Selection Competitions

(Concepts and Best Practices)

Bert van den Berg

February 2021

Overview

There are numerous selection processes where people are recruited to score submissions against a fixed set of criteria. This includes many government-run funding programs, opportunities offered by charities and public competitions such as hackathons. Important concepts for such competitions include:

- defining the selection criteria,
- recruiting experts to score the submissions,
- assigning experts to review specific submissions, and
- combining scores from experts to rank the submissions.

This document explores the concepts important in selection processes, including some best practices and issues. It also raises questions for further research that might provide further insight into competitive selection processes.

This document is intended as a companion to the “Introduction to CoSeT” document, which describes an open-source tool to support data-intensive aspects of competitive selection processes.

Contents

Introduction to Competitive Selection Processes

Organizations offering opportunities that attract many submissions may engage a pool of experts in the selection process. Such competitions often have multiple winners from among these submissions. To increase the consistency and transparency of the selection, these experts[1] are asked to score the submissions[2] against criteria. The scores are then compiled to arrive at a ranked list of submissions.

Examples of competitive selection processes that use the type of processes supported by CoSeT[3] include:

- The grant selection processes used by various programs in government departments and agencies, including those used to award research funding.
- Hackathons that attract large numbers of submissions from informal teams.
- Opportunities offered by charities and other public organizations.
- Staffing processes aimed at building pools of qualified candidates from large numbers of applicants.

This document is intended to introduce important concepts for competitive selection processes and is intended as a precursor to a discussion of CoSeT, an open-source tool which supports data-intensive aspects of competitive selection processes.

Concepts for Selection Processes

When selection processes use experts to score submissions against criteria, the general sequence involves:

- defining the selection criteria and launch the competition,
- recruiting experts to score the submissions,
- assigning submissions for the experts to score,
- gathering scores from the experts,
- compiling the results, and
- distributing feedback on the submissions to the applicants.

These processes are described below, along with potential issues and best practices.

Defining the Selection Criteria

The selection criteria should be few (three to seven), orthogonal (non-overlapping), and readily understood by both those creating submissions[4] and the experts doing the scoring. To understand the

[1] In this document, the people who score submissions will be referred to as markers, reviewers, or experts interchangeably. CoSeT internally uses the term 'Marker'.

[2] In this document, the item submitted to the competition will be interchangeably referred to as application, proposal, project, or submission. CoSeT internally uses the term 'Project'. In some competitions a person (candidate), group or team may be competing, notionally via the submission(s) they have provided.

[3] CoSeT is a Microsoft Excel workbook with tables and macros that help with the management and flow of information for competitive selection processes, particularly assigning markers to projects, and compiling (normalized) scores for projects.

[4] In some types of competitions (staffing pools come to mind) the scoring rubric may not be available to applicants.

challenges involved, consider the design of criteria for a research funding program. Typical criteria would address the qualities of the proposed research team (their expertise, and scope relative to the research proposed), one or more criteria addressing the research to be undertaken (research novelty, management of research activities), perhaps a criterion related to the student training opportunities, and a criterion related to how research results will be shared with those who can potentially use them.

While the experts may have well established terminology, those submitting applications may be coming from other research domains or have less research experience, and thus have a different understanding of terminology.

The challenge of creating orthogonal selection criteria becomes clearer when considering the overlap between the training and other criteria. If the research plan is poorly organized, or the researchers have a past track record of excellent training, or if the research partners are not interacting with the students, it may impact both the score for the training criterion, and the scores for the research plan and the research transfer criteria.

A variety of approaches are employed to improve the design and effectiveness of selection criteria. One approach is to use focus groups from the applicant community, competition domain experts, as well as people who understand criteria design and evaluation processes. Their combined guidance can help create criteria that effectively implement the competition's objectives[5]. Language ladders help structure scoring by offering specific (and typically objective) descriptions of what corresponds to a particular rating for a criterion. Their use can improve the consistency of experts' scoring[6],[7].

Recruiting Experts

Competitive selection processes require people who have expert knowledge of the general topics covered by the submissions and are willing to contribute time to a result they may not directly benefit from. By acting as an evaluator for a competition, experts may gain:

- insight into the people, organizations or projects being proposed,
- recognition from their community as leaders, and
- acknowledgements for their (often voluntary) contributions.

