The Paper Selection Scheme Simulation Analysis

Zheng Yuan Zhu
Jian Liu
Haonan Tan
Fudan University
Shanghai, China

Advisor: Yongji Tan

Summary

We provide five models for selection schemes and, based on computer simulation, we propose optimal schemes for each model.

In analyzing the problem, we use a cost function to evaluate a scheme.

We quantify a judge's capability in terms of the variance d of accidental error and the magnitude e of systematic bias.

We enumerate our assumptions and give the algorithm for computer simulation. We then discuss the possible ranges of parameters d and e, finding that the judges' capability must reach a certain level to accomplish their task.

We discuss the five models. The Ideal Model can be used to explain the selection scheme under ideal conditions. The Round-Table Model and the Classic Model produce expensive solutions. To save money, we put forward the Cutoff Model and the Advanced Round-Table Model.

The Cutoff Model is based on numerical scoring and rejects papers under a certain cutoff level in each round. Its flexibility on changing the rejection proportion in each round, depending on the capability of the judge, leads to a more economical scheme. The Advanced Round-Table Model is a combination of rank-ordering and numerical scoring; it can generate a scheme that is both economical and easy to operate.

We compared all the models (see **Table 6**). The schemes produced by the Cutoff Model and the Advanced Round-Table Model reduce the expense drastically.

Later we generalize and find that all models except the Round-Table Model are suitable for different values of P, J, and W. For the purpose of classifying winners to various ranks, the Cutoff Model works best.

We find that the expense depends on the capabilities of the judges. A small decrease in the capability of the judges can lead to a great increase in expense in each model. So the best advice that we can give to the contest committee is

Do not hesitate to choose the best judges!

Assumptions

- There is an absolute rank-ordering and numerical scoring to which all judges would agree.
- The absolute numerical scores of all the papers are integers from 1 to 100, and are distributed $\mathrm{N}(70, 100)$ ($\mu = 70$, $\sigma^2 = 100$).
- A scheme is accepted if and only if it guarantees with 95% probability that the final W winners are among the "best" $2W$ papers.
- The judges score papers individually and do not influence one other.
- The judges' scoring has accidental errors that have the normal distribution; the magnitude of errors can be obtained from a judge's past records.
- Some judges have systematic bias for a specific kind of paper, hence they will mark higher or lower scores on such papers.
- In using the rank-ordering method, only the bottom 30% that each judge rank-orders are rejected.

Analysis of the Problem

Our main task is to provide a scheme that can reliably select the W winners and significantly reduce the number of papers for each judge to read.

The Evaluation Method

We can adopt either rank-ordering or numerical scoring to select the best papers. We elaborate the case of scoring, since a rank-ordering can be produced on the basis of the scores.

Expense

The purpose of reducing the number of papers read by each judge is to save on the expense of the contest. According to the theory of marginal utility, more papers to read means more money to be paid for each paper. So the reading number for different judges should be as equal as possible. The cost function depends on the actual situation, but we use the following function: Papers 1 to 20 cost \$$m$ each, 21 to 50 cost \$$2m$ each, and 51 to 100 cost \$$4m$ each. In mathematical form, the function is

$$C = m \cdot \sum_{i=1}^{J} \{a_i + (a_i - 20) \cdot u(a_i - 20) + 2 \cdot (a_i - 50) \cdot u(a_i - 50)\},$$

where a_i is the number of papers that judge i reads and

$$u(x-a) = \begin{cases} 0, & x < a; \\ 1, & x \geq a. \end{cases}$$

Small variation in the cost function has little effect on the scheme; with so little information, we might as well let $m = 10$.

The Judge

After preliminary simulation, we found that the capability of the judge is the most important factor in deciding the selection scheme. We use two parameters to describe the capability of a judge:

- The variance of accidental errors in scoring. The smaller the variance, the more abundant the judge's experience and the more accurate the judge's scores, and vice versa. The variance can be obtained from the judges' past performance.

- The magnitude of the systematic bias. This is hard to quantify, since it is difficult to determine the type and the magnitude of an individual's bias in real life. So we simplify the situation and classify both the judges and papers into three types: radical, neutral, and conservative. A radical judge gives higher scores to radical papers and lower scores to conservative papers and has no bias towards neutral papers. A conservative judge stands opposite to the radical one. A neutral judge has no bias at all.

