



Improving Reactions to Forced-Choice Personality Measures in Simulated Job Application Contexts Through the Satisfaction of Psychological Needs

Tristan C. Borman¹ · Patrick D. Dunlop¹ · Marylène Gagné¹ · Matthew Neale²

Accepted: 21 February 2023
© The Author(s) 2023

Abstract

There is a wealth of evidence justifying the use of personality assessments for selection. Nonetheless, some reluctance to use these assessments stems from their perceived vulnerability to response distortion (i.e., faking) and the somewhat negative applicant reactions they elicit, when compared to other assessments. Adopting a forced-choice personality assessment format appears to alleviate the former problem but exacerbates the latter. In this study, we introduce basic psychological needs as a theoretical foundation to develop interventions to improve reactions to forced-choice personality assessments. We propose that the forced-choice format impedes respondents' desire to respond to items in a preferred way, interfering with autonomy need satisfaction, and constrains respondents' opportunity to show their capabilities, interfering with competence need satisfaction. In this pre-registered between-subjects experiment ($N = 1565$), we investigated two modifications to a ranked forced-choice personality questionnaire and compared these to traditional forced-choice and single-stimulus (Likert) formatted questionnaires. One modification, where participants could write a free-text response following the assessment, did not show significant effects on reactions. The second modification allowed participants to view all items they had ranked last (first) and then identify any the participant believed in fact described them well (poorly). That modification positively affected perceived autonomy- and competence-support, and fairness perceptions, bridging approximately half of the gap between reactions to forced-choice and single-stimulus assessment formats. This study suggests that a modification to forced-choice personality questionnaires may improve applicant reactions and that basic psychological needs theory may be a fruitful lens through which to further understand reactions to assessments.

Keywords Personality · Forced-choice assessment · Applicant reactions · Basic psychological needs · Pre-registration

Introduction

Personality questionnaires are recommended for use in selection as they measure traits that (a) have known associations with work outcomes (Barrick & Mount, 1991; Hurtz & Donovan, 2000; Tett et al., 1991), (b) provide unique information not provided by other assessments (Schmidt & Hunter, 1998), and (c) exhibit relatively small group differences, providing an avenue to reduce adverse impact (Sackett et al., 2021). Nonetheless,

job candidates often regard personality questionnaires relatively less favorably than some other types of assessments such as interviews and work samples, with estimates of mean “favorability” ratings falling close to or just below the mid-point of rating scales (Anderson et al., 2010; Hausknecht et al., 2004; McFarland, 2013; Nikolaou & Judge, 2007; Rynes & Connerley, 1993). Applicant reactions describe the “attitudes, affect, or cognitions an individual might have about the hiring process” (Ryan & Ployhart, 2000, p. 566) and are of interest to employers because these reactions can influence applicants' decisions to accept job offers, their decisions to recommend the organization to other potential applicants, and their perceptions of an employer's brand generally (Chapman et al., 2005; Hausknecht et al., 2004; McCarthy et al., 2017a, 2017b). In the context of growing social media use including anonymous platforms such as Glassdoor, Seek, and Reddit, applicant reactions — whether positive or negative — can now be easily expressed more publicly than

✉ Patrick D. Dunlop
patrick.dunlop@curtin.edu.au

¹ Future of Work Institute, Faculty of Business and Law, Curtin University, Kent Street, Bentley, Western Australia 6102, Australia

² Criteria Corp, Brisbane, Queensland, Australia

ever before, with potentially major implications for employers' reputations or brands (Nikolaou et al., 2015; Woods et al., 2019). Meanwhile, increasing scarcity and competition for qualified candidates has put pressure on employers' ability to attract and select talented personnel (Michaels et al., 2001; Serrat, 2017), creating clear incentives to ensure a positive candidate experience during a selection process (McCarthy et al., 2018).

