Psychological Bulletin
2002, Vol. 128, No. 6, 997–1004

0033-2909/02/$5.00 DOI: 10.1037//0033-2909.128.6.997

"Getting in" Revisited: An Analysis of Manuscript Characteristics, Reviewers' Ratings, and Acceptance of Manuscripts in *Psychological Bulletin*

Nancy Eisenberg, Marilyn S. Thompson, Susan Augir, and Elizabeth Harris Stanley
Arizona State University

Getting published in the premier journals in psychology can be a daunting task, especially for the novice. Young scholars or those with limited experience publishing often ask editors what factors affect the probability of a manuscript being accepted and what they should do to enhance their chances of getting published. Of course, many factors affect whether a manuscript is accepted, including who the acting editor and reviewers happen to be and the quality and orientation of the journal to which a paper is submitted. Nonetheless, it is highly likely that certain characteristics of manuscripts contribute in a systematic manner to reviewers' evaluations of them and, ultimately, to a final editorial decision.

At the close of Robert J. Sternberg's editorship of *Psychological Bulletin*, Sternberg, Hojjat, Brigockas, and Grigorenko (1997) published an analysis of reviewers' ratings and their relation with the final editorial decisions for manuscripts handled by the editor and one of the associate editors between January 1, 1993, and February 29, 1996. In their analyses, they considered each item on the reviewer rating form as addressing an analytical, creative, or practical aspect of the manuscript. The analytically oriented items concerned the extent to which "the authors completely and correctly analyzed, evaluated, and reflected on the extant literature reviewed" (p. 321). These items included completeness of coverage, balance and fairness in the coverage of alternative views, and accuracy of information. Thus, to score high on this dimension, authors needed to accurately determine what articles should be reviewed, analyze their views and findings in a balanced manner, and provide an accurate summary of the issue in question.

Sternberg et al. (1997) included among the creativity items those that assessed whether the article made a novel contribution to the field. The items that tapped this dimension included scientific importance of the topic, theoretical orientation of manuscript, and contribution to the field reviewed. A creative article, then, was viewed as one that involved "taste and judgment in topic selection; . . . the formation of a new theory or application of an existing theory to diverse scientific findings; . . . and a substantial incremental contribution to a field at the time of publication" (p. 321).

The items Sternberg et al. (1997) labeled as *practically oriented* concerned the extent to which manuscripts reflected authors' tacit knowledge of the features of a manuscript that make it publishable in a particular journal. For *Psychological Bulletin*, these included "generalized writing skills needed to communicate effectively and persuasively to a diverse audience" (p. 321). The items considered as tapping this dimension were scope of review; appropriateness of manuscript for journal; interest for a broad audience of psychologists; existence and clarity of a take-home message; existence of and persuasiveness in arguing for a well-articulated point of view; organization of manuscript; and clarity, coherence, and conciseness of prose.

Although Sternberg et al. (1997) initially classified their rating items into the aforementioned three classes, an exploratory factor analysis of reviewers' item ratings for initial manuscript reviews (not revisions) suggested two factors, which they described as the overall appropriateness of manuscripts for any scientific journal and the overall appropriateness of manuscripts specifically for *Psychological Bulletin.* In addition, they found that all items had statistically significant polyserial correlations with the final editorial decision regarding acceptance versus rejection of a manuscript. The items most strongly correlated with the final editorial decision (in order of strength) were the reviewer's recommendation regarding acceptance/rejection and contribution to the field reviewed, followed by appropriateness of the manuscript for the journal, existence of and persuasiveness in arguing for a well-articulated point of view, completeness of coverage, and theoretical orientation of the manuscript.

Our goal in this article is to reexamine manuscript characteristics that affect reviewers' recommendations regarding acceptance versus rejection of manuscripts and the acting editor's final decision. Early in Nancy Eisenberg's editorship, she devised a rating sheet for reviewers that was used subsequently by all associate editors. Many items on the full 13-item scale used under her editorship paralleled those used on Robert J. Sternberg's rating form with minor wording changes, although multiple items were replaced or rewritten to capture more specific technical aspects of the manuscript (e.g., quality of the analyses [if relevant]). Building on Sternberg et al.'s (1997) analysis and considering the changes in the rating form, we expected to find three dimensions of manuscript characteristics to be useful for predicting publication in

Editor's Note. Daniel Reisberg served as the action editor for this article.—NE

Nancy Eisenberg, Marilyn S. Thompson, Susan Augir, and Elizabeth Harris Stanley, Department of Psychology, Arizona State University.

We gratefully acknowledge the assistance of various associate editors (Scott M. Monroe, Daniel Reisberg, Lynn C. Robertson, Kenneth J. Sher, Arthur P. Shimamura) who collected the data analyzed in this article, as well as Christi Coxsey's assistance in entering and organizing some of the data.

Correspondence concerning this article should be addressed to Nancy Eisenberg, Department of Psychology, Arizona State University, P.O. Box 871104, Tempe, Arizona 85287-1104. E-mail: nancy.eisenberg@asu.edu

Psychological Bulletin: fit of the manuscript for *Psychological Bulletin,* technical and scientific quality, and strength and clarity of the written argument.