Constructing the Model

Because of too many random factors in the judgment process, it is difficult to solve the problem theoretically. To solve the problem, we adopt computer simulation based on theoretical analysis.

The Algorithm of the Simulation

1. Generate 100 random integers from 1 to 100, as the papers' "real" scores, from the distribution $\mathrm{N}(70, 100)$. Put them in the array `paper_score[1,i]`.

2. Take a constant d as the upper bound of all the judges' accidental errors. Generate 8 random integers d_j as the standard deviations of the judges' accidental errors, using a discrete uniform distribution on the integers from 0 to d. Put these numbers in the array `judge[1,j]`.

3. Take a constant $e > 0$ as the systematic bias value. Let 1, 0, and -1 represent radical, neutral, and conservative, respectively. Give every paper a

number in $\{1, 0, -1\}$ and put them in array `paper_score[0,i]`. Give every judge a number in $\{1, 0, -1\}$ and put them in `judge[0,j]`. We calculate the systematic bias s from the expression

$$s = e * \texttt{paper\_score}[0, \texttt{i}] * \texttt{judge}[0, \texttt{j}].$$

For example, when a conservative judge meets a radical paper, $s = -e$.

4. The method that judge j scores paper i: Let

$$u = \texttt{paper\_score}[1, \texttt{i}] + e * \texttt{paper\_score}[0, \texttt{i}] * \texttt{judge}[0, \texttt{j}].$$

Generate random integers in [1,100] as scores that have the normal distribution $\mathrm{N}(u, d_j^2)$. Put them in array `judge_score[i,j]`. Thus we generate the judges' score matrix.

Determining the Parameters

We need to determine the values of d and e. First, we discuss how to determine the range of d.

From probability theory, we have

Lemma. *Let X_1, X_2,..., X_n be independent random variables with variances σ_j^2. Then $\overline{X} = \sum X_i/n$ has variance*

$$\sigma^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2.$$

So we have

Corollary 1.
$$\frac{1}{\sqrt{n}}\min_{1\le i\le n}\{\sigma_i\} \le \sigma \le \frac{1}{\sqrt{n}}\max_{1\le i\le n}\{\sigma_i\}.$$

From the corollary, we conclude that the accuracy of judgment can be improved if several judges work on each paper and average their scores.

Using the Cauchy inequality, we have

$$\sigma^2 \ge \frac{1}{n^3}\left(\sum_{i=1}^{n}\sigma_i\right)^2,$$

that is,

Corollary 2.
$$\sigma \ge \frac{1}{\sqrt{n}}\frac{\sum_{i=1}^{n}\sigma_i}{n}.$$

Since by our assumption σ is distributed discrete uniform on $[0, d]$, we have

$$\frac{\sum_{i=1}^{n} \sigma_i}{n} \approx \frac{d}{2}.$$

For $n \leq 8$, this becomes

$$\sigma \geq \frac{1}{2\sqrt{2}} \cdot \frac{\sum_{i=1}^{n} \sigma_i}{n} \approx \frac{\sqrt{2}}{8} d.$$

Now we have

Conclusion 1. *Generally speaking, accidental errors can be reduced when several judges work on the same paper. The more judges involved, the more accurate the result.*

Conclusion 2. *For the most part, in scoring a single paper, the standard deviation of the mean accidental error will not be lower than* $\frac{\sqrt{2}}{8} d$.

We have not proved the two conclusions; in fact, there are exceptions. But the probability of exception is too slim to have effect on the practical problem. So we grant these conclusions in our later discussion.

Based on lots of experiments with computer simulation, we find the experimental law listed below.

Law 1. $$d < 10,$$

where d is the upper bound on the standard deviation of the accidental judging error.

Verification: We need only explain that when $d = 10$, there is no selection scheme that guarantees with 95% probability that the final W winners are among the "best" $2W$ papers. We consider the ideal situation, under which each judge reads all the papers, for the case of 6 papers and 8 judges.