In addition to the relatively unfavorable applicant reactions to personality questionnaires, compared to alternative assessment types, these assessments pose the challenge of "fakability" (Hough & Oswald, 2008; Morgeson et al., 2007). Specifically, when personality assessments are used for job applications, motivated candidates might provide inaccurate responses, for example, overplaying how friendly or hardworking they are, to increase the chances of securing a job offer (Donovan et al., 2014; Ziegler et al., 2011). Though there is little doubt that faking can, and does, occur in high-stake settings (Birkeland et al., 2006; Hu & Connelly, 2021; Viswesvaran & Ones, 1999), there remains a lively debate as to whether it poses a threat to the construct and criterion validity of personality assessments (e.g., Hogan, 2005; Marcus, 2009, 2022; Tett & Simonet, 2021; Tett et al., 2022), with empirical evidence remaining somewhat equivocal (e.g., Donovan et al., 2014; Dunlop et al., 2022; Huber et al., 2021; Peterson et al., 2011). Nonetheless, concerns around faking have prompted personality questionnaire vendors to invest in techniques designed to minimize its presence or impact.

One popular approach adopted by personality questionnaire designers to mitigate the faking problem is the *forced-choice* questionnaire format (Gordon, 1951; Salgado et al., 2015). In contrast to the *single-stimulus* format, where personality items are presented and rated independently (e.g., Likert scales), the forced-choice format is one where respondents receive blocks (e.g., pairs, triads, or tetrads) of personality items simultaneously. The respondent's task is to consider the items from a block concurrently and either rank the items, identify the items that describe them the most and the least (the "MOLE" format), or in the case of item pairs (the "PICK" format), select the item from the pair that best represents their personality (Cao & Drasgow, 2019; Seybert & Becker, 2019). In terms of reducing faking, the key to this format is that respondents will encounter some blocks of statements where a most desirable response (e.g., the most desirable ranking of items in a block) is unclear (Hughes et al., 2021). For example, if a candidate who wishes to fake is faced with a block of three items measuring the desirable traits of friendliness, diligence, and sociability, they can choose only one of these items to be ranked first and must choose one to rank last. When faced with such a dilemma, candidates are thought to be more likely to rank statements in a way that is representative of their actual trait disposition rather than respond in a positively biased way (Berkshire, 1958; Gordon, 1951). By contrast, with a single-stimulus format, a candidate can simply fully endorse all three items (e.g., by selecting *strongly agree*).

Compared to single-stimulus formats, well-designed forced-choice formatted measures appear to exhibit lower rates and extents of faking (Cao & Drasgow, 2019; Wetzel et al., 2021), while still affording psychometric equivalence to their single-stimulus counterparts (Zhang et al., 2020). However, as implied by its name, the forced-choice format also introduces constraints onto the respondent that may have negative side-effects. In studies of participants completing forced-choice and single-stimulus formatted questionnaires for low-stake research purposes, Sass et al. (2018) and Zhang et al. (2020) found that test-taking motivation and affect were not affected by the different formats. Zhang et al. (2020), however, found that the forced-choice questionnaire was perceived as more difficult by respondents than the single-stimulus counterpart. Furthermore, less favorable test-taker reactions to forced-choice personality assessments, relative to single-stimulus formats, have been observed in at least one study where the assessments were completed for a hypothetical job application (Converse et al., 2008). To the extent that Converse et al.'s results generalize to the assessment of real job applicants, employers may face a trade-off between reducing faking and providing a more positive candidate experience.

In this investigation, we adopt a basic psychological needs theoretical lens to identify two mechanisms by which the constraining format of the forced choice assessment can undermine the satisfaction of respondents' psychological needs, ultimately resulting in relatively more negative evaluations of the assessment's fairness. Specifically, we propose that the restrictiveness of the format limits the satisfaction of respondents' need for autonomy and that the forced-choice format will also create situations where respondents are unable to show their capabilities, undermining the satisfaction of the need for competence (see also Dalal et al., 2019). We then propose and empirically evaluate two modifications to the forced-choice format that aim to improve the satisfaction of the needs for autonomy and competence among test-takers, without compromising the psychometric properties of the assessment. Importantly, and as we explain in more detail below, the interventions we examine are compatible with any forced-choice assessment structure (i.e., pairs, triads, rank, MOLE). Altogether, we contribute to the literature by introducing a novel perspective to study applicant reactions, while also adding practical insights for improving application reactions through the design and evaluation of the two interventions (Ryan & Huth, 2008).