Beyond having appropriate expertise, it is important that those recruited reflect the priorities of the organization hosting the competition. For example, competitions in support of activities that will impact particular groups should have those groups significantly represented among the scoring experts. In recent years, Equity Diversity and Inclusion (EDI) considerations have encouraged the recruitment of experts from under-represented groups. It is also generally seen as important to have a mixture of new and returning experts. The experienced raters tend to have a deeper understanding of the process and can help train the new people. However, experienced raters may also perpetuate unhelpful habits.

---

[5] For further advice on designing selection criteria see: https://facultyaffairs.gwu.edu/guidelines-developing-selection-criteria.

[6] ProGrid offers a full scoring system that includes customized language ladders for specific programs (and other tools to support competitions).

[7] For an example of a language ladder, see NSERC's Discovery Grants Merit Indicators table: https://www.nserc-crsng.gc.ca/_doc/Professors-Professeurs/DG_Merit_Indicators_eng.pdf.

Assigning Experts to Submissions

This step involves deciding which experts will review which submissions. Submissions and the experts recruited generally cover a spectrum of knowledge and experience. The quality of the submission scoring can be enhanced by ensuring each submission has the best quality of expertise assigned to it, and/or asking experts to rate proposals their expertise is better aligned with[8]. However, experts must not be involved in the scoring of proposals where they are in a conflict of interest[9] - which often will exclude those with the most expertise relevant to a submission. Both the number of scoring assignments for each expert, and the number of experts scoring each proposal is important.

Statistically, as the number of assignments per expert increase, the probability increases that an expert receives a group of submissions that is representative of the competition. If markers receive too few assignments, the expert will not have a good sample for self-calibration, and they may be assigned a subset of the competition that is stronger or weaker than the overall competition[10]. If they receive too many assignments experts may not have time to carefully score the submissions.

Research has shown that more than 10 experts scoring each submission is required to ensure that the average of their scores closely estimate the score that would result if all the available experts scored the proposal[11]. However, for practical purposes, the number of experts scoring a proposal is typically between three and five.

Ideally, scoring assignments optimize the fit between the expertise of those evaluating the submissions and the set of submissions received. The process of assigning experts to score submissions is further discussed below.

On Assigning Experts to Submissions

Various strategies can be considered when assigning experts to score submissions. A first order approach is to assign any submission to any marker (i.e., ~randomly).

An improvement on the first-order strategy is to ensure that markers are not asked to score submissions where they there is a reason to exclude them (including conflicts of interest)[12]. Conflicts of interest typically include:

- Personal involvement in the submission, or a competing submission,
- Role in advising on the submission.
- Personal relationship with the proponent(s) of the submission.
- Family relation to the proponent(s).
- A work relationship with the proponent(s) (often this is time-limited, i.e., last N years).

---

[8] CoSeT provides a tool to automate the assignment of markers to submissions.

[9] Experts are also excluded from consideration as markers on individual submissions for reasons beyond conflict of interest. CoSeT use 'X' for all situations where experts are e**X**cluded from possibly being assigned to score a submission.

[10] Important when considering normalization (discussed below).

[11] See the paper by Richard Snell: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0120838

[12] Other reasons to exclude an expert from the scoring of a particular proposal include having previously scored a previous version of the submission, or not being comfortable in the language that the submission is written in. Some competitions have experts (potential markers) mentoring competitors (i.e., in hackathons).

Since this strategy is based on asking people for their expert advice (via scoring), a third level strategy is to assign the experts to score submissions that better match their expertise (while not being in a conflict of interest). There are two general approaches to matching experts with submissions better aligned with their expertise:

1. **Direct:** In the direct approach each marker is asked to rate their confidence in scoring each submission based on a small amount of information about the submission. This might include the title and a short precis.

2. **Indirect:** The indirect approach requires the experts to indicate their expertise on a set of keywords relevant to the submissions in the competition. This requires the competition organizers to develop the list of keywords pertinent to the anticipated (or received) set of submissions.

For example, consider a competition focused on climate change. In the direct approach, each marker would be given the titles and abstracts of each of the submissions and asked to signal their anticipated confidence reviewing each submission (i.e., High, Medium, or Low).