Because of **Conclusion 2**, the standard deviation of the 8 judges' mean accidental error is $\sigma \geq d\sqrt{2}/8$. We may suppose $\sigma = d\sqrt{2}/8$ as well, and here $d = 10$. We set the systematic bias to zero: $e = 0$. Using a simulation of the Round-Table Model and 10,000 iterations, 10,000 times, we observed 9,460 times when the judges chose the 3 winners correctly (i.e., they are among the "best" 6 papers). This probability of 94.6% is a little lower than our standard.

A calculation using Mathematica gave probability of failure (at least one winner not among the "best" 6 papers) as 5.6%. □

By using the simulation program for Round-Table again, with a changed value of the parameter d, we get:

Law 2. *When $d \leq 3$, Round-Table meets our 95% standard if each paper is read by only one judge.*

Comments on Laws 1 and 2

Law 1 points out that to succed in the task of selecting winners in a contest, the judges' capability should reach a certain level. If a judge's standard deviation d is more than 10, even if there is no systematic bias, a paper that deserves the score of 70 may be scored higher than 80 or lower than 60 with probability greater than 30%. The probability of a score higher than 90 or lower than 50 is no less than 5%. Such a person obviously cannot be qualified as a judge in any serious competition.

Law 2 points out that if all the judges are reliable enough—in other words, they are all experienced and have little systematic bias—a single judge's score is sufficient for determining the winners. When $d = 3$, running of Round-Table 5,000 times shows that the average failure rate is about 1.2%.

We conclude that if $e = 0$, we need to consider only d from 3 to 10; if $e \neq 0$, then d ranges from 0 to 10.

The range of e is quite difficult to determine. A reasonable supposition is that e has the same magnitude as d. We ran Round-Table with different values of d and e and used Mathematica to plot a three-dimensional bar chart (see **Figure 1**).

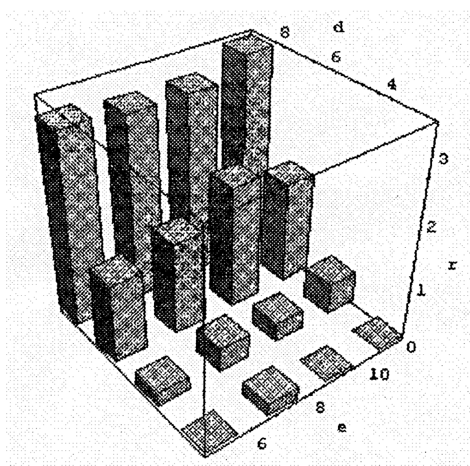

Figure 1. The failure rate $(r/100)$ is much more dependent on accidental error d than on bias e.

From the figure, we find that the standard deviation d has significant effect on the result, whereas the bias e has little effect.

Later we put forward several practical models for paper-judging and find some optimal schemes for different values of d and e, based on computer simulation.

We may suppose that $e \in \{0, 5, 10\}$ and $d \in \{1, 3, 5, 7, 9\}$. These limited ranges suffice to reveal the relationship between the scheme selection and the capability of judges.

The Ideal Model

When $d = e = 0$, every judge can rank-order or score all the papers as correctly as the absolute rank-ordering. This is the "ideal situation."

For 100 papers and 8 judges, 4 judges have 13 papers each and 4 have 12 each; a single reading of a paper, by any judge, suffices. The winners are the papers with the highest scores. The total cost is $1,000, and the 3 winners must be the "best" 3 papers.

A good scheme to rank-order the papers is shown in **Figure 2**. This scheme guarantees that the winners are the "best" 3 papers. The cost is $1,210.

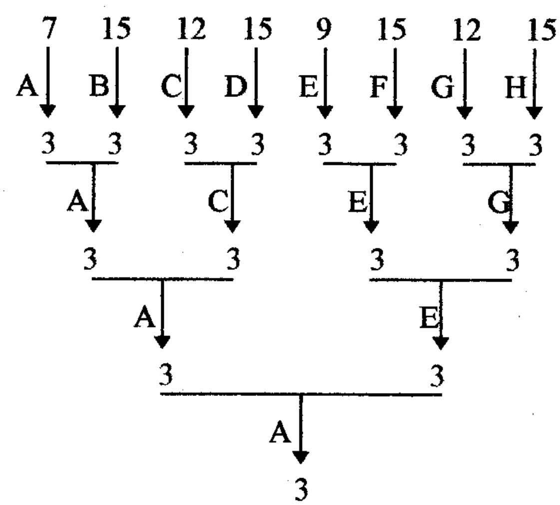

Figure 2. A scheme for rank-ordering the papers. The letters represent the 8 judges. The number at the tail of each arrow is the number of papers that a judge reads, and the number at the head of the arrow is the number that the judge selects.

