sense to the practitioners who consume his results' (Dubin, 1976: 19). Put differently, the definitions of variables, research design, and data should be described in a way that fits with how practitioners would describe the situation being addressed in the study. Again, this requires that researchers understand the practitioners' perspective and that methodological choices be guided by the parameters of practitioner experience. We hope that these suggestions will contribute to the production of more high-quality, highly relevant management research.

ACKNOWLEDGEMENTS

The authors wish to thank Jeffrey Goldberg for contributing to the development of the original idea of this paper and Richard Priem for his insightful comments on earlier versions of the manuscript. Dan Schendel and two anonymous reviewers of *SMJ* also provided helpful comments that significantly improved the paper. We are grateful to Karen Blakeley, Darshan Hiranandani, and Zeki Simsek for their help with data collection.

REFERENCES

Academy of Management Code of Ethical Conduct 2000. *Academy of Management Journal* **43**(6): 1296–1300.

Academy of Management Journal. 2001. Special research forum: knowledge transfer between academics and practitioners. **44**(2): 340–440.

Academy of Management Newsletter. 2002. Business Policy and Strategy Division (Domain Statement). Fall: 3.

Administrative Science Quarterly. 1982. Special Issue, Part I: The utilization of organizational research. **27**(4): 587–685.

Argyris C. 1992. *On Organizational Learning*. Blackwell: Cambridge, MA.

Baird LM, Oppenheim C. 1994. Do citations matter? *Journal of Information Science* **20**(1): 2–15.

Carmines EG, Zeller RA. 1979. *Reliability and Validity Assessment*. Sage: Beverly Hills, CA.

Coe R, Weinstock I. 1984. Evaluating the management journals: a second look. *Academy of Management Journal* **27**(3): 660–666.

Cohen J. 1988. *Statistical Power Analysis for the Behavior Sciences*. Erlbaum: Hillsdale, NJ.

Daft RL, Griffin RW, Yates V. 1987. Retrospective accounts of research factors associated with significant and not-so-significant research outcomes. *Academy of Management Journal* **30**(4): 763–785.

Davis MS. 1971. That is interesting. *Philosophy of the Social Sciences* **1**: 309–344.

Deutsch D. 1997. *The Fabric of Reality*. Penguin: London.

Dubin R. 1976. Theory building in applied areas. In *Handbook of Industrial and Organizational Psychology*, Dunnette MD (ed). Rand McNally College: Chicago, IL; 17–39.

Duncan WJ. 1974. Transferring management theory to practice. *Academy of Management Journal* **17**(4): 724–739.

Dunn WN. 1980. The two-communities metaphor and models of knowledge use. *Knowledge Creation, Diffusion, Utilization* **1**(4): 515–536.

Glick W, McKelvey W, Cooper M, Huber G, Zmud W. 1997. Survey feedback from surveys of journal reputations. Report of the 1997 INFORMS Committee Review of *Organization Science*.

Goldberg LR. 1968. Simple models or simple processes? Some research on clinical judgments. *American Psychologist* **23**: 483–496.

Gomez-Mejia LR, Balkin DB. 1992. Determinants of faculty pay: an agency theory perspective. *Academy of Management Journal* **35**(5): 921–955.

Hitt MA. 1998. Presidential address: twenty-first-century organizations: business firms, business schools, and the Academy. *Academy of Management Review* **23**(2): 218–224.

Huff AS. 2000a. Changes in organizational knowledge production. *Academy of Management Review* **25**(2): 288–293.

Huff AS. 2000b. Citigroup's John Reed and Stanford's James March on management research and practice. *Academy of Management Executive* **14**(1): 52–64.

Johnson JL, Podsakoff PM. 1994. Journal influence in the field of management. An analysis using Salancik's index in a dependency network. *Academy of Management Journal* **37**(5): 1392–1407.

Kelly EL, Fiske DW. 1951. *The Prediction of Performance in Clinical Psychology*. University of Michigan Press: Ann Arbor, MI.

Kunda Z. 1990. The case for motivated reasoning. *Psychological Bulletin* **108**(3): 480–498.

Larwood L, Gattiker UE. 1999. *Impact Analysis: How Research Can Enter Application and Make a Difference*. Erlbaum: Mahwah, NJ.

Lawler EE, Mohrman AM, Mohrman SA, Ledford Jr GE, Cummings TG, & Associates. 1985. *Doing Research that is Useful for Theory and Practice*. Jossey-Bass: San Francisco, CA.

McGrath RG. 2001. Exploratory learning, innovative capacity, and managerial oversight. *Academy of Management Journal* **44**(1): 118–131.

Miles MB, Huberman AM. 1984. *Qualitative Data Analysis*. Sage: London.

Mowday RT. 1997. Presidential address: reaffirming our scholarly values. *Academy of Management Review* **22**(2): 335–345.

Murphy KR, Saal FE. 1990. *Psychology in Organization: Integrating Science and Practice*. Erlbaum: Hillsdale, NJ.

Neter J, Wasserman W, Kutner MH. 1983. *Applied Linear Regression Models*. Richard D. Irwin: Homewood, IL.

Priem RL, Rosenstein J. 2000. Is organization theory obvious to practitioners? A test of one established theory. *Organization Science* **11**(5): 509–524.

Rumelt RP, Schendel D, Teece DJ. 1991. Strategic management and economics. *Strategic Management Journal*, Winter Special Issue **12**: 5–29.

Salancik GR. 1986. An index of subgroup influence in dependency networks. *Administrative Science Quarterly* **31**(2): 194–211.

Schendel D. 1991. Editor's comments on the Winter Special Issue on 'Fundamental Research Issues in Strategy and Economics. *Strategic Management Journal*, Winter Special Issue **12**: 1–3.

Sharplin AD, Mabry RH. 1985. The relative importance of journals used in management research: an alternative ranking. *Human Relations* **38**(2): 139–149.

Shrivastava P. 1987. Rigor and practical usefulness of research in strategic management. *Strategic Management Journal* **8**(1): 77–92.

Shrivastava P, Mitroff I. 1984. Enhancing organizational research utilization: the role of decision makers' assumptions. *Academy of Management Review* **9**(1): 18–26.

Staw BM. 1995. Repairs on the road to relevance and rigor: some unexplored issues in publishing organizational research. In *Publishing in Organizational Sciences*, Cummings LL, Frost PJ (eds). Sage: Thousand Oaks, CA; 85–97.

Strauss AL, Corbin J. 1990. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Sage: Newbury Park, CA.

Tabachnick BG, Fidell LS. 1996. *Using Multivariate Statistics*. HarperCollins: New York.

Tranfield D, Starkey K. 1998. The nature, social organization and promotion of management research: towards policy. *British Journal of Management* **9**(4): 341–353.

U.S. Bureau of Labor Statistics. 2003. U.S. Department of Labor Bureau of Labor Statistics. Washington DC. http://stats.bls.gov.

Weiss CH, Bucuvalas M. 1977. The challenge of social research to decision making. In *Using Social Research in Public Policy Making*, Carol HW (ed). Lexington Books: Lexington, MA; 213–234.

Weiss CH, Bucuvalas M. 1980. Truth tests and utility tests: decision makers' frames of reference for social science researchers. *American Sociological Review* **45**(2): 302–312.