In the indirect approach, each marker would be given a list of keywords related to climate change (e.g., atmospheric modelling, rising sea levels, displacement of people, crop failure, dis-information, financial planning) and asked to signal their expertise in these areas (i.e., High, Medium, or Low). The markers' self-ratings would then be combined with ratings of how relevant the same keywords are to the submissions to arrive at estimates of the likely confidence of the markers to rate a particular submission.

The direct method is expected to produce more relevant expertise information but requires the experts (often volunteers) to provide confidence assessments for all the submissions and thus requires more effort. The indirect method requires more work and technical knowledge by the competition organizers and can be expected to produce less specific expertise insight than the direct method.

CoSeT supports both methods of assigning markers. Its current algorithm attempts to give each submission a similar level of expertise among those assigned, and where possible to maximize the overall use of marker expertise.

In summary, the following general concepts apply to assigning markers:

- Some markers will need to be excluded from review of individual submissions due to conflicts of interest, language of the submission, etc.
- Information about the experts' expertise with regards to the subject(s) of the competition (or its submissions) allows for a better alignment between markers and the submissions they score.
- The markers should each be assigned a statistically representative sample of the submissions.
- Each project should get the same number of markers assigned, ideally a statistically significant sample of the markers participating in the competition.

Challenges with Scoring Submissions

Scoring the submissions is prone to significant challenges. Beyond the general perceptual biases that we all have[13], experts may be more generous for topics new to them. It has been observed that some

[13] Well-structured competition processes anticipate common perceptual biases. Training staff to recognize such biases, and to challenge experts to address them can improve the quality of selections.

experts may also have a personal tendency to score submissions more harshly or generously than their peers.

Scores are generally used to rank order the submissions. In an ideal world, where each person has deep expert knowledge, and applies the same (effective) processes to score the proposals, all scores on a submission would be the same (i.e., the variance between scores on a proposal would be zero). However, there is often significant variation observed in the scores from experts on a single submission. This variance in scores significantly lowers the confidence in the final ranking of submissions in a competition[14]. In this context, better executed selection processes should see lower variance in the experts' scores on submissions.

Typically, the submissions that gain the highest or lowest final rankings tend to have the most consistent scoring[15], while the scoring variance increases substantially between these limits. Unfortunately, this means that the scoring variance is often high at the cut-off between those proposals that are accepted and those that are not.

Techniques to Improve the Quality of Scoring

Given the significant challenges that can be observed with realizing objective and statistically-valid scores from the assigned experts, a variety of techniques are applied to improve scoring quality[16]. These include:

- Using a language ladder (noted above).
- Providing experts with example submissions along with the expected scoring to them help calibrate their scores.
- Making experts aware of cognitive biases.
- Training the experts about the expectations for the competition, and running through the scoring of one or more sample submissions, including the use of the language ladder for the scoring.
- Encouraging experts to self-calibrate their scores across the submissions they score.
- Normalizing scores between experts. If the number of submissions each expert is asked to rate is a significant sample, then statistically, the totals of the scores from the experts should be the same. Normalization produces a consistent average for the total scores provided by each expert. In this way, normalization compensates for experts who tend to be generous or harsh in their scoring.

Selection of Best Submissions

The final set of experts' scores are generally combined in a master scoresheet. Typically, the scores for a submission are averaged and the top-ranking submissions are selected.

---

[14] Some competition processes follow-up the scoring phase with a qualitative discussion and ranking process.

[15] It is comforting to think that experts agree on the best and worst proposals … and thus resulting in smaller variances for these submissions. However, there may be less variation because the span of possible scores available at the high and low and of the scoring range is narrower.

[16] Aside from supporting normalization of expert scoring, CoSeT does not constrain or support any of the above-mentioned techniques for improving submission scoring.

Feedback to applicants is appreciated and helps them improve. One approach to crafting the feedback is to assign a marker with the responsibility of assembling a coherent set of feedback for the submission (this role may be called the 'first-reader').

Funding competitions run by public agencies tend to involve committee meetings to discuss the submissions before the final selections are made. Research indicates that these discussions among experts improves the understanding of strengths and weaknesses of submissions and does see experts revise some scores. However, the research also indicates that such discussions do not meaningfully reduce the variance between expert scores and have minimal impact on which submissions are accepted/funded[17].

Information management

The foundational information elements in a selection competition includes information about the submissions, selection criteria, markers and their expertise. Further information created during the competition includes the assignments, scores, comments from experts and the final list of ranked submission scores.

