ORIGINAL ARTICLE

The Seven Deadly Sins of Communication Research

W. Russell Neuman[1], Roei Davidson[2], Sung-Hee Joo[1],
Yong Jin Park[1], & Ann E. Williams[3]

1 Department of Communication Studies, University of Michigan, Ann Arbor, MI 48104
2 Department of Communication, University of Haifa, Haifa 31905, Israel
3 Department of Communication, Georgia State University, Atlanta, GA 30303

We analyzed anonymized copies of the complete reviewer comments for 120 recent sub-missions to the Journal of Communication *and attempted to identify the scholarly "sins" and "virtues" most frequently mentioned by the reviewers and most closely associated with the decision to publish the submission. We assessed levels of interreviewer agreement and patterns of evaluation in different subfields of communication scholarship. An explicit connection to a clearly identified theoretical corpus and novel findings or perspectives proved to be the most important predictors of publication. We discuss the ramifications of these findings for the current state of communication research.*

One particularly revealing juncture in the conduct and diffusion of research on human communication is the peer review exercise—the gatekeeper process that is designed to determine the scholarly and scientific merit of inquiry in the field. Peer review is central to the allocation of grants and awards, the tenure process, and, of course, acceptance for publication. Peers turn out to be a rather critical lot. The *Journal of Communication*, for example, accepts only about 16% of articles submitted. High-quality research is not easily achieved, and even communication's most distinguished and admired scholars will attest to their personal experience of significant criticism and rejection.

In the interest of probing how the field of communication defines quality, we selected one journal, the *Journal of Communication*, and with the generous cooperation of its editors and staff and the approval of the International Communication Association's oversight publication committee, we obtained carefully anonymized copies of 120 recent sets of peer review comments. Our goal was to discern which sins and which virtues were most frequently identified and which were most closely associated with the recommendation for publication. We base our article title on the notion of the seven deadly sins rather than the seven cardinal virtues because

---

Corresponding author: W. Russell Neuman; e-mail: rneuman@umich.edu

the sinful variant is much better known and iconic, but, of course, our primary motivation is to understand what leads reviewers to identify research as important and significant rather than compile some sort of list of miscues or failures. Sinfulness turns out to be a dominant theme, however, as negative remarks outnumbered positive ones in our analysis by 9 to 1.

The research team initially drafted a list of the sins and corresponding virtues expected to define high-quality communication research and then to explore whether such a list adequately captured the content of the review sets. The exercise was instructive but perhaps naïve. Our subsequent reading of the surprisingly extensive research literature on the dynamics of peer review spanning a broad range of humanistic, scientific, and social scientific fields revealed that drafting such a list may be comparable to reinventing the wheel. The key criteria of scholarly merit are relatively well developed in the literature and, conveniently, there appear to be about seven clusters of criteria.

The peer review literature

This interdisciplinary literature is robust, diverse, and rather large. One analyst tracked down 643 recent academic papers on the peer review process, 101 of which included empirical evidence (Armstrong, 1997). There are extensive studies of reviewers' evaluative criteria and agreement levels in psychology, sociology, economics, educational research, medical research, and the physical sciences but, unfortunately, very few in communication. Among the few communication-oriented studies, there is a tendency to focus on bibliometric data and patterns of citation rather than evaluation of scholarship (e.g., Beniger, 1990; Bunz, 2005; Lin & Kaid, 2000; Rice, Borgman, & Reeves, 1988). Bryant and Miron (2004) is a notable exception.

From the broader literature of the peer review process, two empirical generalizations stand out: (a) the criteria of evaluation are quite similar across academic disciplines and (b) the journal acceptance rates are not. In the physical sciences, most journals accept about 80% of manuscripts submitted, whereas in the social sciences, about 80% are rejected. This was a key finding of the unfailingly cited and seminal study in this literature published by sociologists Zuckerman and Merton (1971). Initially, researchers concluded that this was primarily a function of the fact that such fields as chemistry and physics have well-developed and agreed-upon paradigms of appropriate topics and methods of analysis that the preparadigmatic and quasi-paradigmatic social and humanistic traditions lack (Braxton & Hargens, 1996). This conclusion resonated with our motivating interest in exploring the possible roots of a common paradigm and defining core of critical questions in communication. But as researchers probed further, they discovered that, surprisingly, although criteria of evaluation were quite consistent across diverse fields, agreement among reviewers on quality was remarkably low regardless of level of paradigm development[1] (Bakanic, McPhail, & Simon, 1987; Cicchetti, 1991; Cole, Rubin, & Cole, 1977; Hargens, 1988; Scott, 1970; Wolff, 1970). The average interreviewer agreement on quality and

publishability or worthiness for funding is a correlation coefficient of about .30 or roughly 10% of the variance in evaluation explained by commonly agreed-upon criteria of quality (Cole & Cole, 1981; Hackett & Chubin, 2003). The classic if somewhat controversial demonstration of reviewer divergence was a study that took 12 recently accepted psychology papers and changed the authors' names and institutional affiliations and resubmitted them to the same journals. Three were recognized as resubmissions and rejected, but of the remaining nine that were sent out for review, eight were rejected as unworthy of publication (Peters & Ceci, 1982). One editor admitted: "all who routinely submit articles for publication realize the Monte Carlo nature of the review process" (Campanario, 1998, p. 191). Indeed, in one study, 69% of rejected papers were found to be published elsewhere, sometimes in more prestigious journals (Chew, 1991).

Such results led to a debate in the literature about whether reviewer agreement is actually a worthy goal. Editors, it is argued, solicit diverse reviewers and welcome evaluations that are complementary rather than replicative (Harnard, 1979). Diversity fosters creativity and serves as a counterpoint to an otherwise conservative process that may stifle innovation (Armstrong, 1997).

The seven sins typology

We proceeded to try to identify the most fundamental and commonly used criteria of evaluation evident in the peer review literature. Terminologies vary, but the underlying criteria were surprisingly consistent across studies. A representative sample of evaluative criteria (including some criteria lists that have been routinized as checklist forms in several journal evaluation procedures) are summarized in Table 1. The studies are identified by principal authors across the top banner of the table, and the seven terms we ultimately used to identify our seven sins and corresponding virtues are listed on the vertical axis with an additional eighth "miscellaneous" category.

Importance, the first criterion, appeared in some form in virtually all evaluative typologies and ranks among the most frequently utilized in evaluations. But it may also be among the least precise. Some used the term to identify the importance of the underlying topic addressed, some the novel contribution of the reported research to a given topic, which we identified (as did others) under a separate heading of originality.[2] Other analysts appeared to equate importance with perceived reader interest. In the review sets analyzed here, this criterion attempts to capture the reviewer's judgment of the general significance of the broadly defined topic undertaken in the article. Such language often appears as almost ritualized commentary used by editors and reviewers when rejecting a paper to acknowledge that yes, the issue is important, but the paper does not say anything significant about it.