But the most economical method is presented in **Figure 3**. Here, judges A, B, and C each rank-order 14 papers, and the other judges rank-order 13 papers each. We suppose that A is the Head Judge, who is responsible for picking out 3 winners from the 8 papers left after the first screening round. The cost of this method is $1,070. Though it cannot ensure that winners are the top 3, it ensures that the "best 3" papers are among the 6 winners with probability of 99.3%.

We prove this assertion. If the "best" 6 papers distribute among at least 3 groups (i.e., are judged by at least 3 different judges at the initial screening), then the 3 winners are "qualified" (i.e., among the "best" 6). If the top 6 distribute among no more than 2 groups, there must be some unqualified winners. The number of the unfavorable events is

$$8 + \binom{6}{1} \cdot 8 \cdot 7 + \binom{6}{2} \cdot 8 \cdot 7 + \binom{6}{3} \cdot 8 \cdot 7 \cdot \frac{1}{2} = 1{,}744.$$

The number of ways to distribute 6 papers to groups arbitrarily is $8^6 =$ 262,144, and $1{,}744/262{,}144 = 0.66\%$. In other words, all the winners are qualified in 99.3% of all the cases.

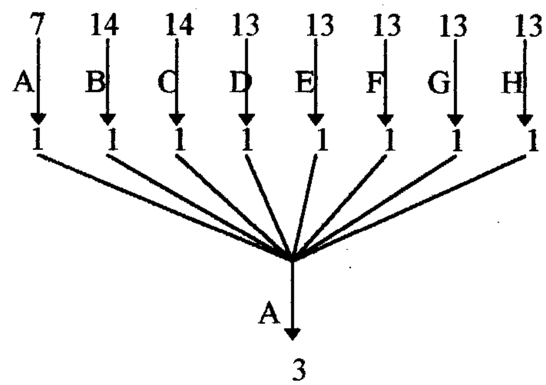

Figure 3. A more economical scheme for rank-ordering the papers.

The Ideal Model sets a lower bound for the cost function. When d and e are small (e.g., d=3, e=0), which means the judges are experienced and have almost no bias, the ideal model can work as the selection scheme.

The Round-Table Model

The distinguishing feature of this model is its simplicity.

1. Determine n, the number of rounds, according to the specific d and e.
2. Have all judges sit around a round table (hence the name of the model) and share the papers equally. At each round, after every judge has finished scoring, each judge passes the papers already read to the the right, where the neighboring judge scores these papers again in the next round.
3. After n rounds, each judge averages the n scores on a paper; the average score is the final score for the paper. The final score determines the rank-order of the paper.

The key question for this method is how to determine the value of n. Through numerical experiment, we found that systematic bias exerts only slight influence on n. So all our discussion later is based on the assumption that n is thoroughly determined by d.

When the distributions of the judges' accidental errors are all $\mathrm{N}(0, d^2)$, it is not hard to find that after n rounds the error of final score will be $d_n = d/\sqrt{n}$. For 8 judges, and with $d < 10$, we have $d_n < \sqrt{2}/8 \cdot 10 \approx 1.77$. Further simulation shows that the scheme is desirable if $d_n \leq 1.6$. We have

Law 3. *If all judges' error are distributed $N(0, d^2)$, and if $d_n = d/\sqrt{n} \leq 1.6$, i.e., $n \geq d^2/1.6^2$, then an n-round scheme is desirable; if $d_n \geq 1.77$, i.e., $n \leq d^2/1.77^2$, then an n-round scheme is not desirable.*

When all the conditions of **Law 3** are satisfied, we can easily find the optimal value of n. But the requirement on the error distribution is too harsh. When the variance of error has the uniform distribution on $[0, d]$, there is an empirical formula:

$$n = \min_{K \in N} \left\{ K \geq \left(\frac{d}{2 \cdot 1.6} \right)^2 \right\}.$$

The values of n obtained from the formula agree with the optimal n obtained from computer simulation (see **Table 1**).