According to the Sternberg et al. (1997) article, 312 manuscripts were included in their analyses; these papers were handled by two different acting editors. In our analyses, we used a larger sample of manuscripts and included data from six acting editors (the editor and four associate editors, plus a small number of manuscripts handled by an additional associate editor). In addition, we evaluated extensively the factor structure of the item ratings and the relation of three manuscript dimensions (fit, technical and scientific quality, and strength and clarity of the argument) to reviewers' recommendations and to the acting editor's final action.

Method

Sample

Our sample consisted of 706 manuscripts submitted to *Psychological Bulletin* from 1996–2001 that were externally reviewed using an 8-item (1996 only) or 13-item evaluation form. Of these, 131 were initially submitted in 1996, 125 in 1997, 126 in 1998, 139 in 1999, 121 in 2000, and 64 in 2001. The manuscripts were handled by six different action editors; the distribution of papers among them included 354 manuscripts handled by the editor (50% of the total), followed by 145, 88, 55, 49, and 15 papers for five associate editors. In addition to the first versions, we also summarized separately reviewer ratings from the final (i.e., last reviewed) versions of 147 of these manuscripts that were sent out for external review at least a second time. Of these 147 externally reviewed revisions, 120 were second revisions, 26 were third revisions, and 1 was a fourth revision.

We excluded from our analyses initial or revised versions of manuscripts that did not undergo peer review. There were three general situations in which a version of a manuscript was not externally reviewed: (a) the manuscript was rejected by an editor upon initial submission (approximately 20% of initial submissions), typically because it clearly was inappropriate for the journal (e.g., single empirical submissions, very scant reviews, or position papers not containing a review of the research literature); (b) the manuscript was an invited commentary or reply to a previously published article (approximately 4% of initial submissions); and (c) it was a minor revision of a previously reviewed manuscript (approximately 54% of revised submissions). We note that 89% of these revised manuscripts that were not sent for additional external review were eventually accepted for publication. Moreover, although nearly all reviewers provided some type of written evaluation of the manuscript, some neglected to rate the manuscript using the item rating form provided. Consequently, their evaluations are not represented in our quantitative averages of ratings for a manuscript, although their written evaluations were used in the editor's assessment of a piece.

Invited commentaries and replies were excluded because they did not undergo the typical review process. The editor or an associate editor selected individuals to write invited commentaries, which were usually reviewed either solely by the solicitor or by the editor and additional reviewers who knew about the special nature of these manuscripts. No invited commentaries were rejected due to the careful selection process and the context of these reviews. However, we retained in our sample seven commentaries that were unsolicited and externally reviewed.

Measures

All manuscripts were reviewed using either an 8- or 13-item manuscript evaluation form, as summarized in Table 1. Reviewers were asked to rate manuscripts on these items on a 5-point scale ranging from 1 (*low*) to 5 (*high*). In addition, they were asked to give a recommendation to the editor regarding publication. The five recommendation options included: (1) *strongly recommend acceptance* (as is or with minor revisions), (2) *recommend acceptance* (with some reservation), (3) *reject/revise/potentially publishable,* (4) *doubtful* (probably should be rejected), or (5) *reject.* For our analyses, options were recoded numerically such that recommendations ranged from 1 (*reject*) to 5 (*strongly recommend acceptance*). The numbers of external reviews requested and received for each manuscript version were noted, as was the final editorial decision to reject or accept the manuscript (coded 0 or 1, respectively). Editorial reject/revise decisions were recoded as rejections if they were never revised and resubmitted and it had been at least 1 year since the last review. The gender of the first author was coded, and if unknown, we attempted to obtain this information from the Internet or the acting editors.

Analysis

Thorough descriptive analyses of reviewer responses were conducted separately for both the initial and final set of reviews for externally reviewed manuscripts. For each item and for the overall reviewer recommendation, we computed a manuscript's average ratings across the multiple reviewers. Revised manuscripts may or may not have been sent again to the initial reviewers (although at least some of the initial reviewers usually reviewed a revision), and occasionally a revision was sent to one or more reviewers who had not previously evaluated the manuscript. In forming editorial decisions about a particular revised form of a manuscript, the editor typically considered the last review received from each reviewer in addition to those from new reviewers. Therefore, for our analyses of ratings for the 147 final externally reviewed revisions, we used the last ratings contributed by each particular reviewer in computing an average rating for any item.

Across manuscripts within each set, we computed means and standard deviations for average item ratings, average reviewer recommendation, and numbers of reviews requested and received. We computed Pearson correlations between item ratings and reviewer recommendation as well as point-biserial correlations between item ratings and the final action. Distributions of average reviewer recommendation and average item ratings were also examined. We also computed two-way contingency table analyses to assess whether the gender of the first author was related to resubmission or acceptance rates.

