# Editor's Page

Stanley S. Siegelman, MD

# Assassins and Zealots: Variations in Peer Review

## Special Report[1]

***Radiology*** **reviewers are required to assign numerical grades of 1 to 9 (1 = accept, 9 = reject) in the rating of manuscripts. The mean ratings for the 660 referees who were assigned 10 or more reviews over a 4½-year period were analyzed. The mean score was 4.8 ± 0.8, and 87.4% of reviewers (the mainstream) had ratings of mean ± 1.5 standard deviations. Categories of reviewers with greater deviation from the mean were identified: zealots and pushovers, whose ratings of manuscripts were more favorable, versus assassins and demoters, who supplied less favorable ratings. To exclude the possibility that the referees who were classified as more critical had actually been sent less meritorious papers, the scores and rejection rates of 859 papers co-reviewed by assassins, demoters, and mainstream referees were compared. Significant differences were confirmed. Deviant referees were widely distributed in the pool of reviewers, including 13 members of the Editorial Board and representatives in each of 19 subspecialty areas. Failure to recognize and control for reviewer variation may be unfair to authors. An Editor has the capacity to reduce unfairness by monitoring reviewer variation and by modulating the review process accordingly.**

**Index terms:** Radiology and radiologists, research • Radiology (journal)

**Radiology 1991;** 178:637-642

Forbear to judge, for we are sinners all.
[William Shakespeare]

*Radiology* uses the peer-review system. The Editor chooses appropriate individuals to pass judgment on the merit of manuscripts submitted for publication. The reviewer's task is to rate each work in terms of originality, documentation, and clinical impact (1) and to offer suggestions for improvement where possible. Reviewers may favor a manuscript or find flaws in the conception, execution, or interpretation of a project and, thus, conclude that the work is invalid. To some extent, such an evaluation is apt to be subjective, and there is a potential for major differences among reviewers as each applies a personal standard for judging a paper. Variations among reviewers in the perception of what constitutes an acceptable paper would make the peer-review system unfair to authors whose work happened to be sent by chance to an "assassin," a referee with stringent standards who advises rejection much more frequently than the norm. Conversely, uncritical acceptance of a manuscript would also constitute unfairness. Do such variations exist?

*Radiology* has the means to assess variations among its reviewers. Data from the review process are available in our computerized record systems and may be subjected to analysis. The goals of this communication are (*a*) to analyze reviewer ratings of manuscripts received over a 4½-year period, (*b*) to categorize reviewer variations, (*c*) to assess various factors as they relate to variations in reviewer ratings, and (*d*) to discuss the phenomenon of reviewer variation and the means by which it may be managed.

## MATERIALS AND METHODS

The Baltimore, Maryland, office of *Radiology* operates a computer-based manuscript tracking system utilizing a customized data base program. The program contains a record of every *Radiology* reviewer, including a reviewer rating for each manuscript. Referees prepare a report in which they rate manuscripts on a scale of 1 to 9 (2). Lower numbers indicate higher ratings; thus, 1 is outstanding, 2 represents excellent, 3 indicates good work, and 4 designates borderline acceptability. A rating of 5 is assigned when the reviewer is uncertain of the acceptability of a paper. Grades of 6 through 9 represent progressively decreasing merit for unacceptable manuscripts. Each manuscript is routinely sent to three reviewers.

1. A series of reports was generated from the computer system.

*a*) A tabulation was made of ratings of all manuscripts evaluated between November 1985 and May 21, 1990.

*b*) The mean ratings for referees who had been sent 10 or more manuscripts ($n$ = 660) during the period of investigation were computed. The standard deviation of the mean ratings was calculated. On the basis of the deviation from the mean score, reviewers were classified into five categories: zealots, pushovers, mainstream, demoters, and assassins. (Criteria for classification are given in Table 1.)

*c*) The mean ratings, standard deviation, and categories for those with 25 or more reviews ($n$ = 313) were obtained.

*d*) Similar data were compiled for the 107 members of the Editorial Board.

