

Talent selection, and the funding of research

Peter van den Besselaar and Pleun van Arensbergen

We very much welcome the reply of the Netherlands research council NWO (Maessen 2012) on our recent paper on talent selection. As Maessen correctly states, the focus of our paper is different from the report we wrote for NWO. In the report (Van den Besselaar 2011) we analyze the relations between the different steps in the procedure and between the criteria deployed, and we are satisfied that the council could make use of the findings. In our paper (Van Arensbergen et al 2012) we use the data for a better *general* understanding of the possibilities and problems of talent recognition and selection. The paper analyzes grant selection from the perspective of uncertainty: what are the sources of uncertainty in the procedure, what are the consequences of it? Briefly, Maessen claims our analysis is wrong in four aspects:

- Where we conclude from changing and diverging assessments that the process is full of uncertainty, Maessen claims that during the process the assessments improve because of the additional information that comes available in the different steps of decision making;
- One of the steps is the external peer reviewers' assessment. Maessen claims that the peer reviewers' role is strong, where we argue the opposite.
- We show that there are no clear boundaries of excellence and the differences between success and failure are small, which leads to uncertainty in the selection. Maessen disagrees, and claims the awarded applicants score substantially higher than the non-awarded.
- Finally, Maessen claims that the procedure is designed to avoid a too early (and wrongly) rejection of good applications, and that this is successfully done.

As research funding is a crucial element of the research system, we find it important to use this opportunity to clarify a few issues. Firstly, the NWO procedure consists of a sequence of steps, in which new information reaches the panel step by step. Increasingly better informed, the panel should be able to improve the prioritization through the process. According to Maessen, this would explain the changing rank order: with the additional information someone may end up higher or lower in the rank order than earlier in the procedure. To some extent we agree that additional information may improve the ranking. However, we show that changes are considerable, suggesting that the information in the different phases is contradictory. In other words, the rank order is not improving throughout the procedure, but is 'contextually' changing, by giving a higher weight to the recent information. This supports the claim that the way the procedure is organized does influence the outcome. It is one of the difficulties in identifying talent, and underlines the importance of considering how various current selection procedures are organized.

Therefore, new information may actually be counter productive. For example, we show that the scores after the interviews are very different from those before the interviews. But what information the interview precisely brings is unclear, as this may lead to stage performance skills being interpreted as scholarly

excellence. Apart from *what* information is added, the question is *how* panel members use that information, also in relation to the other information already available. How systematically does the panel weigh all available information, and is this done in the similar way for all applicants? Or is the most recent information (e.g. stage performance) actually dominating and is earlier information (the reviews) forgotten (Van Arensbergen et al., forthcoming)? The latter possibility is supported by our observation that panelists are not able to position applications on an absolute scale, but deploy the scale always relative to the set of applications under discussion. This results in low scores for earlier high scoring application, simply because the lower scoring applications are not any more under discussion. Furthermore, if it is correct that more information means a better assessment and decision, why then not collect all information about all applicants? That would inform the panel maximally, and would lead to a real balanced and well-informed decision. Of course this takes more time, but a fair comparison between proposals would require this. On the other hand, it is interesting that e.g., the prestigious advanced grant of the ERC does not use interviews in the selection process.

Secondly, we found a large difference between the rank order based on the peer review scores and the first panel scores. According to Maessen, panelists use the reviewers scores, together with the rebuttal of the applicants. Having in this way more information, the rank order of the panel may differ from the rank order following from the peer reviewers. However, the panel (not being experts) may neglect (and in fact does) peer review scores when they are dissatisfied with the quality of the reviews or disagree with them (Van Arensbergen et al., forthcoming). The relative low correlation between panel and reviewers' scores indicates that this occurs regularly. Below we plot the reviewers' score by success for two fields, one for the early career grants and the other for the intermediate career grants. In the first case (figure 1), the peer review scores seem to provide a kind of threshold (about a 2.3 score) and only better scoring proposals are eligible for funding – and the panel selects among those. In the second case (figure 2), the review scores are not used as threshold at all, and even a low review score of 3.5 can end up in being funded. This supports our conclusion about the low influence of peer review.