Theoretical integration refers to the strength of connection between the empirical or analytic results and the recognized elements of theory. Some typologies simply used the term "theory" to identify this clearly central criterion; others used such phrases as the "linkage of concepts to execution."

Table 1 Principal Criteria of Scholarly Quality in the Peer Review Literature

Reviewer Criteria	Bonjean & Hullum (1978)	Bakanic, McPhail, & Simon (1989)	Cicchetti (1991)	Fiske & Fogg (1990)	Straub, Ang, & Evaristo (1994)	Petty, Fleming, & Fabrigar (1999)	Eisenberg, Thompson, Augir, & Stanley (2002)
Importance	Important contribution	Topic	Importance	General—Importance	Topic selection	Importance/ significance	Significance, importance, reader interest
Theoretical integration	Theory	Theory		Linkage of concepts to execution	Theory	Conceptualization/ theory/hypothesis	Theoretical contribution
Clarity	Presentation	Writing style		Editorial & writing, presentation	Professional style and tone		Quality of writing
Methodology	Methodology	Design, data, sample, measurement, analysis, results	Scientific reliability	Design, procedures, measurement, statistics	Statistical analysis, research design, replicability	Methodology/ procedures/ analyses	Quality, accuracy of analysis
Completeness					Coverage of significant literature	Quality of literature review	Adequacy of literature review
Normative connection							
Originality	Important contribution		Originality		Contribution to knowledge		
Miscellaneous	Editorial discretion		Suitability		Contribution to practice, adherence to scientific ethics, length, reputation, future research		Appropriate length, persuasive point of view, fairness to alternative view

Clarity was our label for the variety of terms used to evaluate the quality of writing or organization of the argument. We did not use this category to code minor grammatical snafus and violations of APA style conventions.

Methodology, of course, was included in everybody's typology. Many analysts chose to list four to six separate dimensions of methodological quality. We lumped them all together. As a result, this criterion included comments on the design and execution of the reported research including sampling, sample size, validity, reliability, inference of causation, and, of course, a long list of statistical issues.

Completeness comes up frequently in reviewer comments when reviewers note that something has been left out. Most frequently, reviewers identify gaps in the cited literatures or in explaining what was done. Other times, reviewers explicitly refer to an alternative explanation which was not considered.

Normative connection is a centrally important category for our analysis but one that is infrequently cited in the peer review literature (note the empty row in Table 1). By normative connection, we refer to the evaluation of the potential relevance of the reported research to social, cultural, or political values and matters of real-world public concern. Understandably, this is an unstated norm in many of the physical sciences and a delicate topic for some of the social sciences but often a more frequently articulated dimension in the humanities side of the house. The studies by Lamont and colleagues on peer evaluation in the humanities note from their empirical studies that "moral judgments play a central role in peer evaluation." (Lamont & Mallard, 2005, p. 11) And they note that "these associations have gone entirely unnoticed by the literature on peer review" (Guetzkow, Lamont, & Mallard, 2004, p. 203; see also Lamont & Molnar, 2002). This criterion resonates with Gerbner's (1983a) original call for research that makes a difference in the original *Ferment in the Field* volume, in Levy and Gurevitch (1993), and the International Communication Association's recent initiative on *Communication in the Public Interest*.

Originality is listed last but certainly not to indicate lesser importance as a criterion. As commonly used in the peer review literature, this category emphasizes the reviewers' identification of original perspectives, methods, and findings (Guetzkow et al., 2004).

The identification of sins and virtues in the *Journal of Communication* review process

Our central research questions focused on the distribution of positive and negative reviewer evaluations on each of these seven dimensions, the structure of those dimensions, and any patterns of correlation between the evaluations and the ultimate recommendation to publish the submitted paper. We were also curious about levels of interreviewer agreement compared with published reports from other fields although we were reluctant to simply characterize any apparent agreement as evidence of some sort of paradigmatic unity. Given the celebrated fragmentation of the field,[3] we wanted to explore whether patterns of evaluation were markedly different

among subspecialties, although our limited sample size constrained our ability to make too fine a set of distinctions. We had hunches on which dimensions would prove to be most important for predicting acceptance but did not have grounds for more formalized hypotheses, given limited prior research.

Method

We assembled copies of all review comments for the 100 most recent submissions that had completed the review process and had been formally accepted or rejected in the fall of 2005. Names of reviewers and reviewees were vigilantly blacked out. Given that we would expect only about 16 acceptances within the primary sample, we requested a supplemental sample of 20 additional accepted submissions by simply going back further in time and including only "accepts" ignoring the others. This would hopefully provide us with sufficient variance to study the linkages between evaluative dimensions and acceptance for publication. Due to some processing difficulties, our resultant usable sample consisted of 117 review sets—98 from the original sample (82 rejects, 16 accepts) and 19 from the supplemental sample (19 accepts.) We judged 86 of these studies to be quantitative in character and 27 to be qualitative (4 not categorized). Using the journal's category system, within this sample, there were 22 experimental studies, 19 content analyses, 38 surveys, 23 qualitative research, 2 critical, 3 theoretical, and 10 other. This appeared to be a pretty typical collection of submissions matching to the journals' internal records with perhaps a few more surveys and a few less qualitative and critical studies than usual.

We did not request and did not review the actual submitted articles, just the review sets. Typically, two or more reviewers would read and provide comments on the submission in addition to the editor. Among acceptances, most were required to revise and resubmit. This sample of 117 review sets included 381 unique reviews, on average 3.3 reviews per submission. The editor's comments were always identified as such, and in context, it was easy to match up a reviewer's first and second review on a revise-and-resubmit. But we do not know how many unique reviewers were engaged in this process because, of course, they were anonymous. Our sample in aggregate consisted of 213,437 words, an average of 1,824 words per review set. The shortest review set was 368 words and the longest 6,265 words.

Most editors return at least a few submissions as simply inappropriate for the journal and these submissions are not sent out for review. All submissions in this sample were in fact sent out for review, and in such cases, of course, the editor relies heavily on the judgments of the reviewers, weighing in with an independent judgment usually only when outside reviewers disagree. We made no distinction between the first and (as appropriate on revise and resubmit) the second or third rounds of reviews. The unit of analysis was the set of accumulated review commentaries, in most cases two review letters and a brief cover letter from the editor that only occasionally added substantive and evaluative editor's comments. When editor comments were substantive, we coded them as well.