2. The distribution of reviews and the categories were analyzed as a function of the number of assigned reviews.

3. Referees are designated with one, two, or three subject headings based on the topic(s) of the material they review. Common associations for those with two subjects are pediatric/ultrasound, neuroradiology/magnetic resonance (MR) imaging, and gastrointestinal/body computed tomography (CT). Data on the re-

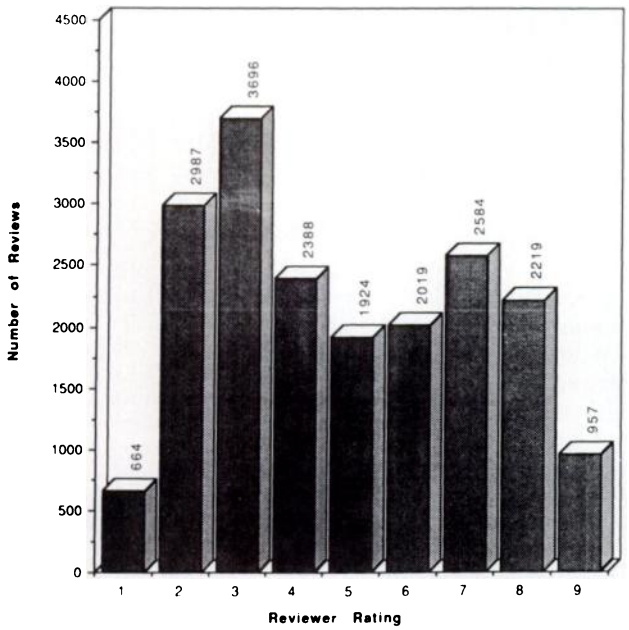
[1] From the *Radiology* Editorial Office, Johns Hopkins Medical Institutions, 550 N Broadway, Ste 206, Baltimore, MD 21205. **Address reprint requests to** the author.

[c] RSNA, 1991

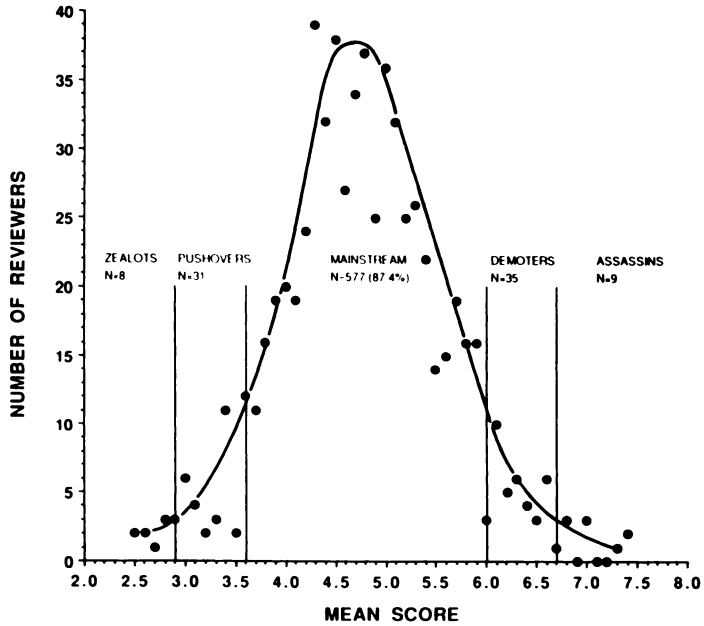
**Table 1**
**Criteria for Five Reviewer Categories**

| Category | Requirement | 10 or More Reviews ($n$ = 660) | | 25 or More Reviews ($n$ = 313) | |
|---|---|---|---|---|---|
| | | Mean Score | No. | Mean Score | No. |
| Zealot | 2.5 SD below mean | ≤2.8 | 8 (1.2) | ≤3.0 | 2 (0.6) |
| Pushover | 1.5 SD below mean | 2.9–3.5 | 31 (4.7) | 3.1–3.7 | 15 (5.8) |
| Mainstream | Mean ± 1.5 SD | 3.6–6.0 | 577 (87.4) | 3.8–5.8 | 268 (85.6) |
| Demoter | Mean ± 1.5 SD | 6.1–6.7 | 35 (5.3) | 5.9–6.5 | 22 (7.0) |
| Assassin | Mean ± 2.5 SD | ≥6.8 | 9 (1.4) | ≥6.6 | 6 (1.9) |

Note.—SD = standard deviation. Numbers in parentheses are percentages.