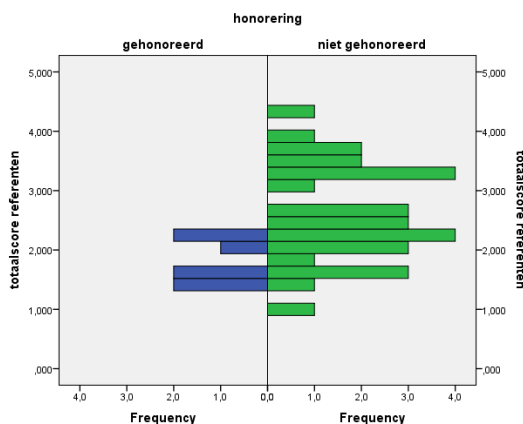


Fig 1: Success by external review scores for (field)

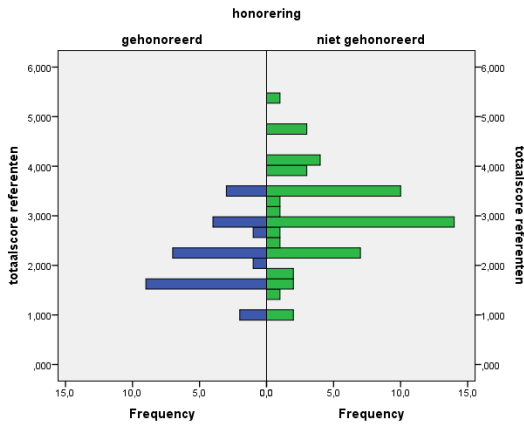


Fig 2: Success by external review scores for (field)

Thirdly, we argue that the difference between success and no-success is very small in many cases. We did not find a clear ‘excellence boundary’, which distinguishes the successful from the unsuccessful applicants. Maessen states that generally the awarded applicants score quite high compared to the others, implying that the top is correctly selected. However by analyzing the whole sample or applicants as one set, we showed in our paper that this is not the case. For a large set of the applications the review scores of selected and rejected ones hardly differ from each other. Recently we also checked this at the level of the individual 27 panels (Van Arensbergen et al 2013). In total, some 260 applicants were selected for a grant. However, only some 50 had a score that clearly differed from the scores of the non-selected applicants at some point of the evaluation process, e.g. in both phases (figure 3) or in the last interview round (figure 4). Although it is good that these outstanding applicants were among the grantees, it also means that in 210 of the cases, those selected scored about similar as others that were not selected (figure 5). Even if in the third case the applicants with the best scores were selected, one cannot conclude that they were significantly better than others that were not selected.

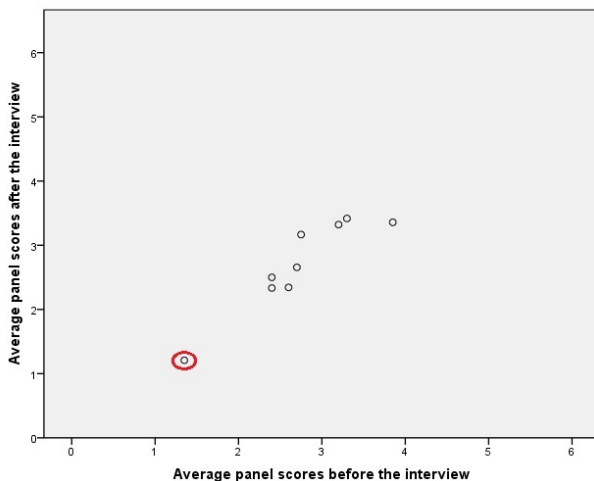


Figure 3. Average panel scores before and after interview in a panel which clearly identified a top talent in both review rounds

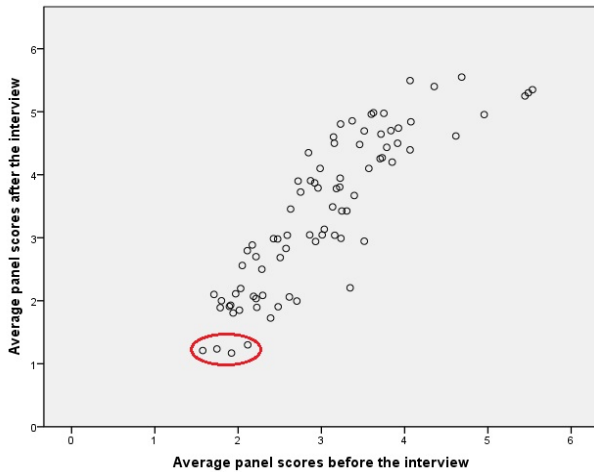


Figure 4. Average panel scores before and after interview in a panel which identified top talents only after the interview