Improving Applicant Reactions to Forced-Choice Assessments

Although there is a speculation as to why the forced-choice personality assessment format is less favorably regarded than the single-stimulus format, we are aware of only one study that has directly investigated the potential causes of, and remedies to, these negative reactions. In that paper, Dalal

et al. (2019) proposed that forced-choice measurement may trigger negative reactions because the assessment format threatens test-takers' positive self-concepts (i.e., their ability to portray their positive self-concept in the assessment). Accordingly, these investigators hypothesized that reactions could be improved by reducing self-concept threat-provoking stimuli or providing ways to portray a positive self-concept through minor variations to the response format. The researchers proposed and examined four interventions and investigated the effects of the interventions on reactions to a computer-adaptive forced-choice assessment where items were presented in pairs (i.e., "PICK").

The first intervention involved replacing the classical pairwise PICK format with a graded 5-point response scale where the items from a pair were placed at either pole of the scale, with the idea being that respondents would be able to, at least partly, endorse one of the items in the pair if they so wished (Dalal et al., 2019). The second intervention involved removing a subset of the most socially undesirable items from the item pool, thus reducing the chances of respondents being forced to endorse an undesirable item if it were paired with another such item. The third intervention involved providing detailed post-assessment feedback to respondents, with the idea being that feedback would give participants the sense of having the chance to personally verify their results. The fourth intervention involved modifying the assessment instructions to adopt a softer, polite tone, as opposed to a more typical standard cold, directive tone. The goal here was to induce more positive immediate "face" reactions to the request to complete the assessment. Across two experiments, Dalal et al. (2019) compared participants' reactions to an assessment with one of the four modifications to the unmodified PICK assessment format.

Altogether, Dalal et al. (2019) had some success with respect to improving reactions. Specifically, they found that the graded response scale modification was associated with significantly more favorable reactions than the standard PICK format. Removing the most undesirable items also showed a significant positive effect on two of the five reaction outcome variables that were studied. Across the two interventions, the observed effect sizes were small, however, with Cohen's *d* among the statistically significant results ranging from 0.10 to 0.14. We also note, first, that the graded response scale modification cannot be easily scaled to forced-choice assessment structures comprising blocks of more than two items, and second, that the omission of socially undesirable items from an item pool is likely to have implications for the psychometric properties of forced-choice assessments that do not use computer adaptive item selection methods. Thus, we applaud the study for identifying these two methods to improve reactions, however, there remains a

need to identify modifications that can be applied to a wider range of forced-choice assessment structures.

We also suspect there may be scope to identify modifications that improve reactions to a greater extent. Indeed, the intervention where test-takers were provided with detailed feedback delivered some relatively strong positive effects on test-taker reactions, with Cohen's *d* ranging from 0.35 to 0.52 (Dalal et al., 2019). While we endorse the delivery of detailed feedback to candidates who have completed a personality questionnaire, Dalal et al. (2019) had unfortunately not included a condition with a single-stimulus version of the personality questionnaire. It is therefore not possible to evaluate the extent that the feedback provision improved reactions, specifically, to the forced-choice format or whether it would improve reactions more generally to all personality assessment formats. We note, finally, that Dalal et al. (2019) did not observe a significant effect of instruction language on their reaction outcomes.

Altogether, there appears to be a clear need to investigate alternative perspectives, coupled with interventions to improve reactions to the forced-choice format. In this investigation, we consider an alternative theoretical lens that grounds applicant reactions in basic psychological needs, thus introducing a novel theoretical perspective from which to develop interventions to improve these reactions. We also attempt to capture the theorized processes by which applicant reactions are shaped.

Basic Psychological Needs Theory and the Experience of Being an Applicant

An influential motivation theory incorporating psychological needs is self-determination theory (Ryan & Deci, 2000). Self-determination theory provides a comprehensive and well-evaluated framework for understanding human motivation, with its claims regarding psychological needs articulated in a central mini-theory, basic psychological need theory (BPNT; Ryan & Deci, 2017). According to BPNT, psychological needs are defined as "innate psychological nutrients that are essential for ongoing psychological growth, integrity, and well-being" (Deci & Ryan, 2000, p. 229). The theory proposes that individuals have a universal set of three basic psychological needs: autonomy, competence, and relatedness, which we define shortly.

