

## RESEARCH ARTICLE

---

# Designing Assessment Reports: How Proper Score Descriptions Can Improve Selection Decisions

---

ROXANA M. SPÎNU

*Department of Psychology and Cognitive Science, University of Bucharest, Romania*

DRAGOȘ ILIESCU\*

*Department of Psychology and Cognitive Science, University of Bucharest, Romania*

### Abstract

This paper explores, through a quasi-experimental design, different ways of building employee assessment reports to determine which elements would lead to more efficient selection decisions. Two types of reports were built: (a) reports containing descriptions of the minimum and maximum scores (i.e., 1 vs. 10), and (b) reports containing the description of the exact score recorded by the participant (i.e., any score on a scale of 1 to 10). The efficiency of these reports was evaluated in two scenarios: a difficult-decision scenario and an easy-decision scenario. A total of 269 hiring managers participated in a simulated selection decision setting. They were asked to choose the best candidate for a specific position, based on a job description and two personality profiles for two fictitious candidates. They were also asked about their perceived levels of comprehension and satisfaction with the report. The model was statistically significant,  $\chi^2(3) = 110$ ,  $p < .001$ ,  $R^2N = .461$ . In both the difficult and the easy scenarios, reports containing descriptions of the minimum and maximum scores were more efficient than reports containing descriptions of the exact scores recorded by the participants, and they also led to higher levels of perceived comprehension and satisfaction with the report. The results were influenced by the participants' familiarity with the used personality questionnaire (the NEO PI-R). This study has both theoretical and practical implications, extending the existing organizational literature by drawing from cognitive psychology, and highlighting the critical role that assessment reports have in the process of organizational selection decisions.

### Keywords

report design, assessment report, score description, selection, hiring decisions, decision-making, employee.

The success of both commercial and non-commercial organizations depends strongly on the quality of their human resources. Therefore, appropriate selection procedures are mandatory to the organization, as well as to the well-being of its employees. Despite extensive attention given to the psychometric properties of different selection procedures, there is a lack of research regarding how the information gained from these measurements is further used in decision-making processes.

Unfortunately, the scientific literature does not offer much guidance on how to structure this information or design the assessment reports to communicate clear and actionable results to decision-makers. A few pointers in this direction can be found in the educational literature (e.g., Miller & Watkins, 2010; Wiener & Costaris, 2012), and even fewer in the I/O literature (e.g., Spînu & Iliescu, 2019; Spînu & Corbeanu, 2023), with a strong call to continue research on this topic (Mastoras et

---

\*Correspondence concerning this article should be addressed to Dragoș Iliescu Department of Psychology, University of Bucharest, Panduri Street, no. 90, Bucharest, Romania. E-mail: [dragos.iliescu@fpse.unibuc.ro](mailto:dragos.iliescu@fpse.unibuc.ro)

al., 2011). We address this issue in a quasi-experimental study by evaluating which elements in the design of assessment reports lead to more accurate selection decisions in organizations.

The way assessment results are reported has a significant impact on the hiring process, directly influencing the quality of selection decisions. While it is generally assumed that reports are error-free and they correctly communicate the intended information, reality seems to contradict this assumption. Based on the same assessment reports, different managers can make completely different decisions (Spînu & Corbeanu, 2023), and this may happen, at least in part, due to the misinterpretation of poorly communicated assessment results. In line with the notion of consequential validity (Messick, 1989; Iliescu & Greiff, 2021), we believe that the same importance that is now given to the selection

process (i.e., the method and measures used) should also be invested in the development of the consequent assessment reports. We advance that the validity of the assessment process is not sufficient in and of itself. Therefore, using good and valid tests may still lead to less-than-optimal decisions if the assessment results are not clearly communicated to hiring managers, who ultimately make the hiring decisions.

In the present study, we examine two different ways of building assessment reports: (a) reports containing descriptions of minimum and maximum possible scores (i.e., 1 vs. 10), which is one of the conventional ways of reporting results (e.g., SHL, 2002), and (b) reports containing the description of the exact score recorded by the participant (i.e., any score on a scale of 1 to 10). Examples of both types of reports could be seen in Figure 1.