Our content analysis resulted in 3,905 codable comments. In aggregate, the reviewers identified 385 virtues and 3,520 sins. On average this comes to 3.3 virtues and 30 sins per review set. As noted, reviewers are a rather critical lot. But the nature of the criticism is most often constructive.

Coding the type of language used in the review process turns out to be extremely difficult. The paper authors, reviewers, and editor are all familiar with the submitted paper, unseen by our coders, and it is difficult sometimes to fully understand the various allusions, abbreviations, and jargon bandied about by what is typically a highly sophisticated set of specialists. Reviewers, after all, are hard-working volunteers, so editors find it in their interest to routinely tolerate a tremendous diversity of informalities, in-jokes, irrelevances, and asides that seem to accumulate in the process. Our seven dimensions are, in the language of the content analysis trade, highly abstract inferences from latent meaning. The coders initially coded the same randomly chosen practice review sets independently and then in four rounds of successive training meetings compared codes paragraph by paragraph discussing coding rules and adding details and clarifications to the codebook. At the end of the coding process, we randomly assigned a dozen coded review sets to be redundantly coded by a second coder for the purposes of calculating intercoder reliability. Across counts for our 14 sins and virtues, the average intercoder correlation was an $r = .72$. The total number of coded comments correlated $r = .84$ across coder pairs.[4]

Lombard, Snyder-Duch, and Bracken (2002), however, warn against using correlations to assess intercoder agreement as perhaps too liberal a measure, so we computed Lin's (1989) Concordance Correlation Coefficient (CCC) index as well.[5] Relatively high reliability was found for theoretical integration, clarity, and methodology (.83, .84, and .83, respectively)—none of these reliability measures included 0 in their two-tailed 95% confidence interval. Fair reliability was found for importance and completeness (.43 and .48, respectively). Reliability for normative connection and originality was too difficult to assess because of the very low frequency with which these domains were evident in the review sets. Given the projective and novel nature of the variables coded in this project, the reliability levels were less than ideal but adequate to the task. The difficulty of inferring these dimensions from the wide-ranging commentaries (in this, intercoder agreement is not just a methodological check but in itself a finding of interest) raises the prospect that various explicit evaluative checklists might fruitfully formalize and clarify the journal submission evaluation process.

Results

We found in our sample, a relatively high level of interreviewer agreement on the ultimate recommendation to accept or reject. Setting the supplemental special sample of acceptances aside, we examined the original sample of the last 98 available completed submissions to the journal and found agreement (almost always two independent reviewers with the concurrence of the editor) 69.4% of the time. Of the

98 original cases, 57 represented agreement to reject, 11 agreement to accept, and 30 differences of opinion, 5 of which were ultimately accepted (via the revise-and-resubmit-process). Does such a pattern provide some evidence of a concurrence among reviewers on what represents high-quality scholarship in the communication field? It is probably not possible to say from such a crude statistic. It would appear to reflect a higher level of agreement than one might expect, given the widely cited interreviewer correlation on quality ratings among scientists of $r = .3$ (Cole & Cole, 1981; Hackett & Chubin, 2003). *Journal of Communication* reviewers are asked to make a dichotomous decision on publication worthiness, and with a well-known acceptance rate of 16%, one might expect to find a large number of agreements to reject based on pure chance if not evaluative convergence. Another complication is the question of when the recommendation to publish is made, and it is made for the same reasons. Another complexity is the subtle process by which the editor selects expert specialists from what are sometimes arcane subfields to do the evaluating. We will take a closer look at substance of the reviewer comments below.

Which are the most prevalent deadly sins and cardinal virtues in the communication field? The results according to our *Journal of Communication* sample are outlined in Figures 1 and 2. On the virtue side (Figure 1), we find that importance dominates, towering over the other dimensions in our histogram. The reported data here reflect the raw numbers of coded comments across our entire sample of 117 review sets. Recalling that most of these papers were rejected, after all, this may be evidence of a somewhat ritualized courtesy reviewers and editors follow which takes the form of noting that the paper "has indeed selected an important topic, but..." as the reviewers then turn to a list of the paper's shortcomings. We see that positive observations on theoretical integration, clarity, and methodology are frequently noted and completeness, normative connection, and originality much less so. Given the significance, particularly of originality, in textbooks, the journal's guidelines, and the lore of the academy, we were surprised and concerned by its apparent rarity as a noted virtue in the review process.
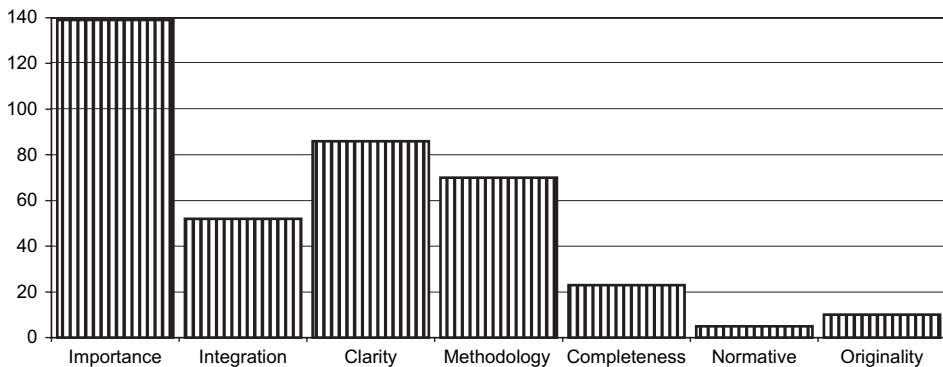

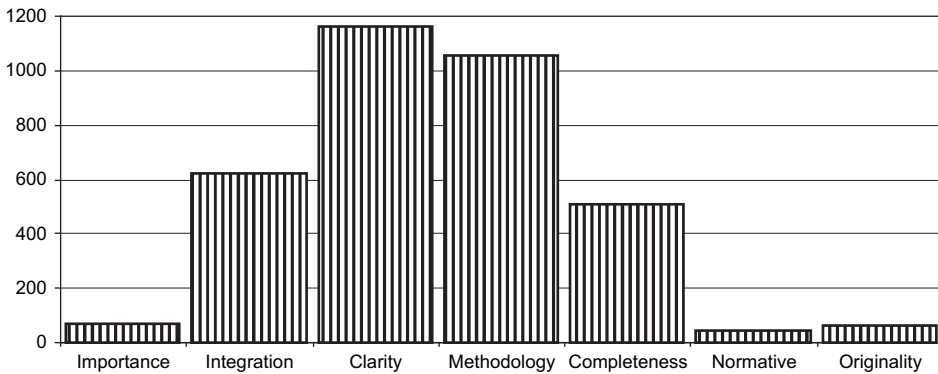**Figure 1** Total positive reviewer comments by virtue.