1.




2.

**Figures 1, 2.** **(1)** Reviewer ratings for manuscript reviews for *Radiology* from November 1985 to May 1990. **(2)** Mean scores for the 660 reviewers who were sent at least 10 manuscripts to review between November 1985 and May 1990. Mean score was 4.8; standard deviation, 0.8.

viewers who evaluated 10 or more manuscripts were analyzed, and, for each subject, a series of items was calculated: the number of reviewers, the number of reviews, the mean score of the reviewers, and the distribution of reviewers in the various categories.

4. All manuscripts that had been rated by reviewers categorized as assassins ($n$ = 12) or demoters (means scores of 6.1–6.7) ($n$ = 32) were further analyzed.

*a*) For papers co-reviewed by assassins and mainstream referees, mean scores and the fraction of manuscripts given a rating of 6, 7, 8, or 9 were tabulated. There were 229 papers reviewed by the 12 "assassins." The mean score for the 229 reviews was calculated. For the group of papers seen by each "assassin," the mean score, the mean of the mean scores, and the standard deviation of the means were calculated. Similarly, for the 387 reviews (of the same 229 papers) by mainstream referees, we calculated a mean score for all reviews, a mean score for each group associated with a single assassin, plus the mean of the mean scores and the standard deviation of the means.

*b*) For papers co-reviewed by assassins and zealots or pushovers, the mean scores (as in 4*a*) and the fraction of manuscripts given a rating of 6, 7, 8, or 9 were tabulated for each group.

*c, d, e*) As in 4*a* and 4*b*, values were tabulated for manuscripts co-reviewed by demoters and mainstream referees, for demoters versus zealots/pushovers, and for mainstream referees versus zealots/pushovers.

*f*) We compiled and compared the mean scores for mainstream reviewers in three categories: co-reviewers with assassins, demoters, and zealots/pushovers.

5. Several statistical tests were performed.

*a*) A comparison of mean reviewer scores (section 3 and section 4) were made by using analysis of variance techniques (ANOVA). Means based on scores of reviewers in different subject classes, such as body CT and breast, were compared by using the Newman-Keuls procedure for multiple comparisons with an adjustment to the sample size to accommodate reviewers belonging to more than one subject class (3).

*b*) $\chi^2$ contingency and trend analysis was used to evaluate the significance of the differences in percentages of acceptance of manuscripts by various categories of reviewers (section 4).

*c*) A two-tailed $t$ test was used to evaluate the difference between mean ratings of manuscripts co-reviewed by reviewers in different categories (section 4).

## RESULTS

1. *a*) During the 4½-year period of study, 19,438 reviews were rated and recorded (Fig 1). The four most common ratings were 3, 2, 7, and 8. The mean score was 4.8. Just over 50% of the ratings (9,735) were votes for acceptance. There were 21,373 papers sent for review; 202 were in the review process at the time of data analysis, and 21,171 had been returned. Of the papers returned, 1,733 (8.2%) had not been reviewed or had not been officially rated by the referee.

**Table 2**
**Data on Reviewers by Category**

| Category | No. of Reviewers | No. of Reviews | Average No. of Reviews/ Reviewer | Mean Score |
|---|---|---|---|---|
| Zealots and pushovers | 48 (7.3) | 1,103 (6.1) | 23.0 | 3.3 |
| Assassins and demoters | 56 (8.5) | 1,342 (7.5) | 24.0 | 6.3 |
| Mainstream | 556 (84.2) | 15,533 (86.4) | 28.0 | 4.8 |
| Total | 660 (100) | 17,978 (100) | 27.2 | 4.8 |

Note.—Numbers in parentheses indicate percentage of total. Mean score with zealots-pushovers excluded: 4.9. Mean score with assassins-demoters excluded: 4.7.