In general terms, the appeal of BPNT in the context of applicant reactions stems from its utility in understanding the relationship between the social contextual environment (in this case, the experience of completing an assessment for a job application) and an individual's well-being and ill-being in broader settings (Gunnell et al., 2013). Put into context, we propose that the format of the forced-choice assessment

likely undermines the satisfaction of psychological needs. If the format can be modified to better satisfy the psychological needs, doing so should create more positive affective reactions, and a more positive appraisal of the assessment process, and potentially, the hiring organization. Although BPNT has rarely been invoked in selection research (Gagné et al., 2022), there is some initial evidence to support the claims above. A recent correlational study that used BPNT in the context of applicant reactions demonstrated that gamified assessment tools' ability to support the needs for competence, and autonomy was associated with a variety of applicant reactions including recommendation intentions, overall satisfaction, and organizational attractiveness (Buil et al., 2020). More broadly, however, evidence for interventions aimed at supporting basic psychological needs have been systemically reviewed in other domains, such as health (Ntoumanis et al., 2021) and education (Su & Reeve, 2011; Vasconcellos et al., 2020), with promising results. Despite this wealth of evidence for the application of BPNT, there have been, to our knowledge, no prior attempts to develop interventions for organizational assessment procedures aimed at supporting or satisfying basic psychological needs. This, we believe, constitutes an opportunity to better understand the mechanism which causes negative applicant reactions.

The need for autonomy is defined as a need to feel volitional, that is, the need to experience one's actions as being congruent with personal values and interests, rather than being pressured to act by external forces (Ryan & Deci, 2017). Satisfaction of the need for autonomy has been meta-analytically associated with higher positive affect, lower negative affect, higher engagement, higher job satisfaction and affective commitment, and lower turnover intentions (Van den Broeck et al., 2016). Thus, ensuring that an assessment format satisfies this need may yield more positive applicant reactions. Previous research has demonstrated that providing choice about aspects of task engagement tend to enhance feelings of autonomy (Patall et al., 2008) and, here, we posit that the forced-choice format restricts choice more so than single-stimulus measures because it is likely to restrict opportunities for applicants to "act"; that is, choose an item congruently with their values, preferences, or self-concept. That is, forced-choice assessment formats present applicants with the occasional dilemmas of deciding which, of a set of statements that *all* describe the applicant well (or poorly), to rank first (or last). In contrast, with single-stimulus measures, applicants can rate each item independently and can thus choose to endorse as many or as few items as they wish, granting freedom to portray the self in a more self-congruent manner. Consequently, we propose that the restriction of choice will undermine the autonomy need and therefore produce poorer perceptions of fairness among test-takers, when compared to a single-stimulus personality measure. Furthermore, we propose that adaptations to the

forced-choice format that allow for greater choice may be an effective remedy.

The need for competence is defined as a need to feel effective in one's environment, or as having a sense of mastery (Ryan & Deci, 2017). Competence satisfaction has been meta-analytically associated with higher positive affect, lower negative affect, higher job satisfaction, and higher task performance (Van den Broeck et al., 2016). In the context of a personality assessment for a job application, we anticipate that a test-taker would feel more effective and masterful to the extent that the personality assessment provides them with the opportunity to describe their positive characteristics, strengths, or skills. For example, Soto et al.'s (2021) social, emotional, behavioral skills framework illustrates how the Big Five major personality dimensions correspond to the five skill domains, namely cooperation, resilience, self-management, social engagement, and creativity, each of which may describe strengths that candidates would want the opportunity to highlight in an assessment. Here, we propose that certain forced-choice item blocks might make candidates feel they are prevented from reflecting their strengths through their answers, leaving them feeling like they could not complete the assessment competently to demonstrate their strengths. For example, a candidate that sees herself as both socially skilled and attentive to deadlines might be presented with a forced-choice block which includes both the extraversion item, "Good at making friends quickly" and the conscientiousness item, "Work hard to achieve deadlines". If she wants to present both traits as strengths, she would be able to do so with a single-stimulus format but would be prevented from doing so in a forced-choice format, potentially resulting in lower satisfaction of the need for competence. We expect, therefore, that a modification to the format that allows respondents to highlight their strengths may be an effective remedy.