To compare our findings with those of Sternberg et al. (1997) and to assist in identifying the underlying manuscript dimensions that affected the reviewers' item ratings, we conducted an exploratory factor analysis of average item ratings (excluding Item 13, overall evaluation) on reviews for the initial versions. We used confirmatory factor analysis to compare statistically the two- and three-factor measurement models. Structural equation models predicting the final editorial decision from these manuscript attributes were specified and evaluated to model the relation of the three dimensions of manuscript attributes with reviewer recommendation and the likelihood for publication. Mplus 2.02 (Muthén & Muthén, 2001) was used for all structural equation modeling analyses.

Results

Descriptive Statistics

Descriptive statistics for reviewer ratings, reviewer recommendation, and the final editorial decision are shown in Table 1. Results are reported separately for initial and final externally reviewed versions of manuscripts. Item ratings for each manuscript reflect the average rating across all reviewers supplying a response for that item. On average, 4.43 reviews were requested (SD = 1.21) and 3.09 were received (SD = 0.70) for initial versions, whereas 2.98 reviews were requested (SD = 1.14)

Table 1
Statistics for Manuscript Ratings by Item on Reviewer Rating Form and Reviewer Recommendation

Variable	Initial versions (N = 706)				Final versions (N = 147)			
	M	*SD*	*r*	r_{pb}[a]	*M*	*SD*	*r*	r_{pb}[b]
1. Significance of paper topic	3.88	0.79	.54	.37	4.32	0.58	.48	.39
2. Interest to a broad audience of psychologists[c]	3.51	0.85	.54	.39	4.05	0.62	.45	.38
3. Appropriateness for *Psychological Bulletin*	3.42	1.02	.64	.43	4.15	0.72	.67	.52
4. Quality of analyses	2.67	1.12	.62	.42	3.53	0.99	.47	.38
5. Balance and fairness in coverage of alternative views[c]	3.01	0.85	.63	.37	3.55	0.73	.62	.34
6. Accuracy of information[c]	3.37	0.81	.64	.38	3.94	0.57	.52	.33
7. Adequacy of literature review	3.01	0.98	.59	.40	3.84	0.69	.59	.37
8. Theoretical contribution of manuscript[c]	2.82	0.96	.64	.39	3.51	0.76	.55	.45
9. Existence of and persuasiveness in arguing for a well-articulated point of view[c]	2.67	0.95	.76	.48	3.52	0.72	.69	.48
10. Importance of conclusions	2.79	0.95	.77	.46	3.72	0.72	.68	.50
11. Quality of writing	3.33	0.86	.54	.39	3.88	0.68	.52	.39
12. Appropriateness of the paper's length	3.19	0.90	.39	.29	3.48	0.83	.46	.32
13. Overall evaluation	2.67	0.91	.88	.55	3.60	0.66	.82	.55
Reviewer recommendation	2.42	1.02		.63	3.60	0.82		.62

Note. All correlations of item ratings with reviewer recommendation and final action were statistically significant at less than .001. r = the Pearson correlation of the average rating for each item with the average reviewer recommendation 1 = *reject;* 2 = *doubtful* (probably should be rejected); 3 = *reject/revise;* 4 = *recommend acceptance* (with some reservation); 5 = *strongly recommend acceptance* (as is or with minor revisions); r_{pb} = the point-biserial correlation of each item with final action (0 = *reject;* 1 = *accept*).
[a] Final decision was not yet complete for 55 manuscripts. [b] Final decision was not complete for 18 manuscripts. [c] These five items were not included on the 8-item version of the review form used in 1996. Sample sizes for these items range from 557–560 for r and 511–513 for r_{pb}, whereas sample sizes for the other 8 items range from 644–647 for r and 597–600 for r_{pb} (excepting Item 4, quality of analyses, for which N = 423; reviewers were instructed to respond to this item only if relevant for the manuscript).

and 2.49 were received (SD = 0.87) for final versions. These differences between the numbers of reviews requested and the numbers of reviews received exist because some reviewers declined to do the review or simply did not respond and additional reviewers often were selected in their stead. The aforementioned numbers do not include reviewers who were contacted by e-mail and declined to review a manuscript before it was sent. (The degree to which acting editors tried to line up reviewers in this manner varied considerably, but Nancy Eisenberg did so only once in a while.)

Initial versions. For the first versions reviewed, average ratings among Items 1–12 were most positive for significance of the paper topic (M = 3.88, SD = 0.79), whereas the lowest average ratings were obtained for quality of analyses (M = 2.67, SD = 1.12) and for existence of and persuasiveness in arguing for a well-articulated point of view (M = 2.67, SD = 0.95). Distributions of the average item ratings were also examined. Particularly for manuscripts that were eventually accepted, distributions were negatively skewed for the first three items—significance, interest, and appropriateness—presumably because of editorial decisions to reject those manuscripts clearly inappropriate for *Psychological Bulletin* without sending them for external review (and because of the fact that reviewers are likely to think that topics they study are important, even if the particular manuscript they reviewed was not of high quality). Boxplots showing distributions of average item ratings and reviewer recommendation are depicted in Figure 1 for initial versions of manuscripts on the basis of whether they were eventually accepted (left panel) or rejected (right panel).