**Table 3**
**Reviewer Category versus Number of Reviews**

| No. of Reviews | No. of Reviewers | Mean Score | No. of Zealots | No. of Pushovers | No. of Demoters | No. of Assassins |
|---|---|---|---|---|---|---|
| 10-12 | 101 | 4.6 | 4 | 10 | 9 | 2 |
| 13-17 | 112 | 4.7 | 3 | 6 | 3 | 3 |
| 18-24 | 134 | 4.9 | 0 | 5 | 10 | 1 |
| 25-35 | 118 | 4.9 | 1 | 8 | 10 | 5 |
| 36-50 | 101 | 4.8 | 1 | 4 | 8 | 1 |
| 51-101 | 94 | 4.8 | 0 | 6 | 4 | 0 |
| Total | 660 | 4.8 | 9 | 39 | 44 | 12 |
| 61-100 | 54 | 4.8 | 0 | 3 | 0 | 0 |

*b*) There were 660 reviewers who had been sent 10 or more papers to review. Of the 19,465 dispatched manuscripts, 17,978 (92.4%) were rated and recorded in our system. The mean score was 4.8 with a standard deviation of 0.8. Display of the mean scores reveals a normal distribution (Fig 2). Objective variations in ratings were observed, and a population of zealots, pushovers, mainstream, demoters, and assassins was identified (Fig 2; Tables 1, 2).

*c*) Analysis of the mean scores of the 313 referees with 25 or more reviews also shows a normal distribution with groups outside of the mainstream (Table 1). (The 313 referees with 25 or more reviews were also included in the group that was analyzed in 1*b*.) The mean score was 4.8, and the standard deviation was 0.7. The lower standard deviation changes the ranges of scores in the various categories. Forty-five reviewers were classified in nonmainstream categories, including 21 who had been mainstream in 1*b* but now were categorized as pushovers (score 3.6, $n = 4$; score 3.7, $n = 5$) or demoters (score 5.9, $n = 9$; score 6.0, $n = 3$).

*d*) The 107 members of the Editorial Board consisted of one zealot, seven pushovers, 94 mainstream, five demoters, and no assassins.

2. The reviewer category was compared with the number of reviews (Table 3). The mean scores of the reviewers were not clearly related to the number of reviews. Zealots or pushovers were found in each of the six groupings and varied from 5% to 13.9% of the reviewers. Assassins and/or demoters were also found in each grouping and varied from 4.3% to 12.7% of the reviewers. The maximum number of papers assigned to an individual in each category was zealot, 46; pushover, 78; demoter, 60; and assassin, 37. The group with the highest proportion of mainstream reviewers (94.4%) was the 54 individuals who were assigned more than 60 manuscripts.

3. Data on the mean scores and the distribution of zealots, pushovers, demoters, and assassins among reviewers specializing in various subjects is presented in Table 4. The most critical groups of referees were those who judge submissions on computer applications (mean score, 5.3) and nuclear medicine (mean score, 5.2). These values were not significantly different from the scores of other groups ($P > .25$, ANOVA). Mean scores of reviewers in all other categories were clearly not significantly different from the overall mean score of 4.8.

4. *a*) *Manuscripts co-reviewed by assassins and mainstream referees.*—There were 229 such papers seen by the 12 reviewers classified as assassins. The mean score for each assassin was higher than the mean score for the corresponding group of mainstream referees. Mean scores for assassins ranged from 6.5 to 7.4. Mean scores for 11 groups of mainstream reviewers ranged from 4.3 to 5.5. There was one assassin with a mean score of 6.8 versus a group of mainstream reviewers with a score of 6.1. For the remaining 11 assassins, the differences in mean scores were 1.4 or greater (>1.5 standard deviation). The overall differences in ratings by the two groups were highly significant (Table 5).

*b*) *Manuscripts co-reviewed by assassins and zealots/pushovers.*—Nine of the 12 assassins reviewed at least one manuscript that was also reviewed by a "zealots/pushover." There were 27 such papers, with 29 reviews from zealots/pushovers, since two papers were seen by two zealots/pushovers. The differences between reviewers in these categories were highly significant (Table 5).

