

Draft

Principles for a multi judge system for competitions (1st, 2nd, 3rd) and exhibitions (gold, silver, bronze).

Principles/assumptions.

1. There is an underlying quality associated with each image that is commonly recognisable by all judges (i.e. judges work from a similar rule book)
2. The score or rank of an individual image assigned by a judge, or the average of scores assigned by a number of judges, represents the quality of each image and is subject to commonly accepted value guidelines
3. Judges are not biased against any particular genre or style of photography
4. In group judging, each judge's score has equal weighting with other judges
5. Judges do not know the author of the work, nor anything about the image apart from what is visible to all judges
6. All images are displayed and judged under equal and optimum conditions
7. For each judge, all images are assessed at a similar time and place
8. Judges must write down their score or rank before any open discussion is held about relative quality.
9. The group average rank of shortlisted images is the default final ranking. Opinions of individual judges disagreeing with this ranking should be confined to technical issues that the judge considers may have been overlooked by other judges. A tie in rank score should be decided by majority vote.

An ideal system should ensure that:

1. An image receives a score or rank commensurate with its quality
2. The scoring system has sufficient separation to rank images appropriately (the WAPJA system uses 8 levels with decimal options within levels, e.g. 5.5, 6.5, 7.9 instead of just 5, 6, 7)
3. Judges do not know scores assigned by other judges until after judging is complete
4. Each image should be anonymous and scored under similar conditions
5. Images may be assessed/reviewed multiple times by each judge
6. Judging should operate under an agreed set of principles with an independent director as arbitrator
7. Judges scores/ranks/opinions are treated equally
8. Judges should sign off on the judging rules and declare conflicts of interest prior to judging

An inferior system:

1. Allows judges to influence each other by openly discussing images prior to scores/rankings being assigned
2. Allows judges to know about the author or the background to the photograph

3. Uses a one-pass scoring system that does not permit re-examination of images after a score or rank is assigned
4. Does not permit examination under ideal viewing conditions
5. Does not allow time to properly assess images
6. Requires judges to evaluate a large number of images continuously and sequentially over an extended time period (mental fatigue)
7. Does not permit all images to be assessed at the same time in the same place.
8. Does not instruct/train judges on the scoring system (judges are working from different rule books)
9. Permits a judge to dominate other judges through strength of personality rather than ability

The Process.

For a system to work effectively, fundamental principles 1 and 2 must be strictly adhered to: that judges all work from a similar set of rules and that the score or rank assigned to each image reflects its quality. If judges do not adhere to these, then the system breaks down and is no better than a lottery. So it is basic for the success of any system that images are sorted into groups ranging from best to worst on the day based on scores or ranks, and that they all work from the same set of scoring guidelines. A Z-score adjustment of scores can be used to minimise differences in the way judges assign scores, but it is not a cure-all.

A. Competitions

Different types of competitions and different practical considerations may require different approaches, but there are basically two main types of competition – those involving a single stage process of selecting winners from prints or digitals, and those involving a two-stage process of first selecting acceptances from digital entries and then selecting winners from prints.

1. Competitions that require a single stage process of selecting winners and placegetters. Examples of this type of competition are typically regional or club events where entrants submit prints and two or more judges select the top one or more.
 - Images are displayed at random without identifying author labels and under uniform light
 - Judges write down their pick of 1st to 5th in rank order for comps involving 2-3 placegetters, or 1-3 for comps asking for only 1st place
 - The adjudicator identifies the top 5 images in average rank order (lowest average rank is best) and asks for confirmation of the top 3 places. If there is strong disagreement the judges discuss their reasoning and re-assign rankings. The adjudicator chooses the top average rank and asks for final agreement on rank order.
 - In the event of a draw, ranking is based on a vote by judges, which may include the adjudicator if judge numbers are equal.

2. Competitions that require a two-stage process. Examples of this are Printwest where a first round uses digital entries to choose acceptances and a second round based on prints to choose a 1st, 2nd, 3rd and a series of merits or commended. The scores in each round are independent, and the process depends upon how the scores are used. For events like Printwest, where digital scores are used to determine club average score, every image must receive a score. For events like the Maritime Museum competition, 50 digital images are required for printing and hanging but final print placings are determined by public voting.

- **Stage one (selection of acceptances).** This is a process whereby images are separated into those that are accepted or rejected for second round judging as prints. While it is necessary to rank or score a predetermined number of images at this stage in order to select the cut-off for acceptances, the number given a specific rank or score depends upon the competition. Typically, the number of images accepted will depend on the hanging space or a predetermined percentage of entries. Images are selected for acceptance on the basis of each judge's rank or score, not average rank or score. For example, if there are 3 judges and a total of 50 acceptances required out of say 300 entries, then each judge will rank or score their top 50 images. There is no discussion amongst judges. The adjudicator will select the 50 acceptances on the basis of first rank of each judge, then second rank until 50 separate images are listed. In cases where an image has already been selected by another judge, the judge misses their turn. It is recommended that if scores are used instead of rank, then adjusted scores be used (see below)
- **For stage 2 (selection of winners and placegetters).** This is on the prints made from the accepted digital images. Judges assign a score (or rank) to each image. The adjudicator calculates rank order based on adjusted score or rank and judges are asked to agree on the top order. Changes from this ranking are to be considered the exception, and based on obvious errors in evaluation rather than opinion. In the event of a requested change in rank, scores are anonymously re-assigned by judges after discussion, and the order re-presented by the adjudicator for final approval.