All correlations of item ratings with the variables of reviewer recommendation and the final editorial action (shown in Table 1) were statistically significant at the $p < .001$ level. Ratings on the persuasiveness and importance of conclusions items had the highest point-biserial correlations with the acting editor's final action (.48 and .46, respectively), as well as the highest Pearson correlations with reviewers' recommendation (.76 and .77, respectively). Reviewers' recommendations also were substantially correlated with the editors' decisions (r_{pb} = .63). For the subset of these data used for the structural equation modeling analyses, none of the correlations between the item ratings and reviewer recommendation or final action differed by more than .05 from the estimates presented in Table 1.

Of the 651 manuscripts that were externally reviewed upon initial submission and for which final decision to accept or reject was complete, 65.3% were rejected, 29.3% were rejected with an invitation to revise and resubmit, 5.2% were accepted with minor revisions, and 0.2%—a single manuscript—was accepted. Twenty-two percent of these manuscripts were eventually accepted (78.0% were rejected).

First authors were more commonly male: 66.8% were male and 31.0% were female, and gender was unknown for 2.2% of first authors. Acceptance rates by gender were 23.4% for papers with male first authors and 20.3% for papers with female first authors. Acceptance and gender of the first author were not significantly

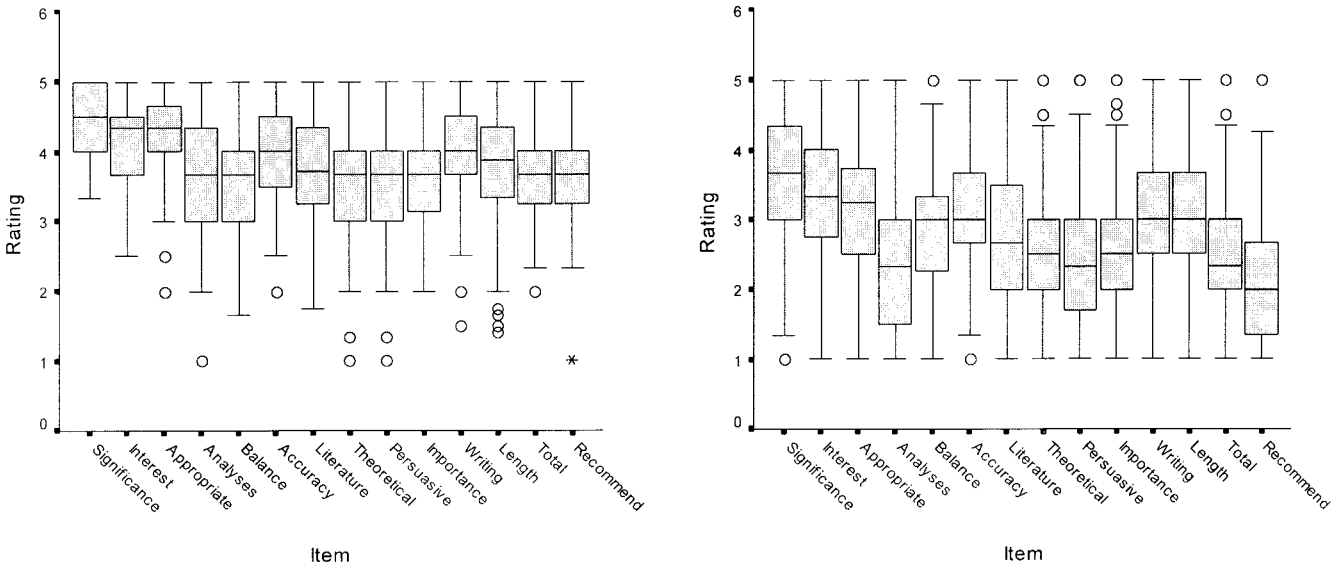

Figure 1. Boxplots of average item ratings and reviewer recommendations for initial versions of manuscripts eventually accepted (left panel; $n = 143$) or rejected (right panel; $n = 508$) for publication in *Psychological Bulletin.* In each boxplot, the box represents the interquartile range (from the 25th to 75th percentiles); the horizontal line indicates the median; the whiskers show the range of scores within 1.5 box lengths from the 25th and 75th percentiles; and circles and asterisks represent outliers that are more than 1.5 and 3.0 box lengths, respectively, from the 25th and 75th percentiles.

related in a two-way contingency table analysis and the effect size for the association was small, Pearson $\chi^2(1, N = 637) = 0.79$, $p = .38$, $\Phi = .04$.