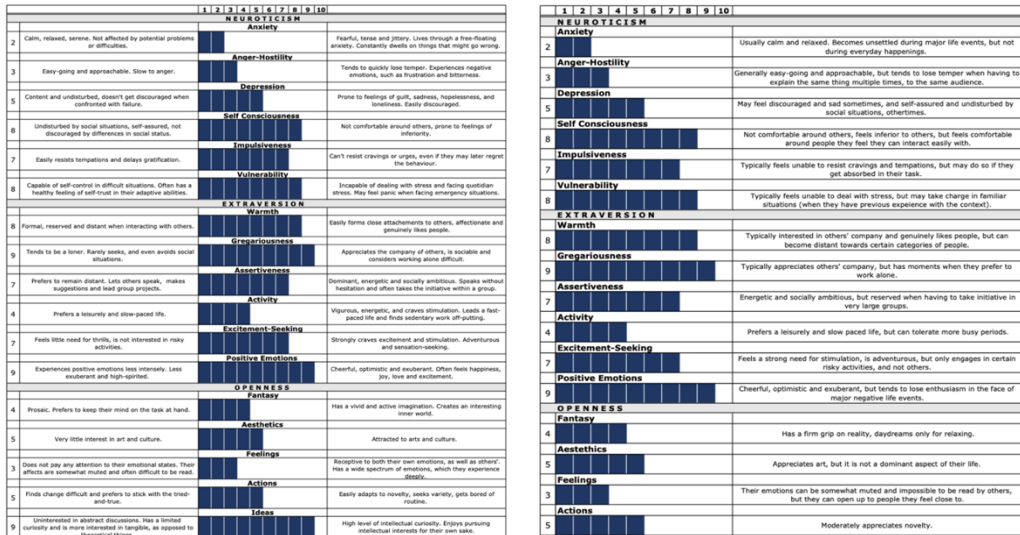


Figure 1. Examples of the Report Types Employed

Unfortunately, there is a scarcity of available literature that directly addresses this topic, and we were able to identify only two studies that have explored the issue of assessment reports so far (Spînu & Iliescu, 2019; Spînu & Corbeanu, 2023). This limited body of research poses a significant challenge

in identifying appropriate theories to serve as a basis for our arguments. Since the specific nature of our research leans more towards the domain of ergonomics and cognitive psychology, rather than traditional work psychology, we believe that drawing upon the cognitive theories, specifically the bounded

rationality, would be a pertinent starting point in our paper, and a source of arguments for the methodological choices we made throughout the study.

### **Bounded Rationality Theories**

In contrast to earlier theories of rational decision-making (Carter et al., 2007), it is now considered that the human ability to process information is governed by heuristics and biases (Kahneman & Tversky, 1979; Shah & Oppenheimer, 2008): decision-makers often encounter limitations in the cognitive resources required for making decisions, and therefore tend to rely on cognitive shortcuts to diminish the cognitive effort, especially in complex decisions or under uncertain conditions (Simon, 1990).

The earlier theories of rationality used to see decision-makers as ideal cognitive processing machines. Accordingly, when confronted with choices and decisions, individuals were expected to have access to all possible alternatives, possess accurate expectations about the future outcomes of their decisions, and use information in a systematic and logical manner (Samuelson, 1947). As such, complex algorithms were used to describe or simulate actual decision-making processes (Shah & Oppenheimer, 2008). An example of a complex algorithm is the weighted additive rule. Without getting into the mathematical details, this algorithm requires decision-makers to follow five steps: (1) identify all relevant pieces of information, (2) process all pieces of information, (3) determine the importance of each piece of information, (4) add up all the available information for each option, so as to gauge its overall value, and (5) compare all alternatives (Payne et al., 1993). However, subsequent research has shown that these complex algorithms cannot be realistically sustained due to the limited capacity of the human cognitive system (Garner, 1982). As a result, individuals tend to use specific strategies (i.e., heuristics) that help them simplify decisions and reduce the mental effort, while still obtaining satisfactory results (Tversky & Kahneman, 1974). This happens especially in conditions of uncertainty, or when the

presented information is unclear (Jasko et al., 2015; Simon, 1976). In other words, heuristics (or biases) are simple mental processes that replace the complex algorithms that would have otherwise required too many cognitive resources (Newell & Simon, 1972).

In the context of decisions based on assessment reports, we believe that solely providing behavioural explanations for the minimum and maximum scores (i.e., 1 and 10) create a challenging and uncertain environment that might determine decision-makers to resort to mental shortcuts when dealing with the exact scores obtained by the candidates. We argue that, without a clear linkage to performance indicators, it may be difficult to understand how an intermediate score (e.g., 4 or 7) corresponds to actual behaviors. According to the attenuation bias, which acts “to oversimplify a decision-making process by ignoring, overlooking or excluding certain information relevant to the decision” (Kiesling & Chong, 2020, p. 43), decision-makers may unintentionally choose to ignore intermediate scores that cannot be easily translated into behaviours. This raises important concerns, as they should ideally process all the information presented in a report, and make an informed decision only after fully understanding the assessment results.

The need to define the scores falling between the two extremes was also visible in the earlier stages of the development of behaviourally anchored rating scales (BARS; Grote, 1996). The frustration of dealing with missing intermediate anchors determined researchers and practitioners to stop defining only the minimum and maximum scores, and to add descriptions for all the other scores as well. Unfortunately, this procedure was not particularly successful as job experts seemed to provide more examples of good and poor behaviours than of average behaviours, with agreement being lower for the latter (Hauenstein et al., 2010). While it is clear that extreme behaviours are more salient than average ones (Debnath et al., 2015), we believe that ceasing to define the intermediate scores is not a justifiable rationale, as it may lead to significant cognitive biases that could be avoided.