*c*) *Manuscripts co-reviewed by demoters and mainstream referees.*—There were 630 such manuscripts. The mean score for each demoter was higher than the mean score for the corresponding group of mainstream referees. Mean scores for demoters ranged from 5.8 to 6.8. Mean scores for the 32 corresponding groups of mainstream referees ranged from 4.2 to 6.1. In seven instances the difference in the mean score between the demoters and the mainstream reviewers was less than 0.8 (1 standard deviation). These differences were 0.1, 0.4, 0.5, 0.6, 0.7, and 0.7. Data for the group as a whole are presented in Table 5. Differences were significant. Table 6 provides data analysis for a single demoter.

*d*) *Manuscripts co-reviewed by demoters and zealots/pushovers.*—Ninety-seven papers were co-reviewed by demoters and zealots/pushovers. The differences in ratings and rejection rates were highly significant (Table 5).

*e*) *Manuscripts co-reviewed by zeal-*

*ots/pushovers and mainstream referees.*—The 859 papers studied included 107 in which there were co-reviews by mainstream reviewers and zealots/pushovers. Differences were significant (Table 5).

*f) Manuscripts co-reviewed by mainstream referees versus other categories.*—The mean score for mainstream reviewers in manuscripts co-reviewed by assassins ($n$ = 387) was 5.1, in manuscripts co-reviewed by demoters ($n$ = 1,037) was 5.2, and in manuscripts co-reviewed by zealots/pushovers ($n$ = 107) was 4.8. The differences were not statistically significant ($P > .25$, ANOVA).

## DISCUSSION

### Uniformity

Although we were seeking to document variations in reviewer ratings, the data can be interpreted as indicating an unusual degree of uniformity. A magic number, 4.8, was revealed. It was the mean score for all the reviewers in our system, as well as for the 660 most productive referees. When reviewers were analyzed by subject, the mean scores for 11 of the 19 classifications were between 4.7 and 4.9 (Table 4). Thus, 4.8 emerges as a standard, an indicator of anticipated reviewer performance when using the *Radiology* nine-point scale.

### Variation

An analysis revealed that there were subgroups of referees whose ratings were significantly more favorable (zealots and pushovers) or less favorable (assassins and demoters) than ratings by the mainstream referees. Before concluding that such referees use different standards, we had to consider the possibility that, by chance, these individuals were assigned manuscripts of greater or lesser merit than the norm. Data were available to examine this question, since each of our manuscripts is rated by multiple reviewers. By focusing on manuscripts reviewed by assassins and demoters, we documented that, compared with mainstream reviewers, the assassins and demoters assigned poorer ratings and more frequently voted for rejection. In the same group of papers, zealots and pushovers gave more favorable ratings and more frequently voted for acceptance.

If ratings were attributable to manuscript selection rather than reviewer standards, we should have

