1996 MCM

Problem B
The Contest Judging Problem

Background Information:

When determining the winner of a competition like the Mathematical Contest in Modeling, there is generally a large number of papers to judge. Let's say that there are P = 100 papers. A group of J judges is collected to accomplish the judging. Funding for the contest constrains both the number of judges that can be obtained and the amount of time that they can judge. For example, if P = 100, then J = 8 is typical.

Ideally, each judge would read all papers and rank-order them, but there are too many papers for this. Instead, there are a number of screening rounds in which each judge reads some number of papers ad gives them scores. Then some selection scheme is used to reduce the number of papers under consideration: If the papers are rank-ordered, then the bottom 30% that each judge rank-orders could be rejected. Alternatively, if the judges do not rank-order the papers, but instead give them numerical scores (say, from 1 to 100), then all papers falling below some cutoff level could be rejected.

The new pool of papers is then passed back to the judges, and the process is repeated. A concern is that the total number of papers that each judge reads must be substantially less than P. The process is topped when there are only W papers left. These are the winners. Typically, for P = 100, we have W = 3.

Problem:

Your task is to determine a selection scheme, using a combination of rank-ordering, numerical scoring, and other methods, by which the final W papers will include only papers from among the "best" 2W papers. (By "best" we assume that there is an absolute rank-ordering to which all judges would agree.) For example, the top three papers found by your method will consist entirely of papers from among the "best" six papers. Among all such methods, the one that requires each judge to read the least number of papers is desired.

Note the possibility of systematic bias in a numerical scoring scheme. For example, for a specific collection of papers, one judge could average 70 points, while another could average 80 points. How would you scale your scheme to accommodate for changes in the contest parameters (P, J, and W)?