

The Mediating Assessment Protocol

INTER's 2023 UvA Election candidate selection and ranking strategy

Board: Kate, Aleksandra, Ryan, Isabeau, Lexi, Yeva

Interviewer: Yeva



“But in public conversations about human error and in organizations all over the world, noise is rarely recognized. Bias is the star of the show. Noise is a bit player, usually offstage. The topic of bias has been discussed in thousands of scientific articles and dozens of popular books, few of which even mention the issue of noise.”

(Noise, p.10)

“Wherever you look at human judgments, you are likely to find noise. To improve the quality of our judgments, we need to overcome noise as well as bias.”

(Noise, p.11)

To start off, it must be illustrated what immense importance of the work we are about to conduct in implementing the Mediating Assessment Protocol for the first time at INTER. *Noise – a flaw in human judgment* describes research findings of staggering 90% of managers, employees and HR specialists in the USA who believe that their performance management process failed to deliver the results that they expect. Furthermore, that in job recruitment the true variance, variance attributable to the candidates performance, accounts for no more than 20-30% of the total variance. The rest of the variance is composed of 70-80% noise: unwanted variability. That is because the recruiter is maintaining different understandings of the scale (level noise), because the recruiter is in different moods at each job interview due to being hungry or pissed off or the weather is bad (occasion noise), or because they maintain a different personal attitude to the candidate based on personal liking (pattern noise). This arbitrary variance is what we seek to reduce.

Whenever there is a decision, there is **noise**: unwanted variability in judgement. This is true across *all* fields of human endeavour: forensics, insurance premium setting, medicine, management decisions, political forecasting, job recruitment and even the criminal justice system. This noise leads to incredible amount of arbitrariness, unfairness and injustice; this is not a trivial matter, you might find yourself denied the dream job you worked hard to attain, due to a hostile interviewer who prefers candidates who match his passion for a sport or hobby totally unrelated to the task at hand. We seek to provide a more just world where, in this case, our candidates are judged on merit, and on criteria that we deem important for the job.

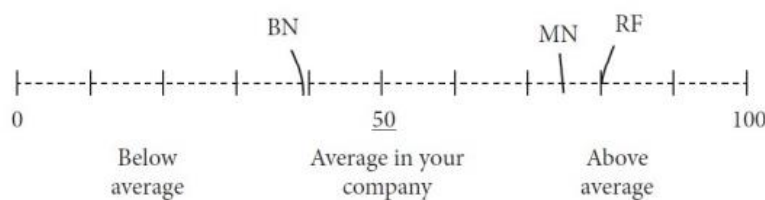
I would like to say, we can entertain any and all amendments, and makeshift stopgap solutions and changes that we need in order to adapt and bring this to a successful close. This is our first time implementing the MAP, and it is bound to not be flawless.

An overview of the Mediating Assessment Protocol

We utilise the **forced ranking method**. The benefit of forced ranking is that we are prevented from giving everyone the highest possible rating. Instead, we are forced to make relative judgments rather than absolute judgments. We as humans are better at relative judgments. Relative judgments (1st, 2nd / top 80% , top 50%) are better, fairer, and more accurate than absolute judgments (‘good’ or ‘bad’). Of course, it is to be noted that forced rankings mean that the candidate who is ranked last is not necessarily a bad candidate. All candidates in our ranking can be good, but the person ranked lowest is the least good. This is not a serious drawback of forced ranking, as what we want to know is who is the best and who is the worst candidate: that is how the party-list system at the UvA elections work

s too. If there were to emerge any serious issues or concerns that may lead the INTER board to eliminate or advise to eliminate, we can still bring this up in our discussion part of the Mediating Assessment Protocol. Other than that, we prefer that the candidate runs, even if they are ranked last after having reviewed all attributes.