Figure 2 Total negative reviewer comments by sin.

Figure 2 charts negative comments and reflects a pattern rather similar to Figure 1 with two exceptions. First, note the much higher values reflected in the raw numbers of the vertical axis as sins outnumber virtues by about 9 to 1. Second, negative comments as coded on the importance dimension are relatively rare. So it turns out that only in the case of the importance dimension do virtues appear to outnumber sins.

Which sins and virtues are most highly associated with the decision to publish a submission? Figures 3 and 4 report the average number of positive and negative comments per review set for each of our seven dimensions. One might expect that acceptance recommendations would be associated with a relative predominance of positive comments and that is indeed what we find in Figure 3. And, accordingly,
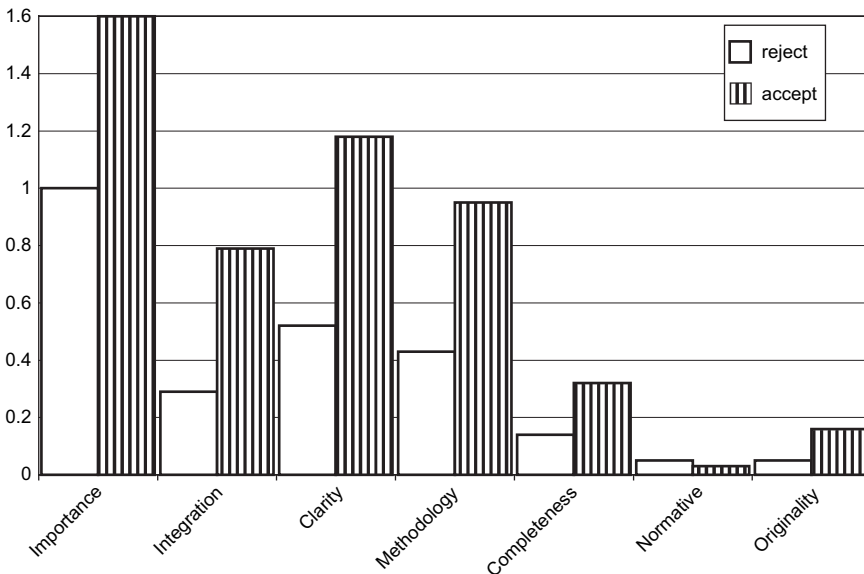

Figure 3 Average distribution of virtues in rejected and accepted manuscripts.
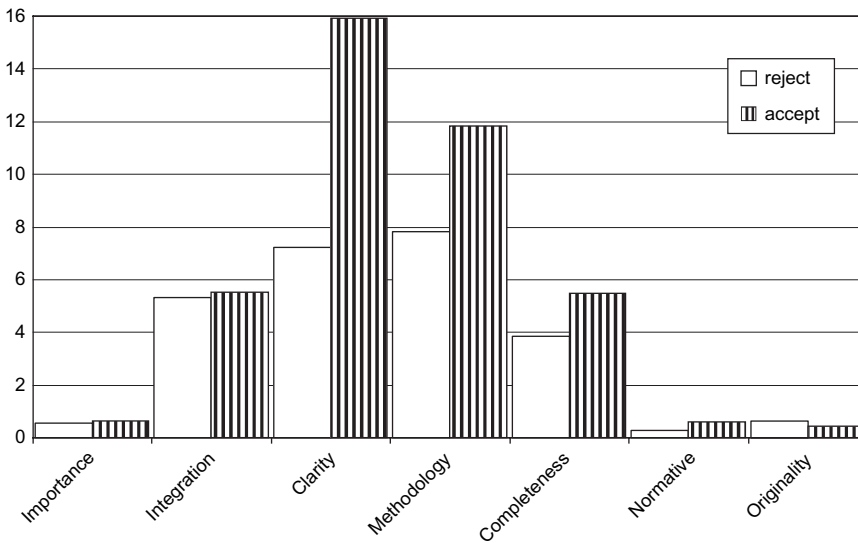
Figure 4 Average distribution of sins in rejected and accepted manuscripts.

one might expect a predominance of negative comments in the reject column, presumably to justify to prospective authors why their submissions do not measure up. But here, a surprise. We find a dramatically larger number of negative comments on average for accepted papers, especially on the dimensions of clarity, methodology, and completeness. On closer examination, it turns out to make sense in terms of the psychology of the hard-working volunteers called upon to provide the reviews. If the overall importance and theoretical integration of a submission is weak, the reviewers simply do not bother to spend a lot of time with constructive but negative comments on such things as clarifying an argument and strengthening the methodological presentation. On papers likely to be published, reviewers may go to greater length to indicate how the authors might be more complete and clear in their presentation.

This appears to be the academic journal review equivalent to the hoped-for thick envelope from a favored college to which one has applied—the more the reviewers say, positive and negative, the more interest in the submitted paper. So we examined the cross-tabulation of the total number of reviewer words and the likelihood of publication and ran a rudimentary linear least squares and determined that roughly for every additional 1,000 words of reviewer comments (including all dimensions), one's chance of acceptance increases approximately 9%.

Next, we generated an overall index for each of the seven dimensions by taking the number of virtues and subtracting the number of sins in each case. Because of the heavy predominance of negative comments, index scores tended to be negative, and the variance of the index is also dominated by the larger number of negative comments, but such an index seemed to us to be the fairest representation of what reviewers say and do. We then regressed the seven indices on the reviewer

recommendation to accept or to reject the submission. Because the dependent variable is dichotomous, we calculated logistic regression coefficients as reported in Figure 5. We repeated the exercise for the ultimate decision to publish and reported the results in Figure 6.

In both cases, as expected, the dimensions of theoretical integration and originality proved to be the best predictors of the reviewers' and editors' ultimate decision that a submission was worthy of publication. So we have found some evidence of an intriguing pattern in the review process involving an interaction among the evaluative dimensions. Broadly addressing an important topic is good but, it would appear, everybody does that. The more difficult challenge is doing so in a way that clearly engages accumulated theory and says something new and original. Getting criticized by reviewers on clarity, methods, and completeness goes with the territory, and such criticism may be tied to a positive evaluation of theory and originality. A quick examination of the bivariate correlations of the number of positive and negative comments on the seven dimensions revealed that, yes, positive evaluations of importance and theoretical integration were positively correlated with the number of negative comments on clarity and methodology.