Finally, the need for relatedness is defined as the need to feel socially connected, often described as belonging and feeling valued by others, and is meta-analytically associated with higher engagement, job satisfaction, and affective commitment (Van den Broeck et al., 2016). Although we do not wish to discount the importance of relatedness need satisfaction for improving applicant reactions to assessments and selection processes, we did not have any strong views about how the single-stimulus versus forced-choice formatting of a personality questionnaire could systematically affect the extent that this need was satisfied. Accordingly, we had no expectations that the satisfaction of the relatedness need would be associated with the format or its modifications.

Summary and Hypotheses

This study draws from BPNT as a theoretical lens through which to understand the negative reactions to forced-choice personality measures, relative to single-stimulus (e.g., Likert)

formatted measures. In doing so, we primarily focus on the basic psychological needs for autonomy and competence, which we then seek to improve through interventions. Accordingly, we evaluate two modifications to the forced-choice personality assessment format that aims to better support the basic psychological needs of autonomy and competence. In addition to the theoretical mechanisms discussed above, in designing these modifications, we also considered two practical constraints: First, the modifications must not impact on the structure of, or responses to, the questionnaire, and hence its construct validity or its ability to prevent intentional response distortions; and second, the modifications must be compatible with forced-choice assessments with any structure (i.e., PICK, MOLE, and rank; any number of items in a block). In so doing, we devised two modifications to the assessment format. Both took the form of an end-of-questionnaire “final say” exercise that provides respondents with the opportunity to identify or clarify aspects of their personality that they felt the preceding forced-choice questionnaire did not allow them to. We hypothesized that providing a “final say” would add an element of choice to the questionnaire — providing autonomy-support and a second chance for respondents to demonstrate missed strengths — competence-support.

To further examine the extent to which constraint influences test-taker reactions, two format variations of this “final say” task were developed. The first took on a mechanistic form. In this “fixed final say” modification, all the items that respondents rated as *most like them* are presented in a list, and respondents are asked to identify any items from that list that they feel do not describe them well. Then, the process is repeated with all the items that were rated as *least like them*, but with respondents being asked to identify any items that did describe them well. The second variation took on an open-ended “free” format involving a text box and the opportunity to clarify aspects of their personality via a written response. To detect effects on test-takers’ perceptions of need-support, these two modifications are compared to a control condition with a forced-choice assessment without any need-supportive post-questionnaire exercise and a condition with a single-stimulus (Likert) format personality questionnaire comprising the same items as the forced-choice formats.

We hypothesized that affording participants with the opportunity to provide a final say (in the form of the fixed and free intervention conditions) after completing a forced-choice personality questionnaire will increase the extent to which they experience autonomy-satisfaction (H1a — fixed; H1b — free) and competence-satisfaction (H2a — fixed; H2b — free), relative to control conditions where a final

say is not requested. These hypotheses were pre-registered.¹ We further hypothesized (not pre-registered) that autonomy (H3a) and competence need-support (H3b) would mediate the relations of the personality questionnaire format and overall perceptions of fairness.

Method

Collection and use of these data for research was approved by the Human Research Ethics Office of Curtin University (Curtin HREC; record number HRE2020-0236). Data were collected over two waves, with a pre-registration lodged prior to each wave. All study materials, including descriptions and explanations of departures from the pre-registered analyses, analysis scripts, output files for factor analyses, additional exploratory analyses, and raw data files are provided on the project’s OSF webpage: <https://osf.io/j8wdy>.

Design

This study employed an experimental between-groups design, with participants randomly assigned to one of four conditions as follows: (1) fixed final say; (2) free final say; (3) control²; and (4) Likert. Details relating to each condition are described in the “[Procedure](#)” section.

Participants

Data were collected over two waves. The first round of data collection involved 902 participants and was undertaken in August 2020 as part of the first author’s Honors dissertation project. An additional round of data collection was undertaken in November 2020 and involved another 899 participants. All participants were recruited from Mechanical Turk (MTurk) using the CloudResearch service (previously TurkPrime; Litman et al., 2016), which provides features that prevent problematic MTurk accounts from accessing the survey. Residents of the USA aged 18 or older, who had completed at least 500 tasks on MTurk, with an approval rate of at least 97% were eligible to participate in the study (Peer et al., 2014). Participants who completed the survey

¹ Original pre-registration for data collection 1 are found via the following URL: <https://osf.io/z3y6m>.