To be concrete. We will be ranking the candidates for each attribute on a described, qualified and meaningful scale. This will be a scale ranging van 0% to 100%. Below is an example relative rating scale:



The interviewer, Yeva, is tasked to position the candidates on this scale, judging them relative to each other on their performance on the specific attribute in question. Again, the benefits of rating the candidates on one scale are manifold: 1) we structure the complex judgement of assessing the candidate’s suitability as a council member into several dimensions, which limits the **Halo effect**, which is the error of excessive coherence that one positive assessment of the candidate on one attribute leads to unwarranted positive assessments on the other attributes. 2) it reduces the **pattern noise** and the **level noise** that we suffer in our decision making. Pattern noise regards the effects of the interviewer’s attitude of the candidate or the board member’s attitude of the candidate on the ratings of the attributes. The interviewer task is solely to collect data and not to discuss anything else in the **structured interview**. Level noise regards differences amongst the board members and the interviewer as to what the extremes of the scale actually mean e.g. whether 100% of the scale means perfection, or simply above average.

The reviewing we're doing here and the Mediating Assessment Protocol we use ensures that every one of us maintains a **common frame of reference**, maintains an **outside view**, that we're using **behaviourally anchored rating skills** to fairly judge all candidates, and that in doing so, we are laying the foundation for a case database in which, every successive year, the INTER board will be even better equipped to rate their candidates on their predicted future job performance: by being able to rely on earlier reference cases of candidate

recruitment and evaluating whether the predictions were correct. Ideally, our candidates would be interviewed and then assessed by four different interviewers conducting their assessments independently, separately. They would do so before they communicate with each other and then we aggregate their assessments. Of course, because at INTER we are restricted to one interviewer, and the MAP is filled in after the interviews have taken place. To counteract **occasion noise**, the interviewer’s assessment and the rankings by the board later on in the **estimate-talk-estimate** method is done at the same time respectively.

It is the board that will make independent, separate judgments by using the **estimate-talk-estimate** technique.

A real life example of these ‘decision hygiene’ strategies being implemented in real life is Google; which has successfully implemented them and uses attributes in their checklist such as general cognitive ability, leadership, cultural fit (lit. ‘googliness’ for us can be ‘INTERiness’) and role related knowledge. These main attributes are then broken down into smaller components. Important here is that any attributes such as the candidate’s good looks, smooth talk, exciting hobbies and any other aspects whether positive or negative that the recruiter might notice in an unstructured interview are excluded. Meaning, only relevant attributes are put on the checklist.

In the final judgement. We introduce our **holistic judgement** that we have delayed thus far. The goal of the Mediating Assessment Protocol is not to remove our intuition or personal judgement. We want to treat the candidate with dignity and respect, as well as allow for the surfacing of ‘skeletons in the closet’ or any factors that the Mediating Assessment Protocol can’t possibly foresee. But we only intuitively judge after all the evidence has been collected and analysed. Thus the tendency of each interviewer and board member to form quick, intuitive impressions and rush to judgement is kept in check.

The three principles of the Mediating Assessment Protocol are thus summarised as:

1. Decomposition of complex judgments into simpler attributes
2. Independent assessment on each attribute
3. delayed holistic judgment

These are principles that are broadly consistent with the recommendations that organisational psychologists have formulated over the years

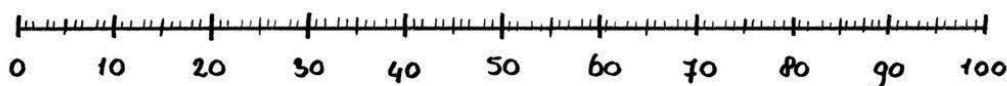
The Mediating Assessment Protocol is the authors of *Noise – a flaw in human judgment* ‘s gift to organisations such as INTER, that incorporates most of the decision hygiene strategies that have been introduced in the book. The research shows that structured interviews produce better results than unstructured ones, and that structuring a hiring decision improves it. It leads to higher accuracy, which unstructured interviews do not even come close to attaining:

“If you are unfamiliar with research on the employment interview, what follows may surprise you. In essence, if your goal is to determine which candidates will succeed in a job and which will fail, standard interviews (also called unstructured interviews to distinguish them from structured interviews, to which we will turn shortly) are not very informative. To put it more starkly, they are often useless.”

(Noise, p. 301)

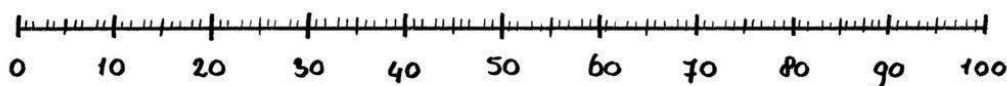
We start the Mediating Assessment Protocol, by deciding in advance on the list of attributes. A **checklist of traits or attributes a candidate must possess**. For the sake of coherence we shall stick to attributes as the term to use. We have decided that the following 5 attributes are important for a candidate to possess.