The next stage of our analysis turned out to be the most difficult. Given the limits of the raw data available and of our coding process, we have thus far come to the conclusion that theory is king. Without a clear connection to an identifiable theoretical corpus and providing an original contribution to that corpus, pristine prose or magnificent methods do not provide the key to scholarly recognition. But the question remains—connection to what theory? If by theory each subspecialty in communication abides by Jensen's (2000) paradoxical dictum—"you have your theory and I'll have mine" (p. 28)—then there may be grounds for something less than optimism about theoretical integration and an evolving, agreed-upon,
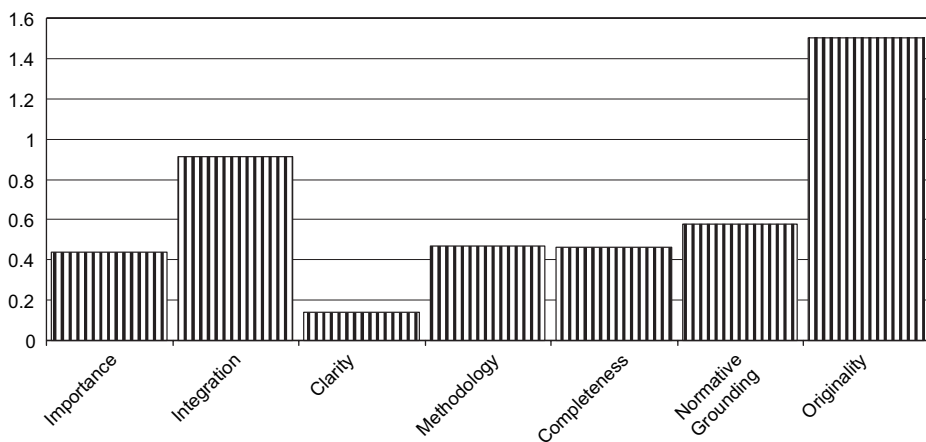

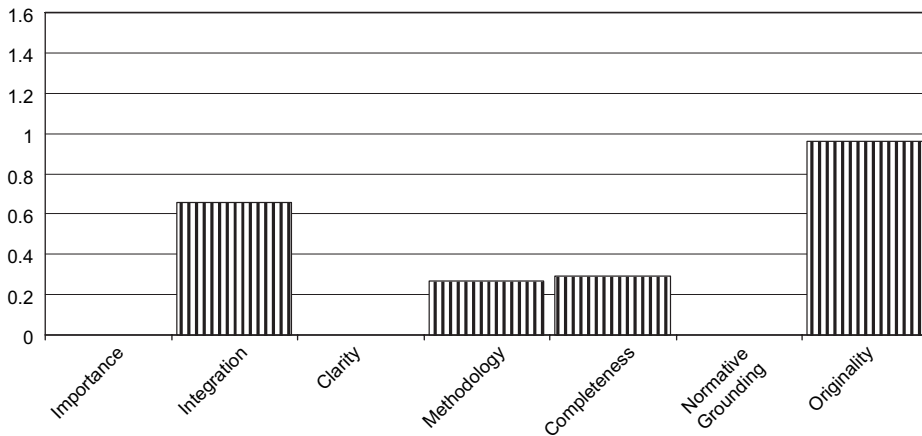**Figure 5** Logistic regression on reviewer accept recommendation.

Figure 6 Logistic regression on ultimate acceptance for publication.

structured relationship among foundational issues within the field. So we ask—when reviewers say "theory," what do they mean?

First of all, there appears to be something of a disjuncture between what communication theorists define as the central theoretical traditions of communication scholarship and what journal article authors write about or at least that is evidently true in the *Journal of Communication*. There may be a different pattern in journals specializing in theoretical issues such as *Communication Theory* or perhaps in other more specialized journals. We coded each review set using Stephen Littlejohn's widely used typology of the central theoretical traditions of the field (Littlejohn & Foss, 2005 and previous editions). The Littlejohn typology was also a central element of Craig's (1999) influential "Communication Theory as a Field." Although the typology has evolved a bit in various editions of the Littlejohn text, it remained focused on six fundamental and philosophically grounded theoretical traditions. Although these traditions may reflect the roots of communication scholarship and comprise an entirely appropriate basis for introducing students to the field, they do not appear to accurately capture the range of what active communication scholars are currently doing as 87% of the submissions in our sample were concentrated in only two of the six Littlejohn categories—the sociopsychological (49%) and sociocultural (38%) traditions. (The semiotic/rhetorical tradition corresponded to 5%, phenomenological to 1%, and cybernetic to 0%.)

Our next step was to try to rework the existing theoretical typologies to find one that was more amenable to the diversity of theoretical traditions evident in our sample. We matched up the submission content as critiqued in the reviews with both the 20 divisions of International Communication Association (ICA) and the 26 theoretical traditions in communication research identified by Bryant and Miron (2004) but given our sample size, both represented an overly fine-grained categorization with too few

submissions possible in any one category for meaningful analysis. We then turned to the graduate-level course syllabi from a half dozen major communication schools and departments in search of categories of course requirements and uncovered four domains of coursework that were prominently used. This typology captures the borderlines in several of the intellectual divides in communication research traditions including relative emphasis on qualitative versus quantitative methods, individual versus institutional, and media effects versus cultural emphases. The four topical clusters are (a) cultural studies (15 submissions), (b) institutional and policy (including political economy and technology) (17 submissions), (c) political communication and public opinion (38 submissions), and (d) mass communication (25 submissions). Several studies did not fit in any of these categories including work on interpersonal communication and organizational communication and those were excluded from the analysis (22 submissions) to try to keep the analytic distinctions as clean as possible. Then, we reran the analysis of the prominence of each of the seven sins and virtues and the correlations of sins and virtues with acceptance. Our categories are ad hoc, but we felt the analysis might reveal whether different subdomains in the fragmented world of communication research exhibited identifiably different criteria of quality scholarship. We knew at the outset from the annual reports of the *Journal of Communication* that the rates of paper acceptance were roughly equivalent in each topical and methodological area, and we knew from the peer review literature that criteria of evaluation are remarkably similar across different sciences, social sciences, and humanities fields, so we were not expecting big differences. But three findings did become evident in the subtopical reanalysis.