Final versions. For final versions reviewed, we again note that the 147 final versions are the subset of the initial 706 manuscripts that were externally reviewed two or more times (120 second revisions, 26 third revisions, and 1 fourth revision). Therefore, the descriptive statistics presented for the final versions are not independent of those for the initial versions. For the final versions reviewed, as for the initial versions, ratings were most positive for significance of the paper topic ($M = 4.32$, $SD = 0.58$). However, the lowest average ratings for final versions were observed for appropriateness of the paper's length ($M = 3.48$, $SD = 0.83$), which also displayed the lowest point-biserial correlation with final action and one of the weaker correlations with reviewer recommendation. Evaluators presumably viewed a length problem as one that could be remedied. Final action by the acting editor was most highly correlated with appropriateness for *Psychological Bulletin* (.52), closely followed by importance of conclusions (.50) and existence of and persuasiveness in arguing for a well-articulated point of view (.48); the reviewer recommendation was again most highly correlated with existence of and persuasiveness in arguing for a well-articulated point of view (.69), importance of conclusions (.68), and appropriateness for *Psychological Bulletin* (.67).

There were 191 initial versions of manuscripts that had been externally reviewed, received a reject/revise decision for the first version, had known gender of the first author (71.4% male and 28.6% female), and had the final decision completed at the time of our analyses. As noted previously, reject/revise decisions were coded as rejections for the final decision if a revision was not submitted by the time of data analysis for this article and within 1 year of the last review. Revised manuscripts were submitted by approximately three quarters of those authors that had received a reject/revise decision for the first version of their manuscript. Including revisions reviewed only by the acting editor as well as those reviewed externally, revised manuscripts were received from 76.3% of male first authors and 72.2% of female first authors. Resubmission and gender were not found to be significantly related in a two-way contingency table analysis, and the effect size was small, Pearson $\chi^2(1, N = 191) = 0.34$, $p = .56$, $\Phi = .04$.

Factor Structure of the Manuscript Rating Scale

Exploratory factor analysis. To follow up on the findings of Sternberg et al. (1997) and to evaluate the effect of rewording and adding items to capture more specific technical aspects of the manuscripts, we conducted an exploratory factor analysis on Items 1–12, excluding the overall evaluation and recommendation to the editor. We used SPSS 10.0 to conduct principal axis factoring for extraction, and we applied an oblimin rotation because of our belief that the manuscript dimensions would be correlated. We used listwise deletion in the exploratory factor analysis such that the results were based on the 382 manuscripts with average ratings across reviewers provided for all 12 of these items. An examination of missing responses indicated that a number of reviewers did not respond to Item 4, which pertained to "quality of analyses (e.g., meta-analysis; respond only if relevant)." Assuming the manuscripts with missing ratings on Item 4 did not involve data analysis, leaving this item unrated was appropriate. We also conducted an exploratory factor analysis based on the sample with Item 4 omitted ($N = 554$ with listwise deletion), which yielded nearly

identical results. In addition, because some earlier manuscripts (received in 1996) were reviewed using an 8-item form, these cases were excluded from these analyses because of a lack of available ratings for 5 of the 13 items. Other methods of handling missing data were explored, including pairwise deletion and mean imputation; however, these methods have their own drawbacks (Fabrigar, Wegener, MacCallum, & Strahan, 1999), and the factor solution did not change markedly when these strategies were used to preserve cases in the analysis.

We examined both two- and three-factor solutions, anticipating that the two-factor solution would approximate the solution observed by Sternberg et al. (1997) and the three-factor solution would be more consistent with our hypothesized factors of fit for *Psychological Bulletin*, technical and scientific merit, and strength and clarity of the written argument. Rather than reporting the specific factor loadings here, we discuss our statistical comparison of the two- and three-factor models with confirmatory factor analysis and present a full structural equation model predicting final decision.

The three-factor solution was readily interpretable and accounted for 71% of the variance in item ratings (eigenvalues = 5.88, 1.67, and 0.96). The items most strongly loading on the first factor (theoretical contribution of manuscript, existence of and persuasiveness in arguing for a well-articulated point of view, importance of conclusions, quality of writing, and appropriateness of length) were associated with the strength and clarity of the arguments presented in the manuscript (with length loading only moderately). The second factor included significance of paper topic, interest to a broad audience of psychologists, and appropriateness for *Psychological Bulletin*—items addressing the fit of the manuscript for *Psychological Bulletin.* Items loading on the third factor were associated with the technical or scientific aspects of the article (adequacy of literature review, balance and fairness in coverage of alternative views, accuracy of information, and quality of analyses—with quality of analyses loading only moderately). The strength and clarity of argument factor had moderate to strong correlations with the fit and technical and scientific factors (r = .48 and .68, respectively), whereas the fit and technical factors were not as highly correlated (r = .29).

The two-factor solution accounted for 63% of the variance in items rated, and although the nature and wording differed for several items, the two factors extracted were similar to those found in the exploratory factor analysis conducted by Sternberg et al. (1997). A strong first factor was extracted on which Items 4–12 loaded—with theoretical contribution of manuscript and appropriateness of length loading only moderately. This factor reflected the general merit of the manuscript, or as Sternberg et al. described, "the overall appropriateness of a manuscript for any scientific journal," (p. 323). The second factor, on which significance of paper topic, interest to a broad audience of psychologists, and appropriateness for *Psychological Bulletin* loaded strongly, reflected the fit or overall appropriateness of the manuscript for *Psychological Bulletin.* The quality and fit factors were moderately correlated (r = .42).