Pre-registration for data collection 2: <https://osf.io/tr7b6>.

² There were in fact two control conditions, with one including an additional self-reported knowledge questionnaire that was completed after the forced-choice personality questionnaire. This additional task was added to one control condition to assess the effect of total assessment time (Ackerman & Kanfer, 2009; Ryan & Ployhart, 2000), which is confounded with both final say conditions. We found no evidence of differences between these two control conditions on any variable, and so, as per the pre-registration, the conditions were assumed to be equivalent and were combined. The data file in the OSF provides information on the control conditions separately.

were paid US\$2.50. An analysis of demographic variables showed no statistically significant differences between these two participant groups.

In line with our pre-registration, participants' responses were excluded from analysis if that participant (1) Responded incorrectly to a directed-response item ($n = 219$); (2) indicated explicitly that we should not include their responses ($n = 6$); or (3) gave a string of identical responses in the questionnaire ($n = 11$). These exclusions were not mutually exclusive and were applied in the above sequence. The final sample was 1565 (792 from the first collection and 773 from the second). Approximately half identified as female (50.6%), 49.2% as male, two participants as non-binary and one did not disclose. Participants' mean age was 41.0 years ($SD = 12.2$ years), 74% held an Associate degree or higher tertiary qualification, and had worked in full-time employment for an average of 17.2 years ($SD = 11.9$).³

Procedure

The study was hosted on the university's Qualtrics platform. After providing consent, participants completed a short demographic survey. They were then instructed to imagine themselves applying for a job that they believe they would be a good fit for and to specify the name of that job in a textbox. Participants were then randomly assigned into one of the four conditions.

In all conditions, participants were informed that as part of their (hypothetical) job application, the employer has invited them to complete some assessments, and that the responses to the assessments would inform the hiring decision. Participants in all conditions were first invited to complete a personality questionnaire (entitled the "Corporate Personality Questionnaire (CPQ)"), but the format of the questionnaire varied by condition. In the fixed final say, free final say, and control conditions, the personality questionnaire was in a forced-choice format (see "Measures" for details). In the Likert condition, participants completed a Likert-format questionnaire comprising the same items.

After completing the personality questionnaire, participants in the fixed final say condition were provided with a list of the 40 personality items that they had ranked *most like me*. Above the list was a message explaining that the employer understood that there may have been occasions that the participant had ranked a statement first, even though the statement did not describe them well. Participants were asked to identify any statements from the set of 40 that they felt *did not* describe them well. Following that participants

were provided with a list of the 40 items that they had rated "least like me" and were asked to identify any that *did* describe them well. See Fig. 1 for the instructions provided to participants on this task along with two sample items.

After completing the personality questionnaire, participants in the free final say condition were provided with the following instruction:

Before you finish the CPQ, we would like to know whether you felt there are any aspects of your personality that you feel were either overlooked or likely to be misrepresented in the previous part of the CPQ (e.g., because you had to rank some statements higher or lower than you would have liked to)?

If so, please use the text box below to explain what you felt was overlooked. While we encourage you to respond, we would ask that you keep your response relatively brief and not to spend too long on this. As a guide, 2–6 sentences or up to three points is sufficient.

Underneath the instruction was with an open-ended text box where participants could type their responses.

After completing any relevant condition-specific tasks, all participants completed an evaluation survey, containing the basic psychological needs measure, an overall fairness perception measure, and an attention-check item.

Measures

Personality Questionnaire

Per their condition, participants completed one of two different formats of the personality questionnaire: forced-choice or single-stimulus (Likert). Full details about this instrument including the items and basic psychometric properties are provided in the online supplement. Briefly, both questionnaires used the same set of 120 items, which were sourced from the International Personality Item Pool (IPIP; Goldberg et al., 2006) and were known indicators of the Big Five personality dimensions. The forced-choice format was derived from the work of Hughes et al. (2021) and comprised 40 three-item blocks asking participants to rank statements in order of most to least like them, via "drag-and-drop." The item blocks were selected to represent a spectrum of social desirability matching, with one-quarter of the blocks containing triplets that were very closely matched on desirability, one-quarter with items that were very poorly matched on desirability (see Hughes et al., 2021), and the matching for the remaining item blocks falling in between. In the Likert condition, participants responded to each item on a 5-point *strongly disagree to strongly agree* response scale. The personality questionnaire was intended only to resemble a realistic personality assessment that might be used in

³ We also re-ran all analyses including all participants' data and found practically identical results. Full data are provided in the paper's OSF page.