First, the cultural studies community is relatively loquacious. (Some readers may have expected as much.) Review sets in that category averaged a full 1,000 words more per review than the other fields. But all that commentary (most of it negative, of course) did not lead to a lower acceptance rate; it was, in fact, the highest at 40% acceptance (recalling that the acceptance rate for our full sample including supplement is 29%.) Analysis of variance with the Tukey honestly significantly different (HSD) procedure revealed that the word length difference was significant, but the acceptance rate for practical purposes was marginally significant; that is, that cultural studies acceptance rate was statistically significantly higher than only one of the other categories. Second, reviewers in the Cultural Studies category are 12 times more likely to use the normative connection dimension than the other subfields. We note that this does not necessarily mean that normative issues are not intertwined in the substance and in the evaluations of other subfields but that in the evolved tradition of scholarly discourse in these fields, such issues are not publicly articulated. The third general finding was that the patterns for political communication and mass communication looked very similar to each other in the dimensions of evaluations used and in the balance of positive and negative comments. Cultural studies reviewers, as noted, generally had the most to say, positive and negative. Reviewers in the institutional/policy subfield were the least positive noting particularly that submissions were particularly lacking in theoretical integration and originality.

When reviewers say theory, across each of these subfields, to what do they refer? Here, we returned to the actual text of the reviews to see if we are able to identify

a pattern in the language of the reviewers. Rather than reading reviews linearly from beginning to end, we return to just those paragraphs or phrases coded, for example, as positive or negative theoretical integration in search of patterns evident in the comparison of language use across reviews. Such a process is not subject to the verification of statistical significance, but we did encounter the following.

When reviewers say theory, they generally invoke one of four patterns of commentary. The first is the use of the word "theory" in communication research as a self-evident term of art with no explication by the reviewer of what does or does not qualify as theory. ("Explain the implications of your results for communication theory.") The second pattern is to equate theory with hypothesis testing about specific variables, usually some ongoing debate in the literature about whether two factors are or are not correlated or causally linked, whether some variable is trending up or down, and whether effects should be characterized as large or small. ("The authors describe two dominant theoretical perspectives that have been invoked *to rationalize the first two of these three causal possibilities*.").

The third pattern is by far the most common. Here, a theoretical corpus is identified by a shorthand phrase or the name of one of the seminal researchers in a specialized field of inquiry ("This manuscript seeks to apply *social comparison theory* to help us understand the uses of reality television."). The fourth pattern, perhaps the least frequently evident of the four, is to invoke theory in the sense of an accumulating paradigm of logically connected, law-like statements, more akin to a textbook definition of theory. This usage, when evident, is most frequently in reference to the ideals of novelty, originality, and challenges to received wisdom ("This paper would be greatly strengthened if the author *proposed a shift* in a single core assumption of the *feminist paradigm*. Specifically, the author should *propose a shift* to a *constructionist/meaning centered paradigm* in which communication is the primary tool in the construction, resistance, and maintenance of meaning."). Note that in this final pattern, the reviewers do not satisfy themselves with ritualistic mention of a theory (as is the case in the third pattern) but instead discuss how the work being reviewed could contribute to advancing theory construction. It is this fourth dimension of the invocation of theory in the journal review process that most attracted our attention. It requires the most energy and attention from the reviewers requiring them to think through if the analysis at hand does in fact contribute to our accumulated understating of human communication and does more than just cite phrases or article titles that contain the word theory in them.

Discussion and conclusion

A note of caution is in order. We do not know if any of these patterns are present in other communication journals or prominently in research areas such as interpersonal or organizational communication which appear less frequently than mass communication and related fields in the *Journal of Communication*. We note further that the less than ideal intercoder reliability and unitization measures indicate

a frustratingly high level of measurement noise mixed in with the "signal" in these findings. Given these limitations, the signal we find indicates that virtuous communication researchers are particularly skillful at linking their analyses to a recognized theoretical corpus of research and in so doing, refining and extending that corpus. Accordingly, they are more likely than their more sinful colleagues to get their papers published. Such an observation is certainly not news and not contrary to what communication researchers should desire and expect in the review process. But the essence of the review process is inadvertently obscured by the longstanding (and certainly appropriate) tradition in scholarship to protect the anonymity of individual scholars. Article reviews, research proposal evaluations, and tenure review letters are carefully protected as private communication restricted to the eyes of those in the institutional processes who need to know.

Our limited effort here represents an attempt to analyze and publicize and critique an evaluative process central to the identity of the field without evaluating specific research or researchers. Our analysis leads us to conclude that strategically theoretical integration and originality ought to be foremost in the minds of prospective contributors as it will be in the minds of their reviewers. But stepping back and thinking about the field as a whole, what strikes us as a weakness in the process is the variability and ambiguity of what is accepted as adequate theory. We hearken back to the literature cited in Footnote 3 which identifies a surfeit of theories but so little agreement on which are significant or how these various epistemic traditions are linked or contribute to each other.

We echo the conclusions of many analysts who precede us in calling for more dialogue and meaningful cross-citation among subfields and subtheories. Rising above one's narrow specialization to make linkages to broader theoretical questions is hard and may not be professionally rewarded in the field proportionate to its difficulty and importance. Perhaps, Merton's (1968) classic endorsement of theory of the middle range merits renewed consideration and emulation. Furthermore, we note with concern that reviewers' identification of a submission as an original or novel contribution to the literature, although it is an important precursor of publication, is, in fact, a surprisingly rare occurrence evident among only 17% of the successful submissions and 5% of the unsuccessful ones. It is difficult to tell from the casual and sometimes ritualized language of the review process, but it may be evidence that we are in something of a rut.

Armstrong (1980) reports on a study from a parallel field in which he asked scholars by random assignment to evaluate one of two versions of a set of analytic arguments. One version was a purposely obscured and obfuscated variant of the original text. His scholars systematically expressed a preference for the obscured version. Not a good sign. Armstrong speculates that subconsciously individuals may attribute higher intellectual merit to arguments they have to struggle to understand. If this is in evidence in communication, it is an impulse to be acknowledged and resisted.

An attentive self-consciousness among communication researchers, an aspiration to clarity about what should and should not qualify as theory, an openness to

falsifiability, and a visibility of the evaluation process while protecting the anonymity of individual researchers would seem to be worthy and achievable goals. The *Journal of Communication* at the behest of Gerbner (1983b) some years ago started a tradition of addressing these questions at an interval of every 10 years. Perhaps, it is not in our interest to patiently wait out the decade.