Table 1.

Failure rate in 1,000 iterations, and expense, for various numbers of rounds n.

Bias e	Max variance d	Number of rounds n	% Failure rate	Expense
0	3	1	1.2	\$1,000
	5	2	4.4	\$2,400
	7	5	3.2	\$10,400
	9	8	2.8	\$22,400
5	3	1	3.6	\$1,000
	5	2	4.7	\$2,400
	7	5	4.8	\$10,400
	9	8	4.4	\$22,400

We see from the table that the expense increases rapidly as d rises. So we had better not use this scheme if the capability of the judges is quite ordinary.

The Classical Model

We put forward a classical model by combining rank-ordering and scoring methods.

1. Distribute the papers to each judge equally as far as possible. If a judge meets a paper already scored by that judge, the judge exchanges it with another judge. The judges score the papers.

2. Every judge rank-orders the papers that that judge has just scored and determines the bottom 30%.

3. Each judge's bottom 30% of papers are rejected.

4. If only 3 papers remain, they are winners. If there have been 8 rounds, then all the papers left have been scored by the 8 judges. So average each paper's scores and select the highest 3 as the winners. Otherwise go back to Step 1.

This model strictly limits the rejection of the papers during each round and refrains as much as possible from rejecting the good papers. The stability and precision of the model are very high, but the flexibility is low. Because of the progressive rejection method, it costs less than the Round-Table Model when d is comparatively large; but in general, this scheme is comparatively expensive. See the result of the simulation of the Classical Model in **Table 2**.

Table 2.

Simulation results for the Classical Model.

Bias e	Max variance d	% Failure rate	Expense
0	0	0.0	\$4,851
0	5	2.0	\$5,022
0	9	4.1	\$5,563
5	0	1.1	\$5,462
5	5	1.9	\$5,779
5	9	4.8	\$6,250
10	0	5.5	\$6,528
10	5	6.3	\$6,653
10	9	10.7	\$7,395

The Round-Table Model and the Classical Model can be used as selection schemes. However, the expenses are not encouraging. So we put forward two models based on them that will be more economical.

The Cutoff Model

This model is based on the Classical Model. However, it changes the cutoff levels in each round. Thus, it is not constrained to reject 30% in each round but can determine the rejection proportion by the circumstances. So it is more flexible.

1. Determine the failing percentage of each round. When there are n rounds, the falling percentage is $x = \sqrt[n]{0.03}$. Each paper is scored in each round, so we can complete the judging work in no more than 8 rounds. Hence $n = 8$.

2. Distribute the papers to each judge equally; each judge should get papers that that judge has not previously scored.

3. The judges score the papers. Average each paper's scores as its this-round-score. Determine this round's cutoff level using the failing percentage and the papers' this-round-score. Any paper below the cutoff level falls.

4. If only three papers remain, they are the winners. Otherwise, go back to Step 2.

For fixed variance d and systematic bias e, we experimented with different values of n to find the optimal scheme, that is, the scheme with the smallest number of rounds that satisfies the criterion of an failure rate less than 5%. **Table 3** gives some results.

Table 3.

Simulation results for the Cutoff Model, giving percentage failure rates for combinations of e and d with values of n.

Bias e	Max variance d	Number of rounds n								Expense
		1	2	3	4	5	6	7	8	
0	0	0								
	3	0								
	5	9	6	1						\$1,189
	7	25	17	8	6	7	4			\$1,661
	9	47	28	16	15	6	6	7	5	
5	0	16	10	7	3					
	3	11	5	4	6	1				
	5	29	7	3	7	2				\$1,725
	7	65	45	18	15	8	6	8	7	
	9	62	33	23	18	18	11	10	3	

Figure 4 shows the rank of the third-best paper for two of the optimal schemes, and **Figure 5** shows that the simulation of these schemes is stable.

We see that the cost is significantly less than for the two previous models; the total reading times are reduced because many of the unqualified papers are rejected earlier. However, the method of distributing papers in each round is comparatively complicated and may cause trouble in practice.