1. Experience as a member of a student council, or student representative body, or similar, at secondary school or higher education or similar, measured in the number of

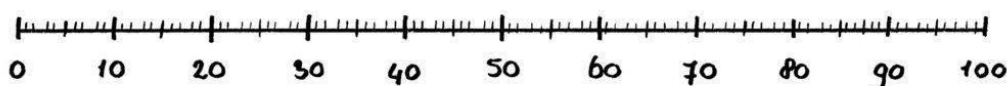


years spent in such positions and the intensity of the office. Experience should count only from 16 years of age onwards.

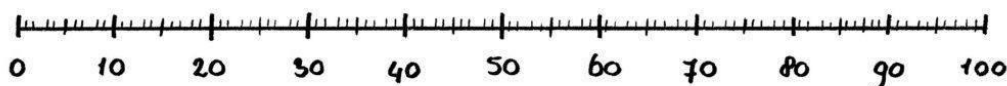
2. Motivation of the candidate to become a council member. What reasons have led the candidate to decide to run for council member, at this time, for INTER and not another party, for that particular council and why?



3. Vision of the candidate and their concrete ideas to strive to realise during their tenure as a council member for the applied for specific council. What is the vision and what are the ideas? How expansive and thought out are they? Are they feasible within the limits and competences of the specific council? Has the candidate put any thought into that feasibility? What distinguishes this candidate from other candidates?

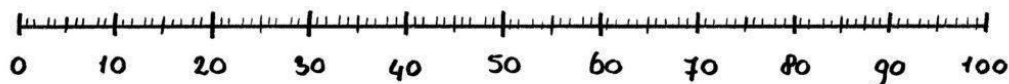


4. CV strength: to what extent has the candidate shown to engage in extracurricular exploits besides representative functions, that demonstrate admirable qualities or skills useful as a council member? Experience should count only from 16 years of age onwards.



5. Cultural fit: does the candidate their opinions on the issues, their vision and ideas for the council, and their values and morals fit with INTER as a party? Is the candidate expected to be able to meet INTER’s values of transparency, democracy, inclusivity and can they be expected to be a visible and accessible member of the council and the

party? Is the candidate likely to attend INTER events and other UvA events where a council member’s presence is warranted?



This checklist of attributes will have to be designed by us only once. From here on, it will become a **routine decision** each year and involve less mental effort and costs, except for if the board considers it necessary to update or tweak the MAP slightly. We will make sure the board discussed these assessments separately, 1 by 1. Just as interviewers in structured interviews assess the candidate on all attributes separately. We are making use of the **collective wisdom of the board**. We are maximising the value of information by keeping the attributes independent of each other. All the attributes have been carefully selected to minimise redundancy. How strongly does the evidence on this attribute argue in favour, or against the candidate?

The result is that with the Mediating Assessment Protocol we have achieved a predictor the candidate’s future job performance from a flip of a coin odds in a situation with an unstructured interview and no protocols, a **63-69% strong predictor with the MAP**. A significant improvement.

The Board assesses the candidates

Ideally we would also have outside experts to weigh in on the quality of the candidates, individuals who are experienced, like Tessa or Sam, but here we are, and we shall have to make do without.

Yeva, as our Head of Policy and interviewer, should not hide any details, especially if they are contradictory. And don't sweep anything under the rug. The interviewer’s job is not to sell the recommendation. It is to represent the truth. If it is complicated, so be it. The truth often is. The interviewer should also be transparent about her level of confidence on each assessment. On what attributes does she feel in the dark? Did any deal breakers or deal makers emerge that the board should know? Lastly, the candidates should be kept **anonymous**: they could be referred to as candidate A, B, C etc.

We use the **estimate-talk-estimate method**. This method combines the advantages of deliberation, and of averaging independent opinions, on each of our 5 attributes. The interviewer, the Head of Policy will briefly summarise the results for each attribute.