B. Exhibitions

Exhibitions are judged the same as competitions except that scores are the sole means of assigning awards (Gold, Silver, Bronze). The scores used for the final determination of awards should be Z score-adjusted, and should have sufficient range to separate divisions (e.g. using the WAPJA 8 level system with options for decimal divisions so that there is a potential range of 89).

Preparation of images and ruling on disqualifications:

Step 1 – for digital entries, the adjudicator or CD will first make copies and store them in a safe place.

Step 2 – for digital and print entries, the CD will check each image and disqualify those that don't comply with competition/subject rules. For digitals, all metadata identifying the author is removed. If titles are not permitted then they will be replaced by a 3 digit number in a random fashion.

Step 3 – The CD will instruct judges on the process to be followed and get sign off, then distribute the images with a deadline for each judging stage. It is permissible for judges to

subsequently downgrade or disqualify images they consider do not meet subject or competition criteria.

Where scores are required, images should first be ranked into broad groups based on scores or rank. If, for practical reasons, images are first scored by judges then the scores should be adjusted using Z scores and the final average score used to rank images. In cases where adjusted scores vary between judges by more than the outlier range ($1.5 \times (Q3 - Q1)$), judges should be invited to comment and then re-evaluate their score without knowing the other's scores.

It is useful for judges to understand the reasoning used by other judges, and thus one of the principles is that judges be given an opportunity at the end to explain their rankings or scores. However, this should not be used as an opportunity by a one judge to try and influence another. The rankings should only be changed if judges agree they missed a technical element and made a mistake.

Z score adjustment (used to remove bias due to skewed results by one or more judges)

1. Use an Excel template with image number down the page with number in column A (one row per image) and judge name across the page (one column per judge)
2. Enter the score (1.0 to 8.9) for each image for each judge (χ).
3. Calculate the overall mean for all judges and images (population mean, μ). Use the AVERAGE formula in Excel
4. Calculate the overall variance for all judges and images using VARP in excel (σ^2)
5. Calculate the mean for the judge (mean judge A $\bar{\chi}$). Use the AVERAGE formula
6. Calculate the overall variance for the judge (s^2) Use VAR.S in excel
7. Calculate adjusted values for each judge for each image using the formula:
$$\mu + \sqrt{\sigma^2} \times ((\chi - \bar{\chi}) / \sqrt{s^2})$$
8. Calculate the average adjusted value for each image by averaging across the judges
9. Rank the images based on their adjusted means (8).

Querying divergent results

A range of 2.0 or more points in an 8 point scale between judges for an image indicates a wide disparity in assessment (2 levels in the WAPJA system) and should be flagged for re-assessment. This is done on the adjusted values. Judges are requested to explain their score for the disputed image without reference to others, and then after all judges have commented, the judges are invited to submit a revised score if they wish.

Systems currently in use:

Print based:

APPAs – images are presented as prints one at a time in sequence to a panel of 5 judges who assign a score from 50 to 100. Challenges are permitted if a score by one judge (usually) exceeds another by 5, and judges may increase their score after debate.

Scores determine Gold, Silver, Bronze awards. Acceptance is “professional standard” (70 points).

The images are displayed under uniform lighting, there is no debate initially and judges scores are entered by computer, with the final scores being shown only after all scores are submitted. There is no going back once an image has been scored unless there is wide divergence between judges. To avoid fatigue, judges are replaced after a certain time. Multiple panels may operate in multiple rooms, so different judges get to assess different images

Pros – Quick and efficient - given the large number of entries, and the fact it is a print-based system, this process is probably the only practical way to group judge prints in a one-step process.

Cons – Judges cannot go back and change their scores, nor can they compare their score to a previous score. There is likely to be a fatigue factor, so luck plays a part in when an image gets assessed in the fatigue cycle. Scores usually get challenged upwards, so a high scorer on the panel will bias the outcome upwards. Likewise a judge with a bias against a genre or style who refuses to shift upwards will bias the results for that panel session.

Royal Show

Panel judging (3) of prints in sequential order, with acceptances based on a minimum total score. The scoring system is 1-5, although 1 is not used as it implies a failure to meet subject specifications. Challenges are permitted in cases of widely divergent scores.

The selection of finalists is based on original scores (say top 10) then by agreement amongst judges.

Pros – as above for APPAs

Cons - as above for APPAs

Digital acceptances, prints final

Printwest – images are scored individually and then the top 50 images (based on total score) in each category are selected for printing. The prints are examined in a room all at once, then scored individually in a sequential display system. The ten top scoring images are displayed together and the winner chosen from these after group discussion

Pro – fast

Con – open to bias (score based system) and influence by dominant judges.