Historically, this information was managed on paper, with information sent to and from experts by mail. This evolved into information managed on office productivity tools and/or customized databases with files shared by email. These approaches have often been supported by templates and customized spreadsheets. More recently, some organizations have invested in bespoke information management systems that include secure extranet websites for sharing information with experts. For organizations with fewer resources, the use of online sheets and online forms have proven to be useful for managing information and sharing it with the scoring experts.

Other Concepts Important for Competitive Selection Processes

Other concepts that are important to the quality of competitive selection processes include EDI, and anonymizing. EDI may be reflected in specific aspects of the scoring criteria and may also be a blanket policy that submissions are expected to address as appropriate.

It is important that those scoring the submissions do not apply their own biases to scoring the submissions they have been assigned. For example, research has begun to show that evaluators are influenced by the identity-linked traits of applicants (gender, ethnicity). This has led to orchestra applicants performing behind a screen, and some scholarship and HR-screening processes being anonymized.

Some organizations ask people scoring submissions to take (online) training in unconscious bias or EDI before completing the evaluation of the submissions they have been assigned[18].

[17] This is also consistent with my years of experience with selection processes.

[18] For example: https://www.chairs-chaires.gc.ca/program-programme/equity-equite/bias/module-eng.aspx?pedisable=false

Topics for Further Research

Ideally, the competition organizers should apply an approach to competitions (criteria design, marker recruitment, marker assignment, and scoring methods) that is likely to optimize the competition results. However, the research to date on selection competitions is limited. Further, what research is available is largely focused on the individual process steps, with little research linking the results of competitions to the competition methods used. Consequently, competition organizers often rely on experience and organizational preferences when designing selection competitions. In this era of 'big data' there is an opportunity to use research to strengthen the performance of selection processes by linking downstream impacts with competition participation[19].

Below are some possible research questions for consideration:

Defining the selection criteria: How does the wording using in competition calls and descriptions influence the decision to participate, and the qualities of the submissions (particularly from different EDI groups[20])?

Recruiting experts to score the submissions**:** Expert participation is generally voluntary, and people who might apply to the same competition or organization appear to volunteer more frequently. Do these volunteers-cum-future-applicants trend towards certain scoring patterns? Do the experts who participate have different personal characteristics that bias against some of the types of submissions that the competition organizers may desire?

Assigning submissions to the experts to score**:** What other information about the submissions, experts, competition domain and criteria leads to assignments that improve the overall impact of the set of submissions selected? What other attributes of people (beyond self-declared expertise) align with high quality scoring of the submissions? What should the objective (metric) be when assigning submissions to markers?

Gathering scores from the experts**:** Does the time delay between experts receiving their assignments and when the experts' assessments are returned align with the professed level of expertise, or the characteristics of the scores returned. What are the impacts of EDI training, unconscious bias training, example submissions with example scores on the results of competitive selection processes?

Compiling the competition's results: Given the significant variance to be expected among the submissions ranked near the win/lose cut-off zone, there are opportunities to explore how other factors correlate with the success of applicants/applications in the competition. For example, submissions that are statistically indistinguishable from others across the win/lose cutoff could be justifiably selected based on other criteria. Subsequently tracking of the results of these experiments could provide insight to improve the impact of the competitions. For example, statistically indistinguishable submissions could be selected:

- randomly … to see if subsequent impact correlates with their nominal ranking, or

[19] The Innovation Growth Lab led by Albert Bravo-Biosca @ NESTA has encouraged organizations to conduct research experiments to improve their understand of, and impact from innovation support programs.
[20] For example, there is research signaling the impact of language on the participation of women in competitions.

- randomly … and tracking a wide variety of demographics associated with the submission to see if any are aligned with better subsequent impact, or
- based on a characteristic of the submission that might align with future success (i.e., young researchers being more innovative).

Feedback from experts on the submissions**:** does the quantity or existence of feedback from experts correlate to the level or variation in scoring, and does the type or existence of feedback improve the success of subsequent applications?

Competitive selection processes as described above have been used for many decades. Developments in information technology are opening the door to new insights into competitive selection processes.

Acknowledgements

While the author is responsible for any errors or omissions, I would like to thank Dawn Davidson, Deepak Gupta, Guillaume Beaulieu-Houle and Mary Lou MacLean for their contributions to the development of this document.