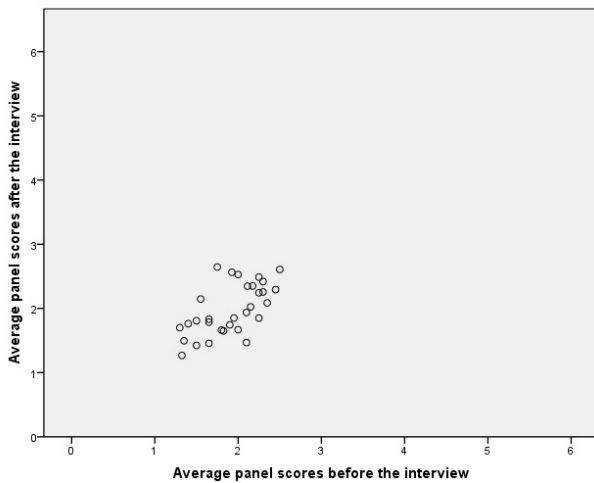


Figure 5. Average panel scores before and after interview in a panel which identified no top talents

Finally, Maessen argues that a main goal of the procedure is to avoid removing possible good applications from the procedure too early. This we actually did not investigate, so our study does not show whether this is the case. The changes in the rank order between the various phases suggest that improvements can be made here: applicants that are rejected early may have ended high with the additional information in the later phases of the procedure. However, this issue could be clarified by studying the so-called predictive validity of the procedure. The question to be answered is whether the group of awarded applicants has a better 'post grant performance' than those that were not selected. And, related to the procedure, is the 'post application performance' of those that were rejected early in the procedure indeed lower than of those that were selected, or rejected later in the procedure? Only a few studies about post-performance are currently available (Bornmann 2011), and some of them suggest a rather low predictive validity (Bornmann et al 2010; Van den Besselaar – forthcoming

With NWO, we feel that regularly studying the way grant decisions are made may help to improve the procedure step by step. The following question should be addressed urgently: (i) In-depth studies of the decision making process are needed, preferably based on observations of the panels. What information is available in the panels, how is it used in the social process of grant decision-making, and what (discipline specific) criteria are in fact deployed? (ii) How does the composition of panels influence the outcomes of the selection process, in terms of e.g., gender, disciplinary and topical bias? (iii) Finally, research is needed to investigate predictive validity: can we say that with hindsight the best researchers were selected – and what procedures do this better than others?

The results of this kind of studies may improve grant selection processes. However, even then there will never be a perfect procedure, as recognizing talent is done by people and therefore is an inter subjective process remaining to a large extent uncertain. But it may increase the legitimacy of the procedure. And that is important, as success in grant application is considered to be a very important indicator of talent, and consequently increasingly decisive for an academic career.

References:

- Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology*, **45**, 199-245
- Lutz Bornman, Loet Leydesdorff, Peter van den Besselaar, A Meta-evaluation of Scientific Research Proposals: Different Ways of Comparing Rejected to Awarded Applications. *Journal of Informetrics* **4** (2010) 3 pp.211-220
- Langfeldt, L. (2001) 'The decision-making constraints and processes of grant peer review, and their effects on the review outcome', *Social Studies of Science* 31(6): 820-841
- Maessen, K.M.H., Comment to the article by Van Arensbergen and Van den Besselaar. *Higher Education Policy* **25** (2012) pp 531-534
- Van Arensbergen P, P. van den Besselaar, The selection of scientific talent and in the allocation of research grants. *Higher Education Policy* **25** (2012) pp 381-405
- Van Arensbergen P, I van der Weijden P. van den Besselaar, Academic talent selection in grant review panels. In: Prpic & van der Weijden, *TITEL TO BE ADDED*, Petersburg: PUBLISHER TO BE ADDED 2013
- Van Arensbergen, P., I van der Wijden & P. van den Besselaar, *Competition for talent. What are the talents we are looking for in science?* (forthcoming).
- Van den Besselaar, P. *A quantitative analysis of decisions about career grant applications – a report for NWO* (in Dutch). Amsterdam, 2011.
- Van den Besselaar, P., *More competition, better science? The predictive validity of grant selection*. 2013 (forthcoming)