**Table 4**
**Reviewer Classification versus Reviewer Category**

| Subject | No. of Reviewers | No. of Reviews | No. of Reviews Per Reviewer | Mean | SD* | No. of Zealots | No. of Pushovers | No. of Mainstream | No. of Demoters | No. of Assassins |
|---|---|---|---|---|---|---|---|---|---|---|
| Body CT | 62 | 2,414 | 39 | 5.0 | 0.8 | 0 | 2 | 52 | 7 | 1 |
| Breast | 33 | 985 | 30 | 4.9 | 0.9 | 0 | 1 | 27 | 3 | 2 |
| Cardiovascular-interventional | 95 | 2,918 | 31 | 4.6 | 0.8 | 2 | 10 | 79 | 4 | 0 |
| Chest | 44 | 1,365 | 31 | 4.6 | 0.7 | 0 | 5 | 36 | 3 | 0 |
| Computer applications | 16 | 504 | 32 | 5.3 | 0.9 | 0 | 0 | 11 | 4 | 1 |
| Contrast media | 10 | 389 | 39 | 5.0 | 0.7 | 0 | 1 | 7 | 2 | 0 |
| Gastrointestinal | 43 | 1,562 | 36 | 4.8 | 0.7 | 0 | 2 | 38 | 3 | 0 |
| Genitourinary | 34 | 1,161 | 34 | 4.7 | 0.8 | 2 | 0 | 30 | 1 | 1 |
| Head and neck | 19 | 667 | 35 | 4.5 | 0.9 | 1 | 3 | 14 | 1 | 0 |
| Heart | 29 | 807 | 27 | 4.8 | 0.8 | 0 | 2 | 23 | 4 | 0 |
| MR imaging | 136 | 4,375 | 32 | 4.9 | 0.8 | 1 | 4 | 118 | 9 | 4 |
| Musculoskeletal | 53 | 1,699 | 32 | 4.9 | 0.8 | 0 | 3 | 44 | 4 | 2 |
| Neuroradiology | 65 | 1,832 | 28 | 4.8 | 0.9 | 3 | 2 | 54 | 4 | 2 |
| Nuclear medicine | 44 | 1,281 | 29 | 5.2 | 0.7 | 0 | 0 | 37 | 6 | 1 |
| Oncology-therapy | 22 | 513 | 23 | 4.8 | 0.8 | 0 | 0 | 21 | 1 | 0 |
| Pediatric | 47 | 1,307 | 28 | 4.5 | 0.9 | 2 | 8 | 33 | 3 | 1 |
| Physics | 60 | 1,590 | 27 | 4.9 | 0.9 | 0 | 2 | 52 | 3 | 3 |
| Statistics | 6 | 156 | 26 | 4.7 | 0.8 | 0 | 1 | 4 | 1 | 0 |
| Ultrasound | 86 | 2,521 | 30 | 4.8 | 0.8 | 0 | 5 | 72 | 6 | 1 |
| Total | 904<br>660 | 28,046<br>19,465 | 31†<br>31† | ... | ... | 11 | 51 | 752 | 69 | 19 |

Note.—Analysis of 660 reviewers with 10 or more reviews by subject classification. Reviewers are listed for one, two, or three subjects. The zealots, pushovers, demoters, and assassins total 150 because 66 have a single listing, 30 have two listings, and eight have three listings. The data, exaggerated by multiple listings, are actually based on 19,465 manuscripts sent to 660 reviewers.
* Standard deviation.
† Average.

**Table 5**
**Comparison of Mean Scores and Rejection Rates of Reviewers**

| Comparison | Mean Ratings | Mean of Mean Ratings | Rejection Rate |
|---|---|---|---|
| Assassins vs mainstream | | | |
| Assassins | 6.9 | 6.0 ± 0.3* | 179/229 (78.2)† |
| Mainstream | 5.1 | 5.0 ± 0.7* | 192/387 (49.6)† |
| Assassins vs zealots/pushovers | | | |
| Assassins | 7.1 | 6.6 ± 1.8* | 23/27 (85.2)† |
| Zealots/pushovers | 3.8 | 3.3 ± 1.2* | 5/29 (17.2)† |
| Demoters vs mainstream | | | |
| Demoters | 6.3 | 6.3 ± 0.25* | 487/630 (67.8)† |
| Mainstream | 5.2 | 5.2 ± 0.5* | 475/1,037 (45.8)† |
| Demoters vs zealots/pushovers | | | |
| Demoters | 6.3 | 6.5 ± 1.4* | 67/97 (69.1)† |
| Zealots/pushovers | 3.7 | 3.7 ± 1.4* | 22/104 (21.1)† |
| Zealots/pushovers vs mainstream | | | |
| Zealots/pushovers | 3.6 | 3.4 ± 1.4* | 19/107 (17.8)† |
| Mainstream | 4.8 | 4.9 ± 1.5* | 42/107 (39.3)† |

Note.—Data are based on an analysis of reviews of 859 manuscripts. Numbers in parentheses are percentages.
* $P < .001$, $t$ test.
† $P < .001$, $\chi^2$ analysis.