Then, Kate should direct the board to use a voting link to give their own ranking of the candidates, secretly and without any discussion! The voting tool we use is from *StrawPoll*:

https://strawpoll.com/create/?gclid=Cj0KCQjw8qmhBhCIARIsANAtbocIXLvaAGdbtuSFq5fTmV3E1S477NQxMTATuhgrhJEwunphmHzC2dIaAgKtEALw_wcB

The voting tool creator will then display the distribution of ratings on the screen without identifying who made what rating. We have now an immediate read on each INTER board

member’s independent opinion before we start the discussion. **We have reduced the danger of social influence and information cascades.**

We will find in this ranking that for some candidates, the ratings of their 5 attributes point in opposing directions. It won't be an all-systems-go picture. That is good: reality is not as coherent, as others sometimes make it out to be. These discrepancies between the attributes will raise questions or trigger discussions in the board. This is what we need, good debate that will enhance our final ranking as a board. The diverse results that may be produced by the Mediating Assessment Protocol for each of the 5 attributes will not make our decision easier, for sure, but they will make it better. A consequence will also be that we will have more overt disagreements, but we will also be better at constructively resolving them, because few things are worse than problems or issues that are left unaddressed or ignored.

It is now Kate’s job to manage the discussion to spend time on the attributes where the board member rankings were most differing, the most controversy. She should make sure that members on each side of the divide speak up. Encourage everyone to express their viewpoints with facts and arguments, but also with nuance and humility. In this discussion of our rankings one of the attributes, or alternatively, of the whole MAP as a whole, is also where the charisma or electability of a candidate may be brought up.

After discussion has been fruitful and exhausted, Kate will direct the board members to vote again on the assessment or the whole MAP. We should find there will be more convergence than initially. We ought to do this process of **estimate-talk-estimate** for each assessment ideally.

Now, the interviewer or the voting tool creator will show on the whiteboard for each attribute the aggregated average of the rankings the board gave of the candidates on each attribute, or of the whole MAP.

Finally, we are arriving at our final decision and this is where we introduce the board’s last intuitive judgement after we have safely anchored our decision on all these fact-based, thoroughly discussed ratings. The board votes on adopting the ranking advice for the council in question.

The result of all of the work we have done here today is that we have implemented several decision hygiene techniques. We have **sequenced information**, we have **structured the decision into independent assessments**, we have used a **common frame of reference**, grounded in **the outside view** and **aggregated the independent judgments** of multiple board members. We have altered the decision process to reduce noise as much as possible. Decision hygiene, as we have done it here today will not prevent all mistakes. It will not make every decision we make, every ranking of the candidates flawless. But like hand washing prevents diseases, it addresses an invisible yet pervasive and damaging problem: the problem of noise. Wherever there is judgement, there is noise. And we have executed decision hygiene as a tool to reduce it.

Our final product is a report detailing the INTER board’s advice for the ranking of the candidates for the CSR elections and all FSR elections in which we are participating, to be submitted to our general INTER members ahead of the candidate list event. This ranking advice will inform and advise them on their own ranking of the INTER candidates running

for student council, who may rank and vote freely according to their right under INTER statutes Article 13 paragraph 4.

Appendix A

On the concern that the whole Mediating Assessment Protocol and similar noise reduction strategies are in the end entirely subjective

In *Noise – a flaw in human judgment*, the authors rather describe the prevailing problem of noise (unwanted variability in judgment) as one full of subjectivity. A problem where too much discretion and flexibility, and an absence of statistical tools and well-designed rules has led to a lot of arbitrariness and unfairness in a plethora of fields: forensic science, medicine, recruiting, HR, company mergers and acquisitions, and even our criminal justice system: when innocent suspects are incorrectly identified as the perpetrator, when a leg is incorrectly consigned to amputation, or someone gets hired based on their affinity with a football team shared with the interviewer, it is not hard to see the tragic cost of discretion. Not all judges in the USA agree, by far, on the correct punishment for criminals, but what they all agree on is that it is grossly unjust that a heroin dealer may be sentenced for 1, to up to 10 years in prison, or that a bank robber may be sentenced for between 5, up to 18 years in prison, depending solely on what judge is assigned to their case. This is the result of too much discretion.