Confirmatory factor analysis. Because the two-factor model can be derived from the three-factor model by collapsing the strength and clarity of argument and technical and scientific factors into one factor, manuscript quality, the models are nested and can be compared statistically using a chi-square difference test. The three factors composing this measurement model are illustrated on the left side of Figure 2. We excluded Item 4—quality of analyses—from our structural equation modeling analyses because of the lack of applicability of this item for many manuscripts. Furthermore, manuscripts were excluded if they were missing average ratings for any of the remaining 11 items or if the final decision for the manuscript was not yet complete, reducing the number to 504. Using Mplus 2.02, the measurement models were estimated using a robust maximum likelihood method that yields robust standard errors and a mean-adjusted chi-square statistic (the Satorra–Bentler scaled χ^2). The two-factor model did not fit the data well, $\chi^2(43, N = 504) = 375.21$, $p < .01$; standardized root mean square residual (SRMR) = .06; comparative fit index (CFI) = 0.88; Tucker–Lewis Index (TLI) = .85; RMSEA = 0.12. In contrast, the three-factor model demonstrated a reasonably good fit to the data, $\chi^2(41, N = 504) = 163.56$, $p < .01$; SRMR = .05; CFI = .96; TLI = .94; root mean error of approximation (RMSEA) = .08. The three-factor model was a statistically significant improvement over the two-factor model as evaluated using Satorra's (2000) chi-square difference test procedure for nested models based on the Satorra–Bentler scaled chi-square statistics (Satorra–Bentler scaled difference $\chi^2(2, N = 504) = 218.46$, $p < .01$; see also Satorra & Bentler, 1999, for computational details). All standardized factor loadings were greater than .65, except the loading for appropriateness of length on the strength and clarity of argument factor (λ = .45), and all were statistically significant at the .01 level. The strength and clarity of argument factor displayed strong correlations with both the fit and technical and scientific factors (r = .61 and .76, respectively), whereas the fit and technical and scientific factors were more moderately correlated (r = .45).

Structural Model Predicting Final Decision

The three-factor measurement model for the manuscript rating scale was then incorporated into a structural model predicting the reviewers' evaluation and, subsequently, the final editorial decision (see Figure 2). Reviewer evaluation was modeled as a latent variable with two manifest variables: the reviewers' average for overall evaluation (Item 13) and the reviewers' average recommendation to the editor for a manuscript. Final decision was a dichotomous outcome variable reflecting the editorial decision to reject or accept the manuscript. Cases with missing data were excluded listwise, leaving 500 manuscripts for analysis of the full model. The model was estimated with Mplus 2.02 using a robust weighted least squares method with robust standard errors and a mean- and variance-adjusted chi-square goodness-of-fit index. Model fit indices suggested a satisfactory fit to the data, $\chi^2(34, N = 500) = 255.45$, $p < .01$; TLI = .99; CFI = .95; RMSEA = .11. Given our sample size, it is not surprising that the chi-square statistic was statistically significant; however, the RMSEA of .11 suggests marginal fit and the CFI of .95 and the TLI of .99 are generally regarded to indicate good fit. Standardized parameter estimates are reported on the path diagram in Figure 2. All factor loadings, correlations among factors, and paths between factors were statistically significant at the .01 level. We note that because of the substantial correlations among the fit, technical and scientific, and strength and clarity of argument factors, the relative magnitudes of the standardized weights predicting reviewer eval-