**Table 6**
**Actual Data from a Single Referee Classified as a Demoter**

| Paper | Demoter Score | Reviewer 2 Score | Reviewer 2 MS | Reviewer 3 Score | Reviewer 3 MS |
|---|---|---|---|---|---|
| 1 | 7 | 2 | 3.1 | 6 | 3.3 |
| 2 | 4 | 2 | 4.3 | 6 | 5.8 |
| 3 | 4 | 2 | 4.3 | 4 | 4.6 |
| 4 | 8 | 6 | 5.3 | 8 | 5.8 |
| 5 | 2 | 2 | 5.4 | 8 | 5.7 |
| 6 | 4 | 6 | 4.1 | ... | ... |
| 7 | 8 | 3 | 4.8 | 6 | 6.3 |
| 8 | 5 | 3 | 3.0 | 8 | 5.3 |
| 9 | 3 | 1 | 4.2 | 3 | 4.5 |
| 10 | 9 | 6 | 4.6 | 7 | 5.8 |
| 11 | 8 | 7 | 6.3 | 7 | 4.3 |
| 12 | 9 | 5 | 4.2 | 7 | 5.3 |
| 13 | 9 | 2 | 4.6 | 3 | 5.9 |
| 14 | 4 | 6 | 4.3 | ... | ... |
| 15 | 5 | 2 | 4.3 | 3 | 5.5 |
| 16 | 7 | 2 | 4.0 | 3 | 4.1 |
| 17 | 8 | 5 | 4.6 | 5 | 4.6 |

Note.—MS = mean score for all manuscripts reviewed. The demoter rated 18 papers; 16 were co-reviewed by a second and third reviewer and two papers (6 and 14) were seen by a single other reviewer. Seventeen papers were co-reviewed by mainstream reviewers (mean score for demoter, 105/17 = 6.2; mean score for mainstream, 129/29 = 4.4). Two papers (1 and 8) were co-reviewed by pushovers. Two papers (7 and 11) were co-reviewed by demoters. Note that with respect to reviewers 2 and 3, on no occasion was a lower score (higher rating) given by a referee with a higher mean score.

found smaller differences between the ratings of the assassins-demoters and the mainstream referees who co-reviewed papers. Chance selection of poorer-quality manuscripts appears to have been a factor for one of the 12 assassins and for seven of the 32 demoters in that their mean scores were close to those of mainstream co-reviewers. For the groups as a whole, however, manuscript selection was not a major factor.

In selecting criteria for categorizing reviewers, we chose arbitrary standards. Mean ± 1.5 standard deviations seemed like a reasonable range for defining the mainstream. In our population of reviewers, the standard deviations of the mean score decrease as the number of reviews increases. We illustrated the shift by separate analyses based on reviewers with 10 or more reviews (mean, 4.8; standard deviation, 0.8) and 25 or more reviews (mean, 4.8; standard deviation, 0.7). If we consider reviewers with 50 or more reviews (mean, 4.8; standard deviation, 0.6), those with scores of 3.8 and 5.8 would also be excluded from the mainstream.

## Identification

Who are the zealots and assassins? There were no objective factors, other than the ratings of manuscripts assigned, that could be used to identify divergent reviewers. Outliers were included in every reviewer subject classification (Table 4). Reviewers of papers in nuclear medicine and computer applications had the highest mean scores. It is notable that these groups had no zealots or pushovers. Reviewers in various nonmainstream categories differ not in the quality of their reviews but rather in their perceptions of what constitutes an acceptable manuscript. Outliers are apt to produce detailed, complete reviews with carefully documented observations on originality, validity, and pertinence of the work. Thus, pushovers and demoters are included among our most experienced and respected reviewers, and 13 of the 107 members of the Editorial Board are outside of the mainstream.

Our choice of terms for reviewer categories is whimsical, yet "zealot" and "assassin" are somehow appropriate because of the implication of an emotional component to the review process. We have received more than a dozen letters from reviewers in these categories protesting the final editorial decision to accept (by assassins) or to reject (by zealots) contrary to their advice.