Notes

1   These findings leave the dramatic discipline-based differences in journal rejection rates a bit of a mystery, perhaps best explained by differences in academic cultures concerning the decision to submit.
2   Thus, in Table 1, the Bonjean and Hullum study using the phrase "important contribution" is listed twice under our rows "importance" and "originality."
3   Trying to understand the roots of theoretical fragmentation in communication research was the central driving motivation behind the study. Craig's (1999) oft-cited observation spurred us on: "Communication theory as an identifiable field of study does not yet exist. Rather than addressing a field of theory, we appear to be operating primarily in separate domains…Except within these little groups, communication theorists apparently neither agree nor disagree about much of anything. There is no canon of general theory to which they all refer. There are no common goals that unite them, no contentious issues that divide them. For the most part, they simple ignore each other" (p. 119). See also Beniger (1993), Bryant and Miron (2004), Donsbach (2006), Gerbner (1983b), Katz, Peters, Liebes, and Orloff (2003), Levy and Gurevitch (1993), Livingstone (1993), Pool (1983), Rogers and Chaffee (1993), and Swanson (1993).
4   The *r* was calculated for the aggregate number of codes (difference between total sins and total virtues) identified by the two coders for each of the seven criteria per review set. Our content analysis was not based on defining a fixed unit of a sentence or paragraph and then coding each as a zero (no relevant text) or one or more of the seven positive or seven negative categories. Coders simply proceeded until a codable comment was identified, at which point, an appropriate code was recorded. This may be a cause for technical concern about unitization. It leads to the prospect that one coder may interpret a passage of text as generating, say three "clarity" codes, whereas another coder interprets the entire passage as all about the same basic point and records only one "clarity" code. Accordingly, and in response to a concerned reviewer, we calculated the Guetzkow's *U* statistic (see Guetzkow, 1950 for details) representing the proportion of unitizing disagreements (using the index values for the review set as a whole), and it is .149 which is higher than the desired .05.
5   CCC was also calculated for the aggregate number of codes (difference between total sins and total virtues) identified by the two coders for each of the seven criteria (e.g., clarity) per review set.

References

Armstrong, J. S. (1980). Unintelligible management research and academic prestige. *Interfaces*, **10**, 80–86.

Armstrong, J. S. (1997). Peer review for journals: Evidence on quality control, fairness, and innovation. *Science and Engineering Ethics*, **3**, 63–84.

Bakanic, V., McPhail, C., & Simon, R. J. (1987). The manuscript review and decision-making process. *American Sociological Review*, **52**, 631–642.

Bakanic, V., McPhail, C., & Simon, R. J. (1989). Mixed messages: Referees' comments on the manuscripts they review. *Sociological Quarterly*, **30**, 639–654.

Beniger, J. R. (1990). Who are the most important theorists of communication? *Communication Research*, **17**, 698–715.

Beniger, J. R. (1993). Communication—Embrace the subject, not the field. *Journal of Communication*, **43**(3), 18–25.

Bonjean, C., & Hullum, J. (1978). Reasons for journal rejection: An analysis of 600 manuscripts. *PS: Political Science and Politics*, **11**, 480–483.

Braxton, J. M., & Hargens, L. L. (1996). Variation among academic disciplines: Analytical frameworks and research. In J. C. Smart (Ed.), *Higher education*: *Handbook of theory and research* (Vol. 11, pp. 1–46). Boston: Kluwer Academic.

Bryant, J., & Miron, D. (2004). Theory and research in mass communication. *Journal of Communication*, **54**, 662–704.

Bunz, U. (2005). Publish or perish: A limited author analysis of ICA and NCA journals. *Journal of Communication*, **55**, 703–720.

Campanario, J. M. (1998). Peer review for journals as it stands today—Part 1. *Science Communication*, **19**, 181–211.

Chew, F. S. (1991). Fate of manuscripts rejected for publication in the AJR. *American Journal of Roentgenology*, **156**, 627–632.

Cicchetti, D. V. (1991). The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences*, **14**, 119–186.

Cole, J. R., & Cole, S. (1981). *Peer review in the National Science Foundation*. Washington, DC: National Academy Press.

Cole, S., Rubin, L., & Cole, J. R. (1977). Peer review and the support of science. *Scientific American*, **237**(4), 34–41.

Craig, R. T. (1999). Communication theory as a field. *Communication Theory*, **9**, 119–161.

Donsbach, W. (2006). The identity of communication research. *Journal of Communication*, **56**, 437–448.

Eisenberg, N., Thompson, M. S., Augir, S., & Stanley, E. H. (2002). 'Getting in' revisited: An analysis of manuscript characteristics, reviewers' ratings, and acceptance of manuscripts in Psychological Bulletin. *Psychological Bulletin*, **128**, 997–1004.

Fiske, D. W., & Fogg, L. (1990). But the reviewers are making different criticisms of my paper! *American Psychologist*, **45**, 591–598.

Gerbner, G. (Ed.). (1983a). *Journal of Communication*, **33**(3) Ferment in the field [Special issue].

Gerbner, G. (1983b). The importance of being critical—In one's own fashion. *Journal of Communication*, **33**(3), 355–362.

Guetzkow, H. (1950). Unitizing and categorizing problems in coding qualitative data. *Journal of Clinical Psychology* **6**, 47–58.

Guetzkow, J., Lamont, M., & Mallard, G. (2004). What is originality in the humanities and the social sciences? *American Sociological Review*, **69**, 190–212.

Hackett, E. J., & Chubin, D. E. (2003). *Peer review for the 21st century: Applications to education research*. Washington, DC: National Research Council.

Hargens, L. L. (1988). Scholarly consensus and journal rejection rates. *American Sociological Review*, **53**, 139–151.

Harnard, S. (1979). Creative disagreement. *Sciences*, **19**, 18–20.

Jensen, K. B. (2000). On the edge: A meta-analysis of the state of media and communication research. *Nordicom Review*, **21**(2), 23–30.

Katz, E., Peters, J. D., Liebes, T., & Orloff, A. (Eds.). (2003). *Canonic texts in media research: Are there any? Should there be? How about these?* Cambridge, UK: Polity Press.

Lamont, M., & Mallard, G. (2005). *Peer evaluation in the social sciences and the humanities compared: The United States, the United Kingdom, and France*. Ottawa, Canada: Social Sciences and Humanities Research Council of Canada.

Lamont, M., & Molnar, V. (2002). The study of boundaries in the social sciences. *Annual Review of Sociology*, **28**, 167–195.

Levy, M., & Gurevitch, M. (Eds.). (1993). *Journal of Communication*, **43**(3 & 4) [Special issue].

Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, **45**, 255–268.

Lin, Y., & Kaid, L. L. (2000). Fragmentation of the intellectual structure of political communication study: Some empirical evidence. *Scientometrics*, **47**, 143–164.

Littlejohn, S. W., & Foss, K. A. (2005). *Theories of human communication* (8th ed.). Belmont, CA: Wadsworth Publishing.

Livingstone, S. (1993). The rise and fall of audience research: An old story with a new ending. *Journal of Communication*, **43**(3), 5–12.