The Mediating Assessment Protocol, as the authors of the book propose it then, reduces the amount of noise in decision making, by enticing us to deconstruct complex judgements (such as ranking candidates) into smaller attributes to assess, by aggregating independent and separate opinions, and lastly by delaying holistic judgement. We keep the subjective intuition and judgement of the board, but delay it and postpone it until the final part of our process. That way, our intuition is unleashed on facts, instead of the facts being distorted by our premature intuition.

On the risk of noise reducing strategies to allow for evasion of the rules

This is a legitimate concern, the passage from *Noise – a flaw in human judgment* dealing with this concern is paraphrased below, along with my own additions to apply the passage to our situation at INTER.

Noise – a flaw in human judgment pp. 344-345: In a noisy system (one with a lot of discretion, little rules or protocols), judges of all kinds can adapt as the situation requires. This can be handy if there are unexpected developments. By eliminating the power of adaptation, some noise reduction strategies can have the unintended consequence of giving people an incentive to game(abuse) the system. Noise might prevent gaming in this regard.

The tax code is an example of this. Here, there's a debate amongst experts between clear rules, that are fair in treating taxpayers, and on the other hand: deliberate vagueness to

combat those that would seek to abuse the tax code. When we cannot easily design rules that combat conduct that ought to be prohibited, we have a distinctive reason to tolerate noise.

The big question for INTER is: How much evasion there would be and how much noise there would be if we did or did not implement the Meditating Assessment Protocol? If there is only a little of evasion and a lot of noise to be prevented, then we are better off with approaches that reduce noise for INTER’s candidate selection. There is not likely to be a lot of gaming of the rules in our case: our candidates so far are not aware of the exact attributes they are being assessed on, and there is no real monetary incentive or other motivation for us as the board to game the system. We must also remember that our whole process of actively devising attributes that we think a candidate should duly possess, and looking at data first, *then* judging last, will significantly improve our decision making and reduce noise. We are better off with a noise reduction strategy such as the Mediating Assessment Protocol.

Appendix B

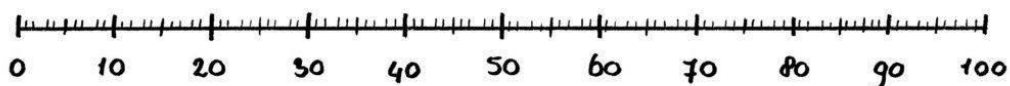
The Mediating Assessment Protocol form for the interviewer

1. Experience as a member of a student council, or student representative body, or similar, at secondary school or higher education or similar, measured in the number of

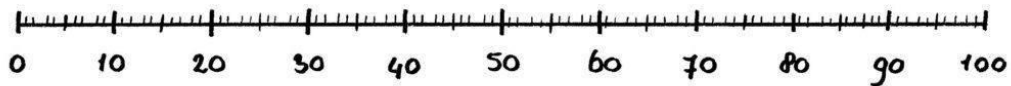


years spent in such positions and the intensity of the office. Experience should count only from 16 years of age onwards.

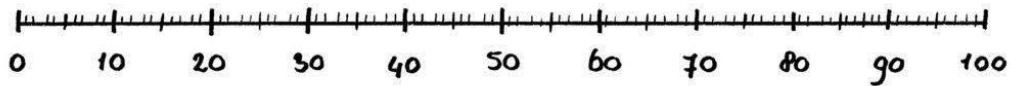
2. Motivation of the candidate to become a council member. What reasons have led the candidate to decide to run for council member, at this time, for INTER and not another party, for that particular council and why?



3. Vision of the candidate and their concrete ideas to strive to realise during their tenure as a council member for the applied for specific council. What is the vision and what are the ideas? How expansive and thought out are they? Are they feasible within the limits and competences of the specific council? Has the candidate put any thought into that feasibility? What distinguishes this candidate from other candidates?



4. CV strength: to what extent has the candidate shown to engage in extracurricular exploits besides representative functions, that demonstrate admirable qualities or skills useful as a council member? Experience should count only from 16 years of age onwards.



5. Cultural fit: does the candidate their opinions on the issues, their vision and ideas for the council, and their values and morals fit with INTER as a party? Is the candidate expected to be able to meet INTER’s values of transparency, democracy, inclusivity and can they be expected to be a visible and accessible member of the council and the party? Is the candidate likely to attend INTER events and other UvA events where a council member’s presence is warranted?