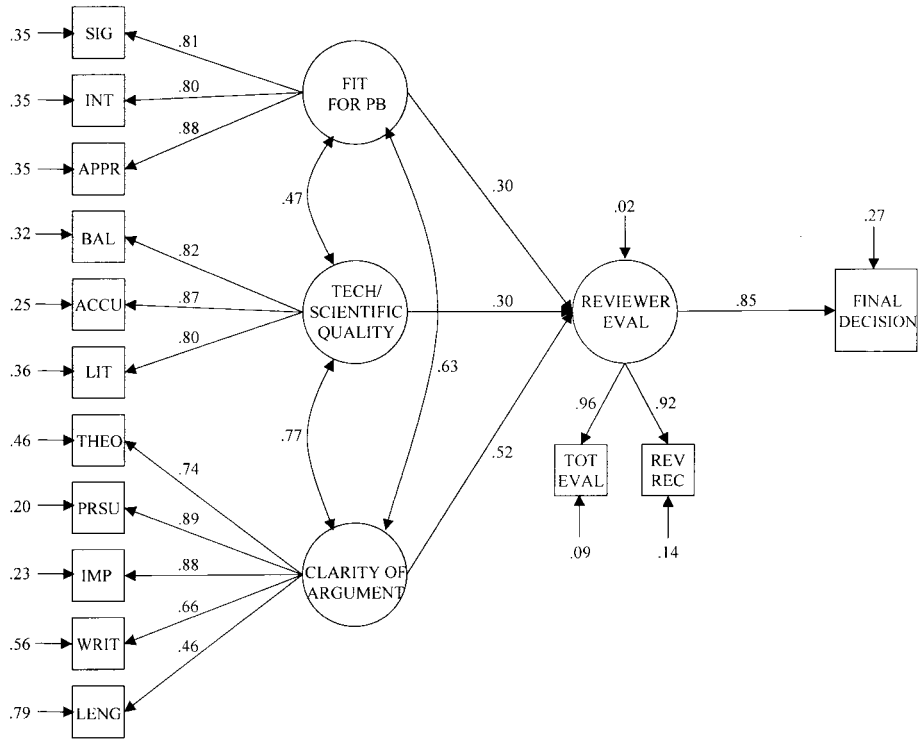

Figure 2. Structural model for final editorial decision ($N = 500$). Standardized parameter estimates are shown on the diagram; all factor loadings, correlations among factors, and paths between factors were statistically significant at the .01 level. SIG = significance of paper topic; INT = interest to a broad audience of psychologists; APPR = appropriateness for *Psychological Bulletin* (PB); BAL = balance and fairness in coverage of alternative views; ACCU = accuracy of information; LIT = adequacy of literature review; THEO = theoretical contribution of manuscript; PRSU = existence of and persuasiveness in arguing for a well-articulated point of view; IMP = importance of conclusions; WRIT = quality of writing; LENG = appropriateness of the paper's length; TECH = technical; EVAL = evaluation; TOT = total; REV REC = revision received.

uation from these factors should not be interpreted as indicative of the importance of these factors in predicting reviewer evaluation. Based on the four-factor measurement model, estimated correlations between the reviewer evaluation factor and the fit, technical and scientific, and strength and clarity of argument factors were .719, .853, and .942, respectively. These correlations are probably the best index of the relative relationships of the three factors to reviewers' recommendations to the editor.

We had also hypothesized a model including direct effects from the fit, technical and scientific, and strength and clarity of argument factors to final decision, in addition to their indirect effects on final decision through reviewer evaluation as specified in Figure 2. As an editor reads a manuscript, he or she conceptually evaluates these attributes and uses this information as well as that provided by the reviewers in making a decision about the manuscript. However, estimation of the model with these additional paths was problematic because of multicollinearity, most notably caused by strong correlations between the reviewer evaluation factor and the three rating scale factors, as described previously.

Discussion

We found that a three-factor measurement model—including fit for *Psychological Bulletin*, scientific and technical quality, and strength and clarity of the arguments—was a good representation of the dimensions underlying average reviewer ratings. These dimensions were all strongly related to reviewers' recommendations to the editor, which in turn was an excellent predictor of the editors' decision. The items existence of and persuasiveness in arguing for a well-articulated point of view and importance of conclusions were strongly correlated with both reviewers' recommendations' and the editors' final decision (especially as rated on the initial version), as was appropriateness for *Psychological Bulletin.* On the basis of the estimated correlations between the reviewer evaluation factor and the fit, technical and scientific, and strength and clarity of argument factors, it appears that the strength and clarity of the argument factor was most strongly associated with acceptance, followed by the technical and scientific quality of the paper and its fit to the journal.

Thus, authors can best increase their chances of acceptance by submitting manuscripts with clear and compelling arguments for major points that pertain to important conclusions. How does one know when these criteria have been achieved? Clearly, they involve subjective judgment. Nonetheless, for conclusions to be important, they generally must go beyond what has already been found or is already known and pertain to an interesting issue. In addition, the arguments must be rational, based on a critical review

of the existing empirical research, and must deal with alternative explanations and points of view. Furthermore, there are issues of timing that are specific to each field and topic. For example, papers sometimes are rejected because a particular field is not yet ready for a comprehensive review due to a lack of sufficient data or because a coherent picture has not yet emerged from the data. On the other hand, a paper may be rejected because a similar review has recently been published. Potential authors should also keep in mind the broad readership of the journal; indeed, some of the reviewers of a manuscript may not be in an author's area of speciality. Sometimes a broad perspective on a field is needed to make these judgments, so young authors often may benefit from the insights of their colleagues with more experience publishing in journals such as *Psychological Bulletin.*

At the same time, it is important to note that the three factors influencing reviewers' evaluations and editors' decisions were correlated, especially the strength and clarity of arguments and scientific and technical merit factors. As one might expect, authors who present clear, persuasive arguments tend to be those who do a careful job of reviewing and analyzing relevant data, have included all (or nearly all) of the relevant research (especially in a meta-analysis) and theory, and are accurate and balanced in their presentation of evidence. Nonetheless, manuscripts that fare well in the review process at *Psychological Bulletin* are not only scientifically and technically sound, but also tend to include a strong theoretical framework for their review, persuasive arguments for a well-articulated view, and conclusions of importance, all clearly communicated to the reader.