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research*, **28**, 587–604.

Merton, R. K. (1968). *Social theory and social structure*. New York: The Free Press.

Peters, D. P., & Ceci, S. J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences*, **5**, 187–195.

Petty, R. E., Fleming, M. A., & Fabrigar, L. R. (1999). The review process at PSPB: Correlates of interreviewer agreement and manuscript acceptance. *Personality and Social Psychology Bulletin*, **25**, 188–203.

Pool, I. d.S. (1983). What ferment? A challenge for empirical research. *Journal of Communication*, **33**(3), 258–261.

Rice, R. E., Borgman, C. L., & Reeves, B. (1988). Citation networks of communication journals, 1977-1985. *Human Communication Research*, **15**(2), 256–283.

Rogers, E. M., & Chaffee, S. H. (1993). The past and future of communication study: Convergence or divergence. *Journal of Communication*, **43**(3), 125–131.

Scott, W. (1970). Inter-referee agreement on some characteristics of manuscripts submitted to the Journal of Personality and Social Psychology. *American Psychologist*, **29**, 698–702.

Straub, D., Ang, S., & Evaristo, R. (1994). Normative standards for MIS research. *Data Base*, **25**(1), 21–34.

Swanson, D. (1993). Fragmentation, the field and the future. *Journal of Communication*, **43**, 163–172.

Wolff, W. M. (1970). A study of criteria for journal manuscripts. *American Psychologist*, **25**, 636–669.

Zuckerman, H., & Merton, R. K. (1971). Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system. *Minerva*, **9**(1), 66–100.

Les sept péchés mortels de la recherche en communication

W. Russell Neuman

Roei Davidson

Sung-Hee Joo

Yong Jin Park

Ann E. Williams

University of Michigan

Résumé

Nous avons analysé les copies anonymes des commentaires complets des lecteurs de 120 soumissions récentes au *Journal of Communication*. Nous avons tenté d'identifier les « péchés » et les « vertus » scientifiques les plus fréquemment mentionnés par les lecteurs et les plus étroitement associés à la décision de publier la soumission. Nous avons évalué les niveaux d'accord inter-lecteurs et jaugé les schémas d'évaluation dans différents sous-domaines de la recherche en communication. Un lien explicite à un corpus théorique clairement identifié et de nouveaux résultats ou perspectives furent les plus importants prédicteurs de publication. Nous commentons les ramifications de ces résultats en ce qui a trait à l'état actuel de la recherche en communication.

Die sieben Todsünden der Kommunikationsforschung

Wir analysierten die anonymisierten Reviewer-Kommentare der letzten 120 Einreichungen für das *Journal of Communication* und versuchten, wissenschaftliche „Sünden" und „Tugenden" zu identifizieren, die von den Reviewern am häufigsten erwähnt wurden und am stärksten mit der Entscheidung der Annahme des Manuskript zur Publikation verbunden waren. Außerdem haben wir das Maß an Übereinstimmung zwischen den Reviewern und Evaluationsmuster in verschiedenen Untergebieten der Kommunikationswissenschaft erhoben. Eine explizite Verknüpfung mit einem klar identifizierten theoretischen Korpus sowie neuartige Ergebnisse oder Perspektiven waren die wichtigsten Prädiktoren für eine Publikation. Wir diskutieren die Konsequenzen dieser Ergebnisse für die aktuelle Kommunikationsforschung.

La Agresión Indirecta en las Películas de Dibujos Animados de Disney

Sarah M. Coyne

Brigham Young University

Emily Whitehead

University of Central Lancashire at Preston, England

Resumen

Los dibujos de caricaturas para niños han sido documentados como alguno de los programas más violentos de televisión actualmente en el aire, mostrando cerca de 3 veces más que la cantidad de violencia por hora de los programas que no son para niños (Wilson et al., 2002). No obstante, la violencia no es la única forma de agresión en televisión. La agresión indirecta (por ejemplo, el chusmerío, la distribución de rumores, la exclusión social) fue hallada en TV en proporciones que exceden los niveles de violencia corrientes (Coyne & Archer, 2004). A pesar de que la agresión indirecta (también llamada agresión relacional) ha sido examinada en programas populares de TV para adolescentes, no ha sido examinada en programas de TV ó películas para niños. Por lo tanto, el presente estudio tiene como objetivo examinar la frecuencia y la representación de la agresión indirecta en las películas animadas de Disney para niños. En general, las películas de Disney representaron 9.23 veces por hora de agresión indirecta. Cuando el conteo por personaje fue tenido en cuenta, la agresión indirecta fue representada por mujeres y hombres en niveles similares. La agresión indirecta fue representada más probablemente como injustificada y por los personajes "malos". Los personajes "altos en SES" entablaron agresión indirecta con más frecuencia que los "bajos en SES" ó aquellos personajes "de clase media." Comparados con la cantidad de violencia en los programas de TV para niños, la cantidad de agresión indirecta de las películas de Disney fue mucho menor y usualmente representada en formas que no facilitan la imitación.

传播学研究的七宗死罪

W. Russell Neuman

Roei Davidson

Sung-Hee Joo

Yong Jin Park

Ann E. Williams

密歇根大学

我们分析了最近 120 篇《传播学期刊》稿件之完整的、匿名的评审意见，以界定评审员最经常提及的、与他们决策（即是否发表）最相关的学术上的"罪"和"德"。我们评估了评审员之间的一致度以及在传播学研究不同子领域的评估方式。预测投稿能否发表的最重要的因素为论文是否与明确界定的理论文集相关和是否有创造性的发现/角度。我们讨论了这些发现对当前传播学研究的意义。

커뮤니케이션 연구에서의 7대 죄악

W. Russell Neuman

Roei Davidson

Sung-Hee Joo

Yong Jin Park

Ann E. Williams

University of Michigan

요약

우리는 최근 *Journal of Communication*에 투고된 120개 논문들에 대해 익명으로 완결된 논문 심사평을 분석하였는바, 이는 논문 심사자들에 의해 가장 빈번하게 언급되었으며 논문의 출판결정에 가장 직접적으로 연계된 학문적인 죄악과 미덕을 확인하려고 한 것이다. 우리는 커뮤니케이션 학문의 다양한 하부 영역에서 상호 평가자 동의 수준과 평가 형태를 분석하였다. 명료하게 확인된 이론적인 집성과 참신한 발견들 또는 전망들에 대한 명백한 연계가 투고된 논문의 출판을 결정하는 가장 중요한 예측도였다. 우리는 현재 커뮤니케이션 연구를 위하여 이러한 발견들의 효과들 논의하였다.