

TEST-RETEST RELIABILITY OF THE POLITICAL ELITES AVERSIVE PERSONALITY SCALE (PEAPS)

DOI 10.21697/sp.2024.24.1.03

JÜRGEN MAIER¹, CORINNA OSCHATZ², MONA DIAN¹, INGO ZETTLER³

¹Department of Political Science, RPTU Kaiserslautern-Landau, Landau, Germany

²Amsterdam School of Communication Research (ASCoR), University of Amsterdam, Amsterdam, the Netherlands

³Department of Psychology, University of Copenhagen, Copenhagen, Denmark

ABSTRACT

Although aversive (“dark”) personality characteristics of politicians have increasingly become a focus of research, a suitable self-assessment inventory for political elites has only recently been introduced. This article examines the test-retest reliability of the Political Elites Aversive Personality Scale (PEAPS) in a unique natural setting using a panel study of candidates running for the 2021 Berlin state election *and* the 2023 Berlin repeated election ($N = 106$, 37.7% female). The average age was 47.5 (2021) and 49.5 years (2023), respectively ($SD = 14.7$ years). 14.2% of the participants ran for the Green Party (Bündnis 90/Die Grünen), 11.3% for the Liberal Party (FDP), 8.5% for the Social Democrats (SPD), 7.5% for the Alternative for Germany (AfD), 6.6% for the Christian Democrats (CDU), 4.7% for the Left Party (Die Linke), and 48.1% for smaller parties not represented in the parliament. Across various methods, we find moderate to high levels of reliability. Especially in the light of the relatively long time lag between the measurement occasions, results thus support using PEAPS for self-assessment of aversive personality of politicians. Given that PEAPS comprises six items only, it appears to be suitable for measuring aversive personality in situations in which time or cost prevent the use of more comprehensive personality measures. We also suggest that future research might aim to test the usefulness of PEAPS for other elites (e.g., corporate leaders).

KEYWORDS

aversive personality; self-assessment; political elites; test-retest reliability; Germany

INTRODUCTION

Especially since the presidency of Donald Trump, the role of personality traits in the actions of political elites has increasingly become the focus of social and behavioral science research. According to existing studies, politicians' personality correlates with their communication behavior, their electoral success and, once in office, with their political performance, including their likelihood of unethical behavior and their tendency to foster democratic backsliding (e.g., Joly et al., 2019; Nai & Maier, 2024; Watts et al., 2013). Outside the political context, it has been shown that in particular aversive (or "dark") personality traits are associated with aggressive or antisocial behavior more broadly (e.g., Du et al., 2022; Zhu & Jin, 2021). This also applies to politicians who, for example, are more likely to "go negative" on their political opponents or to use populist rhetoric (e.g., Nai & Martínez i Coma, 2019), or to "spice up" their communication with additional confrontative elements such as incivility (e.g., Nai & Maier, 2020).

Measuring the personality of political elites is anything but trivial in practice, specifically when it comes to aversive characteristics. First, available inventories aimed at assessing aversive characteristics are often too long to present them to politicians or to request assessments from others (e.g., experts, voters) on many politicians. Second, available inventories often include items that are (apparently perceived as be) too self-incriminating to be used for self-reporting by political elites, thus increasing the risks of selection bias and dropout. As a consequence, to the best of our knowledge, available inventories have typically been administered as observer reports (e.g., Nai, 2019). However, limiting the inventories to expert or voter ratings restricts research to the study of top politicians, because it is difficult to provide observer-ratings of the lesser known politicians. Moreover, assessments of experts and voters are likely to be affected by media reports, because these raters usually do not know the politicians personally, at least not in depth (for a critical review of expert assessments of politicians' personality see, e.g., Nai & Maier, 2024).