## The Present Study

The process of decision-making can be a cognitively demanding activity for individuals (Kool et al., 2010). Typically, assessment reports include only the descriptions of the minimum and maximum possible scores of a scale and miss descriptions for the specific scores obtained by the candidate. Besides the lack of evidence supporting the effectiveness of this traditional reporting approach, we believe that it actually makes information processing unnecessarily difficult.

In this paper, we evaluate the efficiency of designing assessment reports that enumerate the specific behaviours associated with the exact scores obtained by the test taker, compared to reports that define and describe only the minimum and maximum scores of the scale (pole-end scores). To this end, we have built two personality profiles corresponding to two fictitious candidates, having in mind two different scenarios, an easy-decision scenario (where the scores of the two candidates are very different apart) and a difficult-decision scenario (where the scores are very similar). We introduced these scenarios based on the premise that in difficult decisions, critical information necessary for making a correct decision would not be as readily available as in easier scenarios: while in easy scenarios almost anyone can reach the correct conclusion, more difficult scenarios require much sharper information. Therefore, in more difficult contexts, an effective design is particularly important to ensure a correct and complete understanding of the presented information.

We test these reports in a selection decision setting, and evaluate their efficiency from the perspective of decision accuracy, as well as from the perspective of participants' levels of comprehension and satisfaction with the reports. Selection accuracy holds particular significance for organizations (Kerr, 2021), as poor hiring decisions can have detrimental impact on various organizational areas (CareerBuilder, 2017). Additionally, empirical findings from educational testing domain showed that different types of reports lead to different levels of comprehension and satisfaction (e.g., Bucknavage, 2007). This is significant because it can influence the

likelihood of practical implementation of specific designs, given that individuals are more inclined to use materials they like and understand (Cowburn & Stockley, 2005; Konradt et al., 2006).

Based on the information presented above, we advance the following hypotheses:

*H1: Regardless of the report type, participants will make more accurate decisions in the easy scenario than in the difficult scenario.*

*H2: Regardless of the scenario, participants will make more accurate decisions when the report includes exact score descriptions, as opposed to reports that describe only the pole-end scores.*

*H3: In the easy scenario, both types of reports are efficient in selection decisions.*

*H4: In the difficult scenario, compared to the easy one, participants (a) will make better selection decisions, and will report higher levels of (b) comprehension and (c) satisfaction when the report includes exact score descriptions, as opposed to reports that describe only the pole-end scores.*

## Method

### Participants

We targeted participants who held a managerial position. Invitations to participate were sent via emails, as well as posted on LinkedIn and Facebook groups. Participation was voluntary and anonymous. Participants had the chance to receive, through a random draw, a \$25 Amazon gift card.

We collected data from 280 participants, however after removing the duplicates ( $n = 2$ ), the missing data ( $n = 3$ ), and the participants who did not meet the inclusion criteria ( $n = 6$ ), the final sample consisted of 269 participants. Of the participants, 68% were women ( $n = 183$ ), and 32% were men ( $n = 86$ ), with an average age of 37 years ( $SD = 8.92$ ), ranging from 19 to 61 years old. Participants had an average experience in a managerial position of 6 years ( $SD = 5.16$ ), ranging from 0 to 25 years, and they had made between 0 and 100 hiring decisions in the last 12 months ( $M = 11.71$ ,  $SD = 17.18$ ). Unfortunately, due to insufficient data, we are unable to provide a comprehensive overview of the departments

in which they were employed. Only 15% of participants provided information about their department or role, which included internal audit ( $n = 6$ ), consulting ( $n = 3$ ), financial ( $n = 1$ ), HR ( $n = 23$ ), management ( $n = 1$ ), payroll ( $n = 1$ ), and sales ( $n = 3$ ).

A 2x2 between-subject design was used for this study, with 4 experimental groups. Although we did not randomize participants, we did not observe any differences between the groups regarding participants' age, managerial experience, and their level of experience in making hiring decisions ( $F(9, 640) = 1.25, p = .257$ ). However, we did observe a significant gender difference ( $\chi^2(3) = 18.06, p = <.001$ ) – there were more women than men in the “easy scenario – pole-end score descriptions” and in the “difficult scenario – pole-end score descriptions” conditions.

## **Procedure**

### **Overview**

We simulated a selection decision setting where the participants were first presented with a job description for the role of Market Research Analyst (MRA). After studying the document, they received two personality profiles corresponding to two fictitious candidates, and were asked to choose which candidate they would hire for the MRA position. They were allowed as much time as they needed to make the decision. Additionally, the participants completed a short questionnaire regarding their level of comprehension and satisfaction with the reports.

The study was conducted online, using a proprietary html code. Two variables were manipulated: the report type (i.e., reports containing pole-end score descriptions vs exact score descriptions) and the decision scenario (i.e., participants were confronted with either an easy or a difficult decision). Detailed accounts of the job description, the personality profiles, and the manipulated variables follow below.

### **Job Description**

We used O\*NET indicators to build the job description for the role of Market Research Analyst. The description consisted of one and a half pages of bullet-lists, divided into three categories: responsibilities (e.g., “Measure the effectiveness of marketing programs and strategies”), requirements (e.g., “Background in business administration or social sciences”), and soft skills (e.g., “Persuasion skills, and an ability to convince and motivate others”). The complete job description can be found here: [https://osf.io/eadx6/?view\\_only=343e296dd7644cc8a1ff765f0f2827dd](https://osf.io/eadx6/?view_only=343e296dd7644cc8a1ff765f0f2827dd).

### **The personality profiles**

The personality profiles were developed based on the job description presented above. We developed two pairs of profiles, one pair for the easy scenario, and one pair for the difficult scenario. However, in both scenarios, we made sure that one candidate (Candidate 1) would be a better fit for the job than the other. We used a similar structure as the NEO PI-R questionnaire (Costa & McCrae, 1992), with five factors and six facets each. The scores were expressed in STEN scores ranging from 1 to 10.

To ensure that one candidate is indeed a better fit than the other, we conducted a pilot test with five experienced psychologists who had an average of 11 years of experience ( $SD = 3.28$ ) in HR consultancy. They were asked to choose the best candidate for the job based on our job description. All five chose the Candidate 1 in both scenarios.

### **The decision scenario as a manipulated variable**

The difference between the easy and the difficult scenarios relies on how similar or different apart the two profiles are. For the easy scenario, the difference between the two candidates were glaring, as the profiles were built at opposite extremes. Candidate 1 had high scores on desirable traits, such as extraversion or conscientiousness, and low scores on undesirable traits, such as

neuroticism. In contrast, Candidate 2 displayed the exact opposite pattern.

For the difficult scenario, the two profiles had very similar scores. However, upon careful examination, it became apparent that one profile was slightly better suited for the

job than the other. Compared to the easy scenario, the task required more in-depth analysis and intense cognitive processing. The scores for all four personality profiles may be seen in Table 1.

Table 1. *The Personality Profiles*

	Difficult scenario		Easy scenario	
	Candidate 1	Candidate 2	Candidate 1	Candidate 2
Neuroticism	33	27	12	52
Extraversion	44	34	46	23
Openness	29	32	40	26
Agreeableness	32	32	41	26
Conscientiousness	47	46	49	18

*Note.* In both the difficult and the easy scenario, Candidate 1 is a better fit for the job.

### ***The report type as a manipulated variable***

The report types were designed having two perspectives in mind: (1) reports that include behavioural descriptions for the exact scores obtained (i.e., any score on a scale of 1 to 10), and (2) reports that include behavioural descriptions of only the pole-end scores of the scale (i.e., 1 vs. 10).

The reports based on pole-end score descriptions included a bar chart and the corresponding behavioural descriptions on the right and left sides. Each of these reports covered one and a half pages. The score descriptions were generated using the NEO PI-R standard report that already provides behavioural descriptions of minimum and maximum possible scores. Examples of descriptions for the Impulsivity scale are as follows: at low scores - "Easily resists temptations and delays gratification", and at high scores - "Can't resist cravings or urges, even if they may later regret the behaviour". When completing the task, participants were instructed that the respective descriptions correspond to the minimum and maximum possible scores.

The reports based on exact score descriptions included a bar chart and then the corresponding behavioural description for that

specific score. The exact score descriptions were generated using several resources, including the NEO Job Profiler (Costa et al., 1995), the NEO PI-R standard report (Costa & McCrae, 1992), and the guidelines offered by Lord (2007). We used professional judgment to define each score on a scale of 1 to 10, with special attention given to balancing the need to provide sufficient details and the need to remain within a one-dimensional scale. The scores were defined in everyday language, without using professional jargon. An example for a score of 7 on the Impulsivity scale is as follows: "Typically feels unable to resist cravings and temptations but may do so if they get absorbed in their task". The instructions for participants also indicated that these descriptions correspond to the exact scores obtained by the participants. All the reports can be found here: [https://osf.io/fc24q/?view\\_only=05b61ee773364f2693c93b91d201f462](https://osf.io/fc24q/?view_only=05b61ee773364f2693c93b91d201f462).

### **Measures**

We analyzed the efficiency of the report types employed using three indicators: the decision accuracy, as well as the self-reported levels of comprehension and satisfaction with the report. The decision accuracy was measured on a dichotomous scale, and employed either

a correct or an incorrect answer. The participants' comprehension level was analyzed using a questionnaire adapted from Bucknavage (2007). The questionnaire consisted in seven items measured on a 1-5 Likert scale (from strongly disagree to strongly agree), and had an internal

consistency of .83. Similarly, the level of participants' satisfaction was measured using a questionnaire adapted from Pelco et al. (2009). It comprised six items measured on a 1-5 Likert scale, and had an internal consistency of .77. The comprehension and satisfaction items may be seen in table 2.

Table 2. *The Comprehension and Satisfaction Questionnaires Items*

Comprehension Questionnaire	
1.	These reports were easy to read.
2.	I easily understood the information that was presented.
3.	The writing style was too confusing for me to understand (R*).
4.	I have a clear picture of the candidates' strengths and weaknesses.
5.	The reports were well organized and coherent.
6.	Reading these reports was difficult for me (R*).
7.	I have an overall good understanding of the candidates.
Satisfaction Questionnaire	
1.	I am generally satisfied with these reports.
2.	The reports included all the information needed for me to make a decision.
3.	The information presented in the reports made sense to me.
4.	The reports relied too much on technical terms (R*).
5.	The reports presented information in a logical format.
6.	I learned a great deal about the candidates from these reports.

Note. \*R = reversed items.

### Statistical Approach

A logistic regression was conducted using the jamovi “GAMLj” module (Gallucci, 2019) to estimate the relationship between the decision scenario and the report type, and the accuracy of the selection decision. The level of participants' comprehension and satisfaction with the reports was analyzed using a multivariate multiple regression, with the help of jamovi “PATHj” module (Gallucci, 2021). Additionally, a TOST test (Caldwell, 2022) was conducted specifically for the easy scenario, in order to evaluate the level of equivalence between the two report types.

### Results

A binomial logistic regression was performed to analyse the effects of the report type and the decision scenario on the likelihood that participants made a correct selection decision. The model was statistically significant,  $\chi^2(3) = 110, p < .001$ , and accounted for 46.1% (Nagelkerke  $R^2$ ) of the variance in selection decisions. The model correctly classified 81.8% of cases, with a sensitivity index of 89% and a specificity of 69.1%.

Only the scenario and the report type were statistically significant, but not the interaction between them. The data supports our first hypothesis (*H1*), i.e., participants were more likely to make correct selection decisions in the easy scenario than in the difficult scenario.

However, contrary to our expectations (*H2*), it seemed that, regardless of the scenario, pole-end score descriptions were more efficient than exact score descriptions in enhancing accurate decisions. The odds of making a

correct selection decision are .2 times lower when presented with exact rather than pole-end score descriptions. The results are shown in Table 3.

Table 3. *The Logistic Regression Model*

	<i>B</i>	95% CI	<i>SE</i>	<i>df</i>	<i>p</i>	Odds Ratio	95% CI for Odds Ratio
Constant	.75	[.43; 1.09]	.16	1	.001*	2.13	[1.54; 3.00]
Predictors							
Scenario	2.03	[1.38; 2.71]	.33	1	.001*	7.66	[3.99; 15.07]
Report type	-1.61	[-.95; .20]	.33	1	.001*	0.20	[2.60; 9.81]
Scenario * Report type	.77	[2.10; 2.16]	.67	1	.251	2.16	[.12; 1.73]

Note. Baseline categories were the easy scenario and the exact score descriptions. \**p* < 0.05

There was no significant interaction between decision scenario and the report type in regards to the decision accuracy (*H4a*). Therefore, the probability of making a correct decision based that one or the other type of report does not depend on the difficulty of the decision. In fact, in the easy scenario, although we hypothesized that both report types would be efficient in selection decisions (*H3*), the data suggests that they are not actually equivalent (*p* = .747) - reports containing pole-end score descriptions led to better decisions

than exact score descriptions even in the easy scenario.

Overall, 87.2% of participants made a correct selection decision in the easy scenario, compared to only 35% in the difficult scenario. Considering both scenarios, 44.9% of participants made a correct decision when presented with pole-end score descriptions, compared to only 18.9% when presented with exact score descriptions. The probabilities of making a correct decision in each group may be seen in Table 4, and Figure 2 respectively.

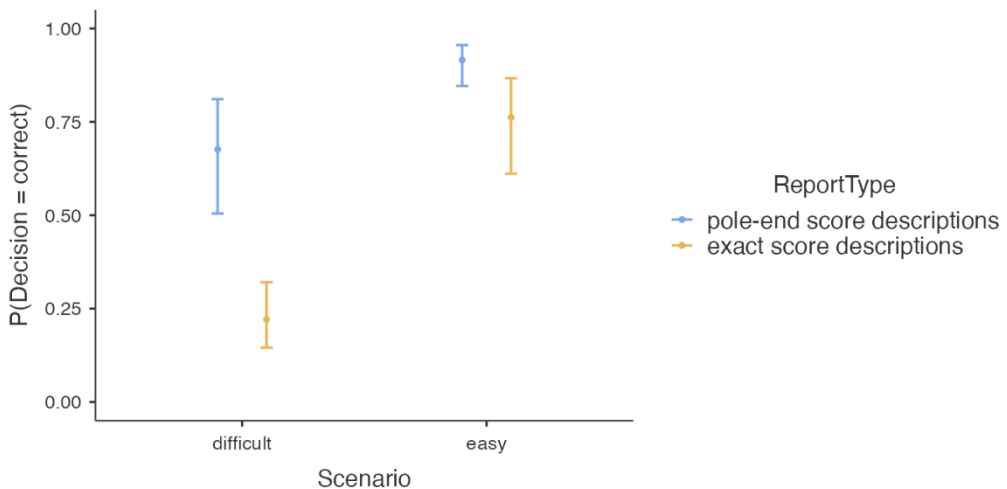


Figure 2. A Visual Representation of the Probabilities of Making a Correct Decision

Table 4. The Probabilities of Making a Correct Decision

Scenario	Report type	Probability	SE	95% CI
Difficult	Pole-end scores	.67	.08	[.50; .81]
	Exact scores	.22	.04	[.14; .32]
Easy	Pole-end scores	.91	.02	[.84; .95]
	Exact scores	.76	.06	[.61; .86]

We additionally controlled for the potentially confounding effects of the participants’ response times, as well as their familiarity with the NEO-PI-R questionnaire, to thoroughly understand our results. The data suggests that there is no significant effect of how much time participants spent in the task,  $\chi^2(1) = 3.31, p = .069$ . However, there is a significant interaction between the report type,

the decision scenario and the familiarity with the questionnaire ( $p = .027$ ). As shown in Figure 3, in the difficult scenario, participants who were more familiar with the questionnaire made better decisions when provided with pole-end score descriptions. Conversely, for exact score descriptions, more familiarity with the questionnaire led to worse decisions.

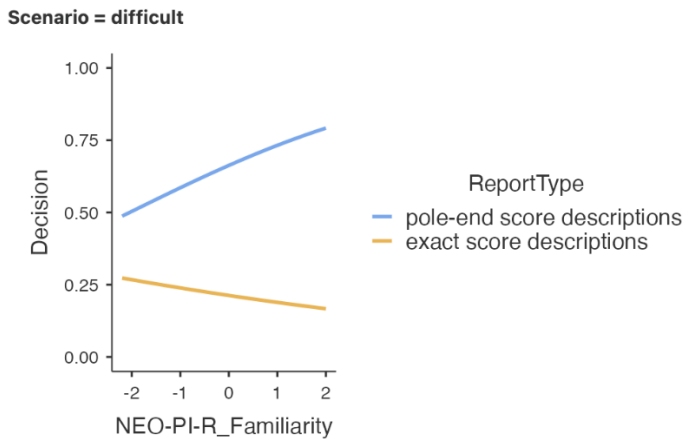


Figure 3. The Probability of Making a Correct Decision When Familiarized with the Questionnaire

To analyze participants’ levels of self-reported comprehension and satisfaction with the report, we conducted a multivariate multiple regression. The model was statistically significant,  $\chi^2(7) = 190, p = .001, CFI = 1.000$ . Contrary to our expectations (*H4b* and *H4c*), the results suggest that participants had a higher level of comprehension when presented with pole-end score descriptions, rather than exact score descriptions, either alone ( $\beta = -.21, p = .008$ ) or in interaction with the scenario ( $\beta = .15,$

$p = .007$ ). Similarly, reports containing pole-end score descriptions led to higher levels of satisfaction compared to reports containing exact score descriptions ( $\beta = -.24, p = .002$ ). However, the interaction between scenario and report type does not significantly influence participants’ satisfaction levels ( $\beta = .08, p = .157$ ).

A post-hoc G\*Power analysis suggested that our design was slightly underpowered (0.60). In order to detect a significant effect for

the exact score descriptions, we would have needed a sample size of 380 people.

## Discussion

### Summary of Findings

The current study explored different ways of designing employee assessment reports to determine which elements would lead to more efficient selection decisions. Two types of reports were evaluated, in two different scenarios.

As hypothesised, participants made more accurate selection decisions in the easy scenario compared to the difficult scenario. However, contrary to our expectations, reports containing pole-end score descriptions were more effective in terms of the accuracy of the selection decision than those containing exact score descriptions. This was true even in the easy scenario, where the most suitable candidate for the job should have been glaring in both reports. Furthermore, participants reported a higher level of perceived comprehension and satisfaction when presented with pole-end score descriptions in both the easy and the difficult scenario. There are several possible explanations for these results.

Firstly, data suggests that familiarity with the NEO-PI-R questionnaire influenced the likelihood of making a correct decision. Participants that were more familiarized with the questionnaire tended to make worse decisions when presented with exact score descriptions rather than pole-end score descriptions. It may be that participants were accustomed to a specific design of the questionnaire, and the fact it was differently structured in our study may have actually made the information processing much more difficult. It may also be that in order to make sense of the dimensionality of a personality trait, managers need the two extreme anchor points of the scale; significance of the points that lie in between the two extremes is then deducted easily – while exposure to only the exact description does not enable the decision maker to actually make sense of the dimensionality of the targeted trait. The questionnaire familiarity and the previous experience in working with it might also

explain the higher reported levels of comprehension and satisfaction with the report when presented with pole-end score descriptions compared to exact score descriptions.

Secondly, according to the picture superiority effect (Stenberg, 2006), people seem to focus more on the visual elements of a document rather than on the text (Schnotz et al., 1993). Thus, it is possible that participants did not read the score descriptions carefully enough. However, a similar eye-tracking study on assessment reports, evaluating how the visual design of the reports can influence information integration from graphics and texts, showed that neither the number of fixations on the graphic or the text, nor the number of integrations between the graphic and the text were significant predictors of the accuracy of the selection decision (Spînu & Iliescu, 2019). It remains unclear how people use information from texts and graphics in decision-making processes, and what elements actually lead to more efficient decisions. We suggest including more eye-tracking designs to better understand this matter.

### Theoretical and Practical Implications

From a theoretical standpoint, the main contribution of our study lies in its interdisciplinary nature, extending the applicability of a cognitive theory into the field of organizational psychology. The study challenges researchers to further investigate how pole-end and exact score descriptions fit with bounded rationality theories and cognitive biases, and which of them actually raises more uncertainty for decision-makers: describing only the extreme ends of a scale without a clear linkage to actual behaviours for intermediate scores, or describing the exact score obtained by the respondent without offering a framework of reference.

From a practical perspective, our study holds significant relevance within the realm of organizational realities. It highlights the importance of effectively communicating results to decision-makers to ensure they are actionable, while also focusing on the consequences of the assessment process. The

correct understanding of the results and the decisions that are made subsequently depends strongly on the quality of the assessment report. Our study shows that current practices in designing assessment reports (i.e., designing reports that describe the minimum and maximum possible scores) are somewhat effective. However, the fact that a non-trivial volume of our sample has made an incorrect decision raises concerns, and suggests that other factors may have influenced the decision.

### Limitations and Implications for Future Research

The main limitation of this study is the use of a standard and well-known assessment tool, the NEO-PI-R questionnaire, as our data has shown that familiarity with the test significantly influenced the outcomes of our design. We recommend that future studies use less popular questionnaires. Another limitation is the lack of randomization in allocating participants to groups. This resulted in significant differences between the groups, that could have been avoided with proper randomization. We also believe that measures that are specifically designed for a work context would yield a higher impact on research results, as they are closer to what actually happens in practice. Furthermore, future studies may include some other types of tests, such as cognitive abilities tests. One final recommendation for future studies would be to include eye-tracking methodologies in their design. This will contribute to a better understanding of how decision-makers acquire and use information in assessment reports.

To sum up, our study offers valuable insights into the realm of assessment reports and contribute to the existing body of knowledge in this area. Although assessment reports are frequently utilized in selection settings, their practical efficacy remains relatively unexplored. Despite its limitations, our research confirms that one of the traditional ways of designing assessment reports is actually effective in selection setting. Nevertheless, it is worth noting that only 87.2% of participants made a correct

decision even in the easy scenario. This suggests the presence of additional factors that may have influenced participants' decisions. Considering this, we advocate for further research and a more in-depth investigation into this subject matter.

### References

- Bucknavage, L. B. (2007). *Psychoeducational reports: Impact of jargon and report length on parent and teacher recall and preference*. The Pennsylvania State University.
- Caldwell, A. R. (2022). *TOSTER R package 0.5.51*. CRAN.
- CareerBuilder, 2017. [The costs and consequences of poor hiring – a survey]. [press.careerbuilder.com](https://press.careerbuilder.com)
- Carter, C. R., Kaufmann, L., & Michel, A. (2007). Behavioural supply management: A taxonomy of judgment and decision-making biases. *International Journal of Physical Distribution & Logistics Management*, 37(8), 631-669. <https://doi.org/10.1108/09600030710825694>
- Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NIT0 Five-Factor Inventory (NEO-FFI) professional manual*. Psychological Assessment Resources.
- Costa Jr, P. T., McCrae, R. R., & Kay, G. G. (1995). Persons, places, and personality: Career assessment using the Revised NEO Personality Inventory. *Journal of Career Assessment*, 3(2), 123-139.
- Cowburn, G., & Stockley, L. (2005). Consumer understanding and use of nutrition labelling: a systematic review. *Public Health Nutrition*, 8(1), 21-28. <http://doi.org/10.1079/PHN2004666>
- Debnath, S. C., Lee, B. B., & Tandon, S. (2015). Fifty years and going strong: What makes Behaviorally Anchored Rating Scales so perennial as an appraisal method?. *International Journal of Business and Social Science*, 6(2), 16-25.
- Gallucci, M. (2019). *GAMLj: General analyses for linear models*. [jamovi module]. Retrieved from <https://gamlj.github.io/>
- Gallucci, M. (2021). *PATHj: jamovi Path Analysis*. [jamovi module]. Retrieved from <https://pathj.github.io/>
- Garner, C. A. (1982). Uncertainty, human judgment, and economic decisions. *Journal of Post Keynesian Economics*, 4(3), 413-424. <https://doi.org/10.1080/01603477.1982.11489303>
- Grote, D. (1996). *The complete guide to performance appraisal*. American Management Association.
- Hauenstein, N. M., Brown, R. D., & Sinclair, A. L. (2010). BARS and those mysterious, missing middle anchors. *Journal of Business and Psychology*, 25(4), 663-672. <https://doi.org/10.1007/s10869-010-9180-7>
- Iliescu, D., & Greiff, S. (2021). On consequential validity [Editorial]. *European Journal of Psychological Assessment*, 37(3), 163-166. <https://doi.org/10.1027/1015-5759/a000664>
- Jaško, K., Czernatowicz-Kukuczka, A., Kossowska, M., & Czarna, A. Z. (2015). Individual differences in response to uncertainty and decision making: The role

- of behavioral inhibition system and need for closure. *Motivation and Emotion*, 39, 541-552. <https://doi.org/10.1007/s11031-015-9478-x>
- Kerr, D. L. (2021). Guide to Hiring Success. *Affintus*.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, 47(2), 263-291.
- Kiesling, J. K., & Chong, D. M. (2020). *Examining the role of decision biases in shaping acquisition decisions within defense acquisition program*. Naval Postgraduate School.
- Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, 139(4), 665-682. <http://doi.org/10.1037/a0020198>
- Konradt, U., Christophersen, T., & Schaeffer-Kuelz, U. (2006). Predicting user satisfaction, strain and system usage of employee self-services. *International Journal of Human-Computer Studies*, 64(11), 1141-1153. <https://doi.org/10.1016/j.ijhcs.2006.07.001>
- Lord W. (2007). *NEO PI-R: A guide to interpretation and feedback in a work context*. Hogrefe.
- Mastoras, S. M., Climie, E. A., McCrimmon, A. W., & Schwean, V. L. (2011). A CLEAR approach to report writing: A framework for improving the efficacy of psychoeducational reports. *Canadian Journal of School Psychology*, 26(2), 127-147. <https://doi.org/10.1177/08295735114097>
- Messick, S. (1989). Validity. In R. L. Linn (Eds.), *Educational measurement* (3rd ed., pp. 13-103). ACE & NCME.
- Miller, J. A., & Watkins, M. W. (2010). The use of graphs to communicate psychoeducational test results to parents. *Journal of Applied School Psychology*, 26(1), 1-16. <https://doi.org/10.1080/15377900903175911>
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). The use of multiple strategies in judgment and choice. In N. J. Castellan, Jr. (Eds.), *Individual and group decision making: Current issues* (pp. 19-39). Lawrence Erlbaum Associates, Inc.
- Pelco, L. E., Ward, S. B., Coleman, L., & Young, J. (2009). Teacher ratings of three psychological report styles. *Training and Education in Professional Psychology*, 3(1), 19-27. <https://doi.org/10.1037/1931-3918.3.1.19>
- Samuelson, P. A. (1947). *Foundations of Economic Analysis*. Harvard University Press.
- Schnotz, W., Picard, E., & Hron, A. (1993). How do successful and unsuccessful learners use texts and graphics?. *Learning and Instruction*, 3(3), 181-199. [https://doi.org/10.1016/0959-4752\(93\)90003-1](https://doi.org/10.1016/0959-4752(93)90003-1)
- Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin*, 134(2), 207-222. <https://doi.org/10.1037/0033-2909.134.2.207>
- SHL (2002). *Motivation Questionnaire: Manual and user's guide*, Version 11. SHL Group Ltd. Thames Ditton.
- Simon, H.A. (1976). From substantive to procedural rationality. In T. J. Kastelein, S. K. Kuipers, W. A. Nijenhuis & G. R. Wagenaar (Eds.) *25 years of economic theory* (pp. 65-66). Springer. [https://doi.org/10.1007/978-1-4613-4367-7\\_6](https://doi.org/10.1007/978-1-4613-4367-7_6)
- Simon, H.A. (1990). Bounded Rationality. In J. Eatwell, M. Milgate & P. Newman (Eds.) *Utility and probability* (pp. 15-18). *The New Palgrave*. [https://doi.org/10.1007/978-1-349-20568-4\\_5](https://doi.org/10.1007/978-1-349-20568-4_5)
- Spînu, R., & Iliescu, D. (2019). Psychological assessment reports in selection decisions: The role of spatial contiguity principle. *Psihologia Resursei Umane*, 17(2), 12-41. <https://doi.org/10.24837/pru.v17i2.290>
- Spînu, R. & Corbeau, A. (2023). Toward a better future for assessment reports. *European Journal of Psychological Assessment*. <https://doi.org/10.1027/1015-5759/a000768>
- Stenberg, G. (2006). Conceptual and perceptual factors in the picture superiority effect. *European Journal of Cognitive Psychology*, 18(6), 813-847. <https://doi.org/10.1080/09541440500412361>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, 185(4157), 1124-1131. <http://doi.org/10.1126/science.185.4157.112>
- Wiener, J., & Costaris, L. (2012). Teaching psychological report writing: Content and process. *Canadian Journal of School Psychology*, 27(2), 119-135. <https://doi.org/10.1177/082957351141848>

## Data Availability Statement

The data that support the findings of this study are openly available in OSF platform at [https://osf.io/fc24q/?view\\_only=b1a53f44a90142a69508ec8f0870cfc5](https://osf.io/fc24q/?view_only=b1a53f44a90142a69508ec8f0870cfc5).