Psychological Bulletin is a relatively unique journal in some respects; for example, it is a premier journal that publishes only reviews and is targeted at a broad audience rather than a subdiscipline of psychology. Thus, the criteria for acceptance in *Psychological Bulletin* may differ somewhat from the criteria for acceptance in other journals, especially less prestigious journals or those with the mission to publish results of empirical studies relevant to a narrow audience. For example, the emphasis on a strong take-home message including important conclusions that are persuasively argued may be even more essential for publication in *Psychological Bulletin* (and *Psychological Review*) because articles often do not contain an alternative empirical contribution (except in the case of some meta-analyses) and because the findings must be of considerable interest to a broader spectrum of readers. Authors are most likely to succeed in publishing in *Psychological Bulletin* when some insights based on a body of findings have "jelled" in their thinking. In contrast, journals that publish empirical articles—even premier journals—may accept papers that used a new and interesting procedure or represented the next logical step in a line of research or discovery, even if there is not a novel conceptual contribution. Thus, the findings presented in this article may not generalize directly to many other journals. Nonetheless, manuscripts faring well with respect to attributes related to publication in *Psychological Bulletin* would have many characteristics regarded as desirable for nearly any journal.

As was noted by Sternberg et al. (1997), reviewers clearly do have an influence on editorial decisions regarding the acceptance and rejection of manuscripts submitted to *Psychological Bulletin.* This should not be surprising. Part of the relation between reviewers' evaluations and editors' decisions surely is due to some similarity in the standards of editors and reviewers in evaluating manuscripts. In addition, there is little doubt that editors are influenced by the content of the reviews they receive. This is likely to be especially true at a journal such as *Psychological Bulletin,* for which acting editors often have to handle papers out of their areas of expertise. The correlation between reviewers' evaluations and editors' decisions was far from perfect; this doubtlessly is because reviewers often differed somewhat in their evaluations and provided additional detailed information in their written reviews and because editors' personal evaluations of manuscripts contributed to their final decisions. Regardless, reviewers' input was fairly highly correlated with the editors' decisions, and it is reassuring that the peer review process played a very substantial role in the editorial decisions of the acting editors.

In sessions at conferences or other venues on publishing in psychological journals, editors and other experienced authors often argue that a reject/revise decision from an editor is good news—that an author's chances of eventual acceptance go up considerably if they receive such a decision and revise in accordance with reviewers' recommendations. Our data strongly support this view. Among those first versions of manuscripts that received a reject/revise decision from the editor, 55% were eventually accepted after one or more subsequent revisions. The figure of 55% acceptance is an underestimate because some people who received a reject/revise decision never submitted a revision; for those who did resubmit, the rate of acceptance was 73.4%. As a slogan for the Arizona lottery (and others) has noted, you can't win if you don't play the game. Especially at a premier journal such as *Psychological Bulletin,* authors may be tempted to give up when they receive a reject/revise decision with numerous suggestions for changes; however, if the flaws in the manuscript are not fatal or largely irremediable, authors are well advised to revise and resubmit when given the opportunity to do so.

The clear majority of first authors on submitted manuscripts were men. This statistic is likely due to a predominance of males in senior academic positions. However, neither acceptance rates nor resubmission rates differed across men and women. This statistic is counter to the common notion that men resubmit more frequently than women, but the figure for resubmissions may vary depending on the quality and type of journal (e.g., journals oriented toward new empirical data or reviews).

Finally, the findings, when compared with those of Sternberg et al. (1997), suggest that the features that predicted acceptance of manuscripts at *Psychological Bulletin* were fairly similar from the Sternberg to the Eisenberg editorship. This is likely due to the similarity in their visions for the journal (Eisenberg, 1997). Our findings also indicate that many of the reviewers shared Nancy Eisenberg's vision and/or were influenced by it through the description of criteria for acceptance of manuscripts that they were sent. An understanding of this vision by potential authors (and any changes in it, as described by Harris Cooper, the new editor, in the January 2003 issue) should help authors to maximize their chances of publishing in this journal.

References

Eisenberg, N. (1997). Editorial. *Psychological Bulletin, 122,* 3–4.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272–299.

Muthén, L. K., & Muthén, B. O. (2001). Mplus 2.02 [Computer software]. Los Angeles: Authors.

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In R. D. H. Heijmans, D. S. G. Pollock, & A. Satorra (Eds.), *Innovations in multivariate statistical analysis: A festschrift for Heinz Neudecker* (pp. 233–247). London: Kluwer Academic.

Satorra, A., & Bentler, P. M. (1999, August 3). *A scaled difference chi-square test statistic for moment structure analysis* (UCLA Statistics Paper No. 260). Los Angeles: University of California, Los Angeles. Retrieved April 8, 2002, from http://www.stat.ucla.edu/papers/preprints/260/

Sternberg, R. J., Hojjat, M., Brigockas, M. G., & Grigorenko, E. L. (1997). Getting in: Criteria for acceptance of manuscripts in *Psychological Bulletin*, 1993–1996. *Psychological Bulletin, 121*, 321–323.

Received April 17, 2002
Revision received June 11, 2002
Accepted June 11, 2002 ■