Tackling these limitations, the Political Elites Aversive Personality Scale (PEAPS) was introduced recently (Maier et al., 2023). PEAPS was developed in line with the conceptualization of the Dark Factor of Personality (D) (e.g., Moshagen et al., 2018, 2020). D was conceptualized based on the idea as well as increasing empirical evidence (see, e.g., McLarnon & Tarraf, 2017; Schreiber & Marcus, 2020) that there is a common core of all aversive traits, which comprises all aversive personality characteristics. D is sought to represent this common core, defined as "the general tendency to maximize one's individual utility—disregarding, accepting, or malevolently provoking disutility for others—, accompanied by beliefs that serve as justifications" (Moshagen et al., 2018, p. 657). Individual aversive traits such as the components of the Dark Triad (Paulhus & Williams, 2002), Amoralism (e.g., Knežević, 2003), Sadism (e.g., O'Meara et al., 2011), or Spitefulness (e.g., Marcus et al., 2014), in turn, are considered to be flavored manifestations of D. Flavored refers to the facts that individual aversive traits (i) can represent the different aspects of D to a different degree (e.g., emphasizing maximizing one's individual utility rather than holding certain beliefs, or vice versa) and (ii) can comprise further, essentially non-aversive personality characteristics (e.g., disinhibition in the case of psychopathy; Bader et al., 2023). In line with its conceptualization, D has been found to predict a large set of behavioral and self-reported criteria representing own utility maximization at the cost of others (e.g., Hilbig et al., 2023; Zettler et al., 2022).

Importantly, similar to the "g factor" in the intelligence domain, D is conceptualized as a fluid construct expressis verbis: "the fluid nature of D implies that the indicators to assess D are ultimately interchangeable, so that any dark trait measure (with sound psychometric properties and sufficient breadth in content) is generally suited to indicate D" (Moshagen et al., 2020, p. 192). Following this conceptualization of D – and aiming to assess the essence of politician's aversive personality (not mingled with rather non-aversive personality characteristics) – thus allowed to (i) develop an ultra-short inventory that (ii) contains not too extremely worded items, (iii) suitable for politicians' self-reports of their aversive personality.

PEAPS comprises six items, originally referring to four aversive personality traits (namely, Machiavellianism, Narcissism, Psychopathy, and Spitefulness). Although first results indicate a high level of reliability and validity (Maier et al., 2023), PEAPS has still not demonstrated its test-retest reliability, i.e., the “consistency, reproducibility, and agreement among two or more measurements of the same individual” (Aldridge et al., 2016, p. 208). Showing the general stability of a trait measurement is important to foster that observed changes are due to changes in the individual and not because the measurement is inaccurate (e.g., Gnambs, 2014).

Whereas test-retest reliability is an important ingredient of trait measures scale developments, the chances of repeatedly question political elites about their aversive personality characteristics are rather low, because it is often not possible to study such a group twice. However, the repeated election in Berlin (Germany) in 2023 provided us with a unique opportunity to test the test-retest reliability of PEAPS by re-assessing the *same* politicians who ran for parliament in 2021. Apart from the fact that we are not aware of any scale than PEAPS that is particularly tailored to ask political elites to self-assess their aversive personality, we are also not aware of any study in which politicians are asked about their (basic or more specific) personality traits more than once.

METHODS

In November 2022, the Constitutional Court of Berlin decided that the 2021 state election was invalid due to several irregularities. Therefore, the election had to be repeated, which meant that voters in 2023 again decided on the candidates who had already run for election in 2021; no new candidates or parties were allowed.

Our analysis is based on post-election surveys of candidates competing in the 2021 and the 2023 Berlin state election. Data was collected in a mixed-mode design using online as well as a regular paper-and-pencil questionnaires to maximize response rates. All candidates that provided an email address in their professional contact details online during the campaign were invited to participate via an online link to our survey (2021: 56.6%; 2023: 45.5%). All candidates without online contact details were invited by mail including a paper-and-pencil questionnaire and a stamped return envelope. Candidates invited by mail were also provided with a personalized link if they preferred to answer the survey online. IRB approval was obtained prior to data collection (GESIS ethics committee, reference number 2020-6). Since our data contain sensitive, non-anonymized information from candidates, the data can be only made available to other researchers on request following data sharing agreements.

Data collection began the day after Election Day and ended two months later. All candidates, including candidates running for smaller parties, were asked to participate in the surveys. We explicitly asked candidates in the invitation letter to fill out the questionnaire themselves. From the initially 1,116 (2021) respectively 1,057 (2023) contacted candidates, 35.0% (2021) respectively 22.1% (2023) participated in the study. For our analyses, we excluded 4 (2021) respectively 2 (2023) candidates who rushed through the online survey by employing the procedure to filter out speeders (Leiner, 2019). This leaves us with 386 (2021) respectively 231 (2023) valid cases. 147 candidates have participated in 2021 *and* 2023, which means that candidates filled out the second questionnaire about 17 months after the first questionnaire. Even though optimal time span for assessing test-retest reliability seems to be controversial, the gap between the two measurement occasions of our study is rather large.