How do divergent reviewers justify their stance? Two trends are apparent. Assassins and demoters claim that they would like to raise standards for papers accepted for publication. They are less likely to conclude that material is innovative. They want larger numbers of cases studied, more convincing documentation for

negative studies, and longer follow-up periods following interventional procedures. They contend that too many marginally acceptable papers are published. Zealots and pushovers would like to have more papers accepted from their areas of expertise. They indicate that impressive progress in their subspecialties justifies efforts to publish as many quality articles as possible. They often provide excellent, lengthy critiques including helpful suggestions for revision. They frequently act as advocates by emphasizing the positive features of submitted work.

## Fairness

Who is right? The zealots may be correct in indicating that a high proportion of submitted work is worthy of publication. The assassins may be correct in contending that more stringent standards should be applied and that fewer papers should be published. The key point, no matter which group is correct, is that the existence of disparate categories of reviewers creates the potential for unfair treatment of authors. Those whose papers are sent by chance to assassins/demoters are at an unfair disadvantage, while zealots/pushovers give authors an unfair advantage.

## Management

Since zealots and assassins are a source of potential difficulty, why not eliminate them entirely from the review process? Those whose reviews are not outstanding are dropped from the reviewer corps. The remainder are retained because they are the most knowledgeable experts in their fields, and, as previously mentioned, they provide superb, detailed reviews. Votes to reject by zealots and to accept by assassins are particularly decisive.

Editors should be aware of reviewer variation. Editors of journals with a small corps of referees undoubtedly will recognize their assassins and zealots and will manage to deal with the disparities. For large journals with numerous reviewers, there is a danger that authors will be treated unfairly if no effort is made to record and to recognize differences in reviewer standards.

*Radiology* continuously monitors review ratings. The computer is regularly utilized to provide a printout of reviewers; this list contains the mean rating of the manuscript previously reviewed. There is considerable variation (mean score, 3.6–6.0) among reviewers categorized as mainstream. Just as several referees may have been miscast as demoters because, by chance, they were assigned manuscripts of lesser quality, there may be unrecognized demoters with mean scores of 5.7 or 5.8 who reviewed papers with greater than average merit. In selecting three reviewers for a paper, an attempt is made to choose a balanced combination. This is usually achieved. The choice of three referees with mean scores of 5.6 or greater, for example, is potentially unfair. On rare occasions, however, limited availability of reviewers in a specific category necessitates the use of an inappropriately critical group of referees. Armed with knowledge of reviewer status, an Editor can make a reasonable, informed decision about the acceptability of such a paper.

Thus, as a matter of policy, we seek a balanced group of referees (in terms of mean scores) for *any* paper submitted to peer review. The Editor must not abuse the system. What if the Editor, on preliminary examination of a manuscript, is displeased by impracticability, poor preparation, or a lack of originality? In such cases, we reject the paper without external review, avoiding the temptation to seek a group of assassins for official castigation. Likewise, we avoid the duplicitous practice of assigning papers that please the Editor to zealots and pushovers.

The Editor can modulate reviewers' recommendations in the decision to accept or to reject the paper. Thus, when a single reviewer opts for rejection, the paper is much more likely to be accepted if the rejection comes from an assassin as opposed to a mainstream reviewer. Similarly, papers are seldom accepted on the basis of affirmations from zealots when mainstream co-reviewers do not give the work an appropriate rating.

Unfortunately, authors may be abused by the subjective nature of the peer-review process. Measures to make peer review more objective are highly desirable. *Radiology* employs a computer-based system in which reviewer ratings and rejection rates are continuously monitored. The existence of assassins and zealots, reviewers whose standards are at variance with those of mainstream referees, has been documented. Knowledge of reviewer status yields fairer treatment of authors. ■

### References


1. Siegelman SS. *Radiology* statement of purpose 1990. Radiology 1989; 173:467–468.
2. Siegelman SS. Guidelines for reviewers of *Radiology*. Radiology 1988; 166:360.
3. Winer BJ. Statistical principles in experimental design. New York: McGraw-Hill, 1971; 216–218.