Fig. 1 Instructions in the fixed final say condition. “CPQ” is the name we assigned to the personality questionnaire in the study, and for brevity, only two of the 40 items are shown

Thank you for completing the first part of the CPQ.

In this next section of the CPQ, we have provided a list of **all** of the statements that you **ranked highest** during the first part. These were the statements you had chosen as “most like you”. We understand that there may have been occasions where you had to rank a statement first, even though you did not feel it described you very well.

Please quickly review these statements and identify any that you do not feel accurately describe your personality. Do not spend too long on this; one quick glance through the list is usually enough.

Please click the checkbox next to any statement that you feel **does not describe your personality accurately.**

	This statement does not describe my personality accurately
I don’t understand people who get emotional.	<input type="checkbox"/>
I fear for the worst.	<input checked="" type="checkbox"/>

In this next section of the CPQ, we have provided a list of **all** of the statements that you **ranked lowest** during the first part. These were the statements you had chosen as “least like you”. We understand that there may have been occasions where you had to rank a statement last, even though you felt that it described you very well.

Please quickly review these statements and identify any that you feel do accurately describe your personality. Do not spend too long on this; one quick glance through the list is usually enough.

Please click the checkbox next to any statement that you feel **does describe your personality accurately.**

	This statement does describe my personality accurately
I lack the talent for influencing people.	<input checked="" type="checkbox"/>
I contradict others.	<input type="checkbox"/>

job selection, and although the online supplement provides information about its construction and scoring, we did not use the scored responses in this study.

Basic Psychological Needs Satisfaction

We are unaware of other research using BPNT in the context of measuring applicant reactions to assessments, and we also note that need satisfaction measures typically must be tailored to the context (e.g., work, sport, education; Bartholomew et al., 2011; Longo et al., 2016; Van den Broeck et al., 2010). Thus, for this study, we constructed a set of 16 items that drew from a combination of existing applicant reactions measures (Bauer et al., 2001) and from measures of need satisfaction and frustration in other contexts (Chen et al., 2015; Longo et al., 2016; Ryan, 1982; Van den Broeck et al., 2010). These items, shown in Table 1, are designed to measure the degree to which participants felt the personality inventory satisfied the three psychological needs of autonomy, competence, and relatedness. We conducted a psychometric analysis of the scale properties (see “Results”).

The need for competence items focused on assessing the extent to which participants felt that the CPQ enabled them to demonstrate their competence or capability. The approach of asking about conveying competence is slightly different from typical measures of need satisfaction, which would ask about how competent a person felt *while performing a task or a set of activities*. We felt, however, that the traditional approach was inappropriate here because personality assessments are not performance-based tests (i.e., one cannot be

“good” or “poor” at completing a personality questionnaire). Thus, we drew extensively from the selection procedural justice scale’s *chance to perform* sub-scale (Bauer et al., 2001) in designing these items (i.e., the chance to convey competence). The need for relatedness items focused on the extent to which using the CPQ was interpreted as a sign that the employer was interested in learning more about the candidate at a deep level. As noted above, we did not expect relatedness to be affected by the format of the personality questionnaire but included it in our analyses to provide an initial test of the discriminant validity of the measure of the three needs and of the interventions. Finally, autonomy satisfaction items focused on the extent to which participants experienced choice or freedom while completing the CPQ. Accordingly, we drew extensively from the other contextualized measures of need satisfaction (Chen et al., 2015; Van den Broeck et al., 2010). The first two authors constructed the items and the third author, an expert in self-determination theory, reviewed them and recommended revisions, with the process being repeated until all the authors were satisfied. Responses to all items were collected using a five-point *strongly disagree to strongly agree* scale.

Fairness

Finally, we included a three-item measure of participants’ perceptions of how fair the CPQ was as a selection tool (Bauer et al., 2001), which used a five-point *strongly disagree to strongly agree* scale. Cronbach’s alpha was ($\alpha=0.95$).

Table 1 Three-factor maximum likelihood robust geomin-rotated factor structure of the psychological needs satisfaction 16-item set

Item		Rotated factor loadings		
		Autonomy	Competence	Relatedness
Auto1*	I did not feel like I had much freedom to choose how to answer the questions in the CPQ	–.903		
Auto2*	I felt like I was forced to answer the questions in the CPQ in the way that I did	–.838		
Auto3	I felt that I had choice about how I answered the questions in the CPQ	.777		
Auto4	I experienced a lot of freedom in how I answered the questions of the CPQ	.763		
Auto5*	I did not really have much of a choice about how to disclose my personality with the CPQ	–.740		
Auto6	When answering the questions in the CPQ, I felt free to “be myself”	.566		
Comp1	I believe that, through the CPQ, I conveyed that I am a strong candidate for the job of [job title]		.896	
Comp2	My responses to the CPQ would have demonstrated that I am a better fit, compared to other applicants for the job		.784	
Comp3	Through the CPQ, I was able to show that my personality is the right fit for the job		.694	
Comp4	The CPQ allowed me to show what my competences are		.637	
Comp5	It was clear to me how the CPQ would be able to assess my strengths as a candidate		.578	
Comp6*	I was not able to show my strengths as a candidate for the job of [job title] on the CPQ		–.437	
Rela1	By asking me to complete the CPQ, it felt like the hiring organization was genuinely interested in me as a person			.785
Rela2	By asking me to complete the CPQ, it felt like the hiring organization wanted to get to know me better			.931
Rela3	By asking me to complete the CPQ, it felt like the hiring organization was interested in understanding who I am			.896
Rela4*	By asking me to complete the CPQ, it felt like the hiring organization did not care about me as an individual			–.606

Factor loadings with absolute values less than .3 are not reported; full results with 1 to 4 factor solutions are available in the online supplement. CPQ was short for “Corporate Personality Questionnaire”, which was the title of the questionnaire provided to participants for immersion purposes

Auto autonomy; *Comp* competence; *Rela* relatedness

* item is negatively keyed

Results

Needs-Supportive Feature Use

Prior to undertaking substantive analyses, we first examined the extent to which the need-supportive features were used. In the fixed final say condition, a participant had the opportunity to mark up to 80 personality statements whose rankings had mischaracterized the participant’s personality. On average, participants marked 12 such statements ($SD = 8.1$), with 93% of participants marking fewer than one-quarter of the statements, and only 2 participants not marking a single statement. In the free final say condition, we found that 22.5% ($n = 73$) did not write anything in the textbox, and a further 19.4% ($n = 63$) wrote that they felt the assessment would have represented their personalities well. After disregarding the blank responses, the mean number of words written was 27.3 ($SD = 19.2$).

Needs-Satisfaction Item Analyses

Because the need satisfaction item set is novel, we undertook a combination of exploratory factor analyses on data collected in the first round ($N = 792$) and confirmatory factor analysis on the data collected in the second round ($N = 773$) to assess the three-factor (i.e., three-needs) structure. This approach was a slight departure from our pre-registered analyses. A summary of the key findings is provided here and detailed analyses including comparisons of the three-factor model to model with different alternative structures, and all statistical output files are provided in the supplement.

Initially, maximum likelihood robust (MLR) exploratory factor analysis was undertaken in Mplus 8.5 to investigate the underlying structure of the set of items. An oblique geomin rotation was used as the perceived need-support factors were expected to be positively correlated (cf. Van den Broeck et al., 2010). The first three eigenvalues were greater than 1 (8.69, 1.50, and 1.11), and the 3-factor model fit was sound

overall ($\chi^2(75, N=792)=220.9, p<0.001; RMSEA=0.050, CFI=0.978, SRMR=0.019$), compared to common benchmarks (Browne & Cudeck, 1992; Hu & Bentler, 1999), and considerably improved relative to the 2-factor model ($\chi^2(89, N=792)=773.6, p<0.001; RMSEA=0.099, CFI=0.894, SRMR=0.047$). The item text and rotated factor loadings are shown in Table 1 and the items loaded strongly to very strongly on their respective factors. Nonetheless, there was a very strong primary factor, and the correlations among the three need satisfaction factors ranged from 0.61 to 0.70 and were judged as strong, relative to correlation benchmarks observed by