For the following analyses we only use candidates who provided valid answers for all items of the PEAPS, which leaves us with N=106 candidates. 37.7% candidates were female. Average age was 47.5 (2021) respectively 49.5 (2023) (SD: 14.7) years. Data for gender and age was based on the information of the State Returning Officer (“Landeswahlleiter”). The IRB approval covers linking candidates’ survey responses with external sources; these linking possibilities were explicitly

mentioned to the candidates in the informed consent form. 14.2% of the candidates in our sample ran for the Green Party (Bündnis 90/Die Grünen), 11.3% for the Liberal Party (FDP), 8.5% for the Social Democrats (SPD), 7.5% for the Alternative for Germany (AfD), 6.6% for the Christian Democrats (CDU), 4.7% for the Left Party (Die Linke), and 48.1% for smaller parties not represented in the parliament.

PEAPS was measured with the following six items (in parentheses: aversive trait for which the item was originally conceptualized; for the German wording of the items see OMITTED): “*There have been times when I was willing to suffer some small harm so that I could punish someone else who deserved it*” (Spitefulness), “*It’s wise to keep track of information that you can use against people later*” (Machiavellianism), “*There are things you should hide from other people to preserve your reputation*” (Machiavellianism), “*I insist on getting the respect I deserve*” (Narcissism), “*I want my rivals to fail*” (Narcissism-Rivalry-Supremacy), “*People who mess with me always regret it*” (Psychopathy). All items were measured on a 5-point scale from 1 (“do not agree at all”) to 5 (“fully agree”). The items were each combined into a total score by computing the average score across all answers. A high (low) score on the final scale reflects a high (low) level of aversive personality. The script of our data analyses is stored in the following Open Science Foundation (OSF) repository: https://osf.io/jmxt8/?view_only=bc98954c91af4c22bcc35198fb39fe80

RESULTS

Test-retest reliability can be assessed with different measures (for a discussion see, e.g., Aldridge et al., 2017). A classic indicator is Pearson’s r , showing the correlation between the two measurement occasions. With $r(105)=0.615$, $p<0.001$, the correlation between the two measurement occasions can be considered moderate with regard to the question of test-retest reliability.

Another classic method to analyze test-retest reliability is the paired t -test, which focuses on agreement. The result shows no significant difference between the candidates’ self-reported aversive personality in 2021 ($M=2.54$, $SD=0.72$) and in 2023 ($M=2.61$, $SD=0.62$), $t(105)=1.252$, $p=0.213$, which indicates a high level of agreement.

In addition to r and paired t -test, the Bland-Altman plot has been recommended to analyze test-retest reliability. Following this method, we have plotted the average score of the 2021 and 2023 PEAPS measurements ($M=2.57$, $SD=0.60$) against the difference between the 2023 and the 2021 PEAPS measurements ($M=0.07$, $SD=0.59$). Figure 1 indicates that the large majority (i.e., >90 %) of the data points fall within the Limits of Agreement (i.e., the 95% CI of the mean difference). In addition, candidates are randomly scattered around the mean difference, suggesting “that there is little systematic bias between the measurements and no obvious data heterogeneity” (Aldridge et al., 2017, p. 214). In a nutshell, the Bland-Altman plot suggests an acceptable level of test-retest reliability.

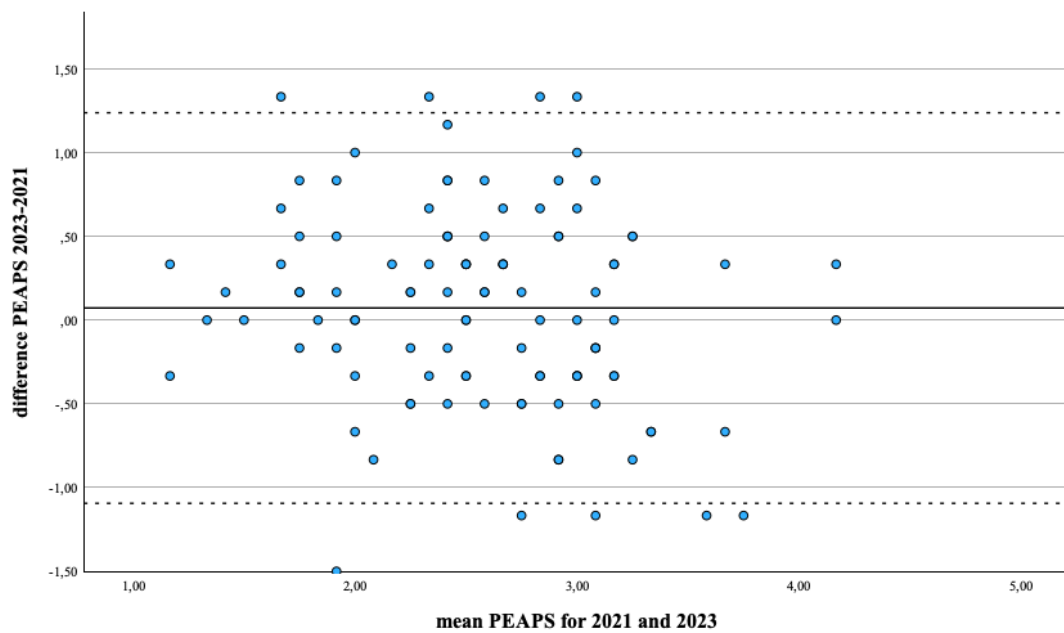


Figure 1. Bland-Altman plot for PEAPS in 2021 and 2023.

Note. Bland-Altman plot for intra-rater agreement ($N=106$). Mean difference is shown as solid line, limits of agreement (i.e., 95 % CI of the mean difference) are shown as dotted lines.

Finally, Intraclass Correlation Coefficient (ICC) is calculated to simultaneously account for both correlation and agreement between measurements. Following the suggestions of Koo and Li (2016), we estimate the ICC and its 95% confident intervals with a mean-rating ($k=6$), absolute-agreement, 2-way mixed-effects model: $ICC=0.755$, with 95% $CI=0.641-0.833$. Based on the ICC, we conclude that the test-retest reliability of the PEAPS scale is “substantial” to “almost perfect” (Landis & Koch, 1977, p. 165; see also Fleiss et al., 2003, p. 619).

DISCUSSION

“Recent years have seen the rise of political figures with a particularly abrasive, controversial, and even antagonistic character” (Nai & Maier, 2024) – prompting several researchers to take a closer look at these politicians. Part of such a focus has revolved around examining the personality, particularly the aversive personality, of politicians in order to assess its influence on their behavior. However, previous measures have not been suitable for self-reporting by political elites because the inventories were often still too long and contained (supposedly) extremely worded items.

To delve into the aversive personality of politicians, the Political Elites Aversive Personality Scale (PEAPS) was introduced (Maier et al., 2023). PEAPS is, to the best of our knowledge, the first inventory that seems in principle suitable for capturing the core of the aversive personality among political elites with self-assessments. First research supported the reliability and validity of PEAPS (Maier et al., 2023), but the scale had not yet demonstrated its test-retest reliability – and thus its stability across at least two measurement occasions of the same individual.

Taking advantage of a unique situation – the 2023 repeated Berlin state election, in which only parties and politicians who had already run in 2021 were allowed to participate – we find that PEAPS shows moderate to good levels of reliability, using different methods for estimating test-retest reliability. More precisely, correlation analyses suggest a moderate (towards strong) association

between the two measurement occasions, a paired t-test shows a high level of agreement between the measurement occasions, the Bland-Altman plot suggests an acceptable level of test-retest reliability, and the Intraclass Correlation (ICC) suggests a “substantial” to “almost perfect” association. Overall, these findings suggest a minimum acceptable stability of the inventory, particularly in light of the time lag between the two measurement occasions, which is with 17 months rather large for examining the test-retest reliability of a psychological trait measure.

This study comes with certain limitations. First, whereas support for the reliability and validity of the PEAPS has been shown for a larger sample of German candidates (Maier et al., 2023), the test-retest reliability analyzed in this paper is based on a smaller sample of candidates running for Berlin state elections, self-selecting into our survey. Further studies should thus test whether aversive personality of politicians can also be repeatedly measured in a reliable way in larger sample. Second, the PEAPS has been only used with German samples, so far. It thus needs to be examined whether the PEAPS is also useful for research in other countries, including an analysis of test-retest reliability. Third, the time elapsed between the first and second measurement points is quite long, which may have a negative impact on test-retest reliability. Further research should use a shorter time interval to produce results that are more comparable with other findings for test-retest reliability.