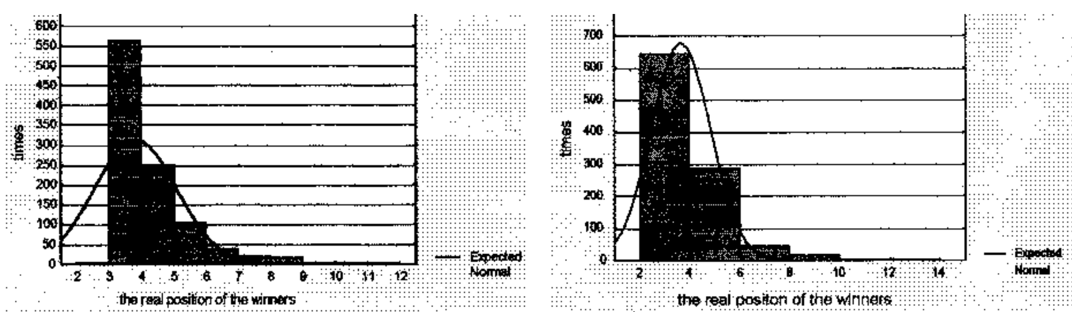

Figure 4. Rank of the third-best paper for two optimal schemes, over 1,000 iterations.

a. $e = 0$, $\max d = 5$, and 2 rounds. **b.** $e = 5$, $\max d = 5$, and 4 rounds.

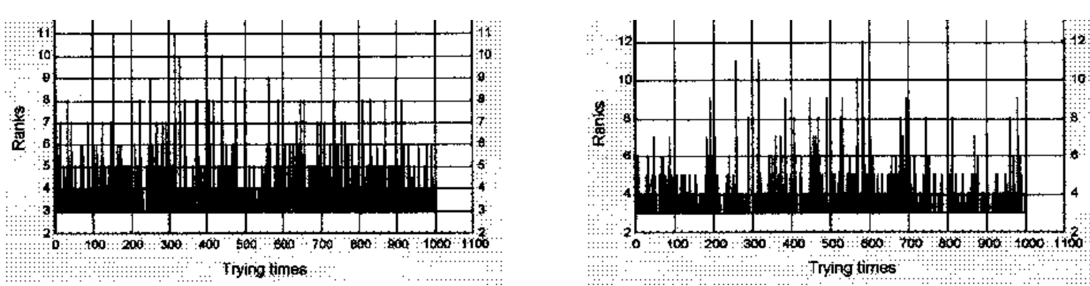

Figure 5. Rank of the third-best paper vs. iteration number, for two optimal schemes.

a. $e = 0$, $\max d = 5$, and 2 rounds. **b.** $e = 5$, $\max d = 5$, and 4 rounds.

The Advanced Round-Table Model

This model use a combination of rank-ordering and numerical scoring. In its early stage, we ingeniously use the method of group sequencing and partially exchanging to realize rejecting in definite proportion without violating the assumption that only the bottom 30% that each judge rank-orders are rejected. This selection scheme is stable in structure and easy to implement. Setting the times of paper-exchanging each round also makes it flexible. In the final stage, we determine the winners from average scores to speed the selection process.

1. We distribute the papers equally to the judges. The rejection proportion in each round is 30%. After n rounds of screening, each judge has only one paper left (when 30% is less than 1, we consider it to be 1). At the first round, we control the roundoff method to ensure that each judge has an equal number of papers left. At other rounds, we round down.) Given the capability of judges, we can determine the exchange times K_i in each round.

   For our problem with n=6, the numbers of papers each judge has after each round are 9, 6, 4, 3, 2, and 1. The numbers of papers rejected in each round are 4, 3, 2, 1, 1, and 1.

2. The judges sit around a round table. Let $K_i = i$ (later we thoroughly discuss the method of selecting the value of K_i). At the first round, $K_0 = 0$, judges do not exchange papers but rank-order them and eliminate the worst 30%.

3. At the second round, $K_1 = 1$. Each judge passes the worst 30% papers to the right. Then each judge scores the new papers received and re-rank-orders *all* current papers (including the ones not passed on) and cuts off the worst 30%.

4. For $K_i \geq 2$, the passing, scoring, re-rank-ordering, and rejecting the worst 30% takes place i times.

5. When each judge has only one paper, the paper is passed K_n times. If the paper has already been scored for K_n times, the other judges do not score it. From the mean value from the K_n scorings, we select the best three papers as winners.