Based on our results, we recommend the use of PEAPS to self-assess the aversive personality of politicians. PEAPS offers the possibility to measure the aversive personality characteristics of political decision-makers with very few items and in connection with a clear theoretical concept (i.e., the Dark Factor of Personality, D). In principle, this short scale also allows the measurement of the aversive personality of politicians in research designs in which participants have little time (e.g., experts) or where the measurement of extensive personality assessments is expensive (e.g., population surveys). Because the PEAPS items have no relation to politics, we believe that the scale might also be suitable for measuring aversive personality traits of other elites, e.g., corporate leaders. Thus, future research might aim to test the reliability and validity of PEAPS in other applied fields of psychology with high-level decision-makers, such as organizational psychology. Given the conceptualization of D as a fluid construct, it is possible to develop scales specifically tailored to a population in focus, potentially using PEAPS as a starting point when elites are considered. Finally, applying PEAPS on other political levels and in other political systems would allow to conduct much-needed, comprehensive comparative research on the personality of politicians (or other elites). This would be beneficial for research in the field of political psychology, as it would allow researchers to systematically integrate an important but not easily measured aspect of personality as an explanatory variable for political attitudes and behavior. We hope that the complementary test of PEAPS’ retest-reliability presented here will further substantiate the basis for this.

REFERENCES

- Aldridge, V. K., Dovey, T. M., & Wade, A. (2017). Assessing test-retest reliability of psychological measures. *European Psychologist*, 22(4), 207–218. <https://doi.org/10.1027/1016-9040/a000298>
- Bader, M., Hilbig, B. E., Zettler, I., & Moshagen, M. (2023). Rethinking aversive personality: Decomposing the Dark Triad traits into their common core and unique flavors. *Journal of Personality*, 91(5), 1084–1109. <https://doi.org/10.1111/jopy.12785>
- Chabrol, H., Van Leeuwen, N., Rodgers, R., & Séjourné, N. (2009). Contributions of psychopathic, narcissistic, Machiavellian, and sadistic personality traits to juvenile delinquency. *Personality and Individual Differences*, 47(7), 734–739. <https://doi.org/10.1016/j.paid.2009.06.020>
- Du, T. V., Miller, J. D., & Lynam, D.R. (2022). The relation between narcissism and aggression: A meta analysis. *Journal of Personality*, 90(4), 574–594. <https://doi.org/10.1111/jopy.12684>
- Fleiss, J. L., Levin, B., & Paik, M.C. (2003). *Statistical methods for rates and proportions*. 3rd edition. Wiley.
- Gnambs, T. (2014). A meta-analysis of dependability coefficients (test–retest reliabilities) for measures of the Big Five. *Journal of Research in Personality*, 52, 20–28. <https://doi.org/10.1016/j.jrp.2014.06.003>
- Hilbig, B. E., Thielmann, I., Zettler, I., & Moshagen, M. (2023). The dispositional essence of proactive social preferences: The dark core of personality vis-à-vis 58 traits. *Psychological Science*, 34(2), 201–220. <https://doi.org/10.1177/09567976221116893>
- Joly, J., Soroka, S., & Loewen, P. (2019). Nice guys finish last: Personality and political success. *Acta Politica*, 54(4), 667–683. <https://doi.org/10.1057/s41269-018-0095-z>

- Knežević, G. (2003). *Koreni amoralnosti [The Roots of Amorality]*. Institut za kriminološka i sociološka istraživanja, Institut za psihologiju.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. <https://doi.org/10.2307/2529310>
- Leiner, D. J. (2019). Too fast, too straight, too weird: Non-reactive indicators for meaningless data in internet surveys. *Survey Research Methods, 13*(3), 229-248. <https://doi.org/10.18148/srm/2019.v13i3.7403>
- Maier, J., Oschatz, C., Stier, S., & Zettler, I. (2023). A short scale to measure self-reported aversive personality in political elites. *Journal of Personality Assessment, 105*(5), 625-635. <https://doi.org/10.1080/00223891.2022.2130341>
- Marcus, D. K., Zeigler-Hill, V., Mercer, S. H., & Norris, A. L. (2014). The psychology of spite and the measurement of spitefulness. *Psychological Assessment, 26*(2), 563-574. <https://doi.org/10.1037/a0036039>
- McLarnon, M. J., & Tarraf, R. C. (2017). The Dark Triad: Specific or general sources of variance? A bifactor exploratory structural equation modeling approach. *Personality and Individual Differences, 112*, 67-73. <https://doi.org/10.1016/j.paid.2017.02.049>
- Moshagen, M., Hilbig, B. E., & Zettler, I. (2018). The dark core of personality. *Psychological Review, 125*(5), 656-688. <https://doi.org/10.1037/rev0000111>
- Moshagen, M., Zettler, I., & Hilbig, B. E. (2020). Measuring the dark core of personality. *Psychological Assessment, 32*(2), 182-196. <https://doi.org/10.1037/pas0000778>
- Nai, A. (2019). Disagreeable narcissists, extroverted psychopaths, and elections: A new dataset to measure the personality of candidates worldwide. *European Political Science, 18*(2), 309-334. <https://doi.org/10.1057/s41304-018-0187-2>
- Nai, A., & Maier, J. (2020). Dark necessities? Candidates' aversive personality traits and negative campaigning in the 2018 American Midterms. *Electoral Studies, 68*. <https://doi.org/10.1016/j.electstud.2020.102233>
- Nai, A., & Maier, J. (2024). *Dark politics. The personality of politicians and the future of democracy*. Oxford University Press.
- Nai, A., & Martínez i Coma, F. (2019). The personality of populists: provocateurs, charismatic leaders, or drunken dinner guests? *West European Politics, 42*(7), 1337-1367. <https://doi.org/10.1080/01402382.2019.1599570>
- O'Meara, A., Davies, J., & Hammond, S. (2011). The psychometric properties and utility of the Short Sadistic Impulse Scale (SSIS). *Psychological Assessment, 23*(2), 523-531. <https://doi.org/10.1037/a0022400>
- Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality, 36*(6), 556-563. [https://doi.org/10.1016/S0092-6566\(02\)00505-6](https://doi.org/10.1016/S0092-6566(02)00505-6)
- Schreiber, A., & Marcus, B. (2020). The place of the "Dark Triad" in general models of personality: Some meta-analytic clarification. *Psychological Bulletin, 146*(11), 1021-1041. <https://doi.org/10.1037/bul0000299>
- Watts, A. L., Lilienfeld, S. O., Smith, S. F., Miller, J. D., Campbell, W. K., Waldman, I. D., Rubenzer, S. J., & Faschingbauer, T. J. (2013). The double-edged sword of grandiose narcissism: Implications for successful and unsuccessful leadership among US presidents. *Psychological Science, 24*(12), 2379-2389. <https://doi.org/10.1177/0956797613491970>
- Zettler, I., Schild, C., Lilleholt, L., Kroencke, L., Utesch, T., Moshagen, M., Böhm, R., Back, M. D., & Geukes, K. (2022). The role of personality in COVID-19-related perceptions, evaluations, and behaviors: Findings across five samples, nine traits, and 17 criteria. *Social Psychological and Personality Science, 13*(1), 299-310. <https://doi.org/10.1177/19485506211001680>
- Zhu, Y., & Jin, C. (2021). A meta-analysis of the relationship between the Dark Triad and aggressive behaviors. *Advances in Psychological Science, 29*(7), 1195-1209. <https://doi.org/10.3724/SPJ.1042.2021.01195>

ACKNOWLEDGEMENTS

The authors would also like to thank all the candidates who took the time to participate in our surveys despite busy schedules.

DECLARATION OF INTEREST

The authors report there are no competing interests to declare.

FUNDING

The first two authors acknowledge generous financial support from the German Research Foundation (DFG) (project number 441574527).

DATA AVAILABILITY STATEMENT

Since our data contain sensitive, non-anonymized information from candidates, the data can be only made available to other researchers on request. The script of our data analyses is stored in the following Open Science Foundation (OSF) repository: https://osf.io/jmxt8/?view_only=bc98954c91af4c22bcc35198fb39fe80