The exchange method in this scheme is somewhat like that of the Round-Table Model, so we call it Advanced Round-Table.

Why Exchange the Bottom 30%?

We assumed that only the bottom 30% papers that each judge rank-orders could be rejected. Hence the number of papers left after each round will not be a fixed number, which makes the method more complicated, more unstable, and increases costs. But this undesirable feature can be avoided if we circulate the bottom 30% papers. For example, consider a paper among the bottom 30% that J_1 has passed to J_2. If it is still among the bottom 30% of J_2's papers after J_2's re-rank-ordering, it is certain that it should be eliminated. If this bottom 30% also contains papers that are not from J_1, the paper in question can be eliminated without violating the 30% rejection rule (it has been rank-ordered only by J_2). Moreover, if J_2 believes that it is even worse than some of the bottom 30% from J_1, it is reasonable for J_2 to reject it.

How Many Times to Pass the Papers?

Since at each round we must determine how many times to pass the papers, the process of searching for the optimal scheme is much more complicated than the Round-Table and Cutoff Models. However, the flexibility of the model increases with the complexity, which makes it possible to find a selection scheme that is both efficient and economical.

First, we note two properties of $\{K_i\}$:

1. $\{K_i\}$ is bounded, i.e., $0 \le K_i < J$.

2. $\{K_i\}$ is monotone increasing, i.e., $i < j \Rightarrow K_i \le K_j$.

There are only J judges, and all the papers are divided into J groups. When $K_i > J$, a judge may find that a paper previously passed on comes back again! That is anything but efficient. So $K_i < J$.

At the last several rounds, it seems more likely that the qualified papers (the top 6 papers) may be eliminated, so there should be more passings at later rounds than at early ones. So $\{K_i\}$ is monotone increasing.

Second, there is a relationship between $\{K_i\}$ and the cost C. That is, C is a monotone increasing with the total number K of papers that all judges have rank-ordered, $K = 8\sum_{i=0}^{n} P_i K_i + 100$, where P_i is number of papers to be eliminated at round i. In this model, the numbers of papers read by each judge are almost equal (the difference is no more than one paper), so C is monotone increasing with K.

It's clear that the expense and the precision of selection scheme are at odds: The less the expense, the bigger the probability of producing an unqualified winner. We might begin as well with the $\{K_i\}$ whose corresponding C is the smallest. By testing the $\{K_i\}$ according to the order of the sequence one by one, we increase the expense step by step and at the same time increase the precision of the scheme. When the precision meets the request, the scheme that this $\{K_i\}$ corresponds to is the optimal scheme. We can use our simulation program to test the precision of the scheme.

We should say that this method of searching for the optimal scheme is time-consuming. To find an optimal scheme corresponding to a group of specific values of d and e, we may spend hours; but compared to the savings, several hours of machine time is nothing.

A more efficient approach is to use binary search to find the "best" $\{K_i\}$.

Table 4.
Failure rate and expense for various sets of $\{K_i\}$.

Bias e	Max variance d	K_n		Iterations	% Failure rate	Expense
0	5	1, 1, 1, 1, 1, 2, 4	*	20,000	1.8	$1,120
	7	1, 1, 1, 1, 1, 4, 5	*	10,000	3.9	$1,560
	9	1, 1, 1, 2, 2, 4, 5		5,000	4.8	$1,960
5	5	1, 1, 1, 2, 2, 2, 4		1,000	0.7	$1,480
	7	2, 2, 2, 2, 2, 4, 8		1,000	2.7	$3,760
	9	2, 2, 2, 2, 2, 4, 8		1,000	6.7	$3,760

Listed in **Table 4** are several $\{K_i\}$ (the two starred sets are optimal).

From the table we notice that under the same conditions this model's cost is less than that of all the other models, while its operating process is clear and definite, easy to understand, and easy to apply in practice.

The weakness is also obvious: finding the optimal $\{K_i\}$, which is time-consuming.

Comparison and Critique of the Models

We have discussed five models, all suited for practical use except for the Ideal Model, which can be used under ideal conditions only. **Table 5** compares the precision and expense of the different models under specific conditions.

Table 5.
Precision and expense of the different models.

Model	Bias e	Max variance d	Iterations	% Failure rate	Expense
Round-Table	0	5	1,000	4.7	$2,400
Classical			1,000	2.0	$5,022
Cutoff			1,000	1.8	$1,414
Advanced Round-Table			20,000	1.8	$1,120
Round-Table	0	7	1,000	4.8	$10,400
Classical			1,000	2.3	$5,389
Cutoff			1,000	4.4	$1,661
Advanced Round-Table			10,000	3.9	$1,560

From the table we see that when the judges's capability is high ($d = 5$), the Classical, Cutoff, and Advanced Round-Table Models' precisions are very sharp, while the Round-Table Model's is a little lower. As far as cost is concerned, the Classical Model is highest, followed by the Round-Table Model, the Cutoff Model, and the Advanced Round-Table Model. The costs of the latter two are extraordinarily low.

When the judge's capability is comparatively low ($d = 7$), the precision of every model is almost the same. The cost of the Round-Table Model is too huge to assume and the Cutoff Model and the Advanced Round-Table Model are both again considerably lower.

To get a comprehensive idea of all the models, we summarize various criteria in **Table 6**. The contest committee can decide which model to use, based on this able and the concrete circumstances. We advise that the committee choose the best judges, even though they may cost more than others, for it would prove to be economical on the whole.

Table 6.
Comparative features of the models.

	Ideal	Round-Table	Classical	Cutoff	Advanced Round-Table
Adaptability	v. low	high	high	high	high
Precision	v. high	medium	high	high	high
Expense (large d)	—	v. high	high	low	low
Expense (small d)	low	low	v. high	low	low
Complexity of initializing	—	low	—	medium	v. high
Complexity of execution	low	low	high	high	low
Determinacy	aboslute	high	low	low	high
Flexibility	low	high	low	high	v. high

Generalization of the Model

Different Values for Parameters (P, J, and W)

The Classical Model can be applied directly with different parameter values. For the other models, all we need to do is to determine the parameters for the optimal scheme based on the new values of P, J, and W. For the Round-Table Model, the parameter is the number of times to pass the papers; for the Cutoff Model, it is the number of rejection rounds n; and for the Advanced Round-Table, it is the sequence $\{K_i\}$ of numbers of times to pass papers in each round.

All the empirical formula and laws in our paper are deduced for the particular given values of P, J, and W (100, 8, and 3), so they do not apply to a new problem automatically. However, using the method offered by our paper combined with computer simulation, we can find new empirical formulas and laws and determine the new parameters of optimal scheme easily and quickly.

For instance, let us take Problem B of the 1995 MCM, which had $P = 174$, $J = 12$, and $W = 4$. Let us assume that $d = e = 5$.

Using the algorithm of the Round-Table Model, we find that $n = 4$ is the best choice; the failure rate is 3.0%, and the expense is $13,440. For the Classical Model, we get 2.8% and $21,320. For the Cutoff Model, the optimal value is $n = 4$, the failure rate is 4.2%, and the cost is $3,502. Finally, for the Advanced

Round-Table Model, the optimal set of parameters is $K_1 = K_2 = K_3 = K_4 = K_5 = K_6 = 0, K_7 = K_8 = 1$; the failure rate is 3.0% and the cost is $1,700.

More Winners

Sometimes a few outstanding papers are not the only result of a contest. We may be asked to classify other papers as Meritorious, Honorable Mention, and Successful Participation, as the MCM does, in order to encourage the participants. Except for the Ideal Model, all the selection schemes discussed in our paper are suitable for this task. Compared with other schemes, the Cutoff Model is best, since it ranks all papers.

Strengths and Weaknesses

Strengths

The Cutoff Model and the Advanced Round-Table Model successfully generate selection schemes that can drastically reduce the cost of judges, and we have given practical methods for determining the optimal selection scheme of both models. Both models and their methods are easy not only to understand but to apply in practice, and they easily generalize to most situations.

Our simulation program requires little memory and runs very fast, so it is especially suitable for practical application.

Weaknesses

Since we largely rely on computer simulation to test our models, to select the optimal selection schemes, and to verify our laws, we cannot assure that our result is 100% definitive. However, we ran the simulation program more than 1,000 times before drawing any critical conclusion, and the result of the simulation are very stable. We believe that all our results are reliable enough for application.

Because of the absence of relevant information, our cost function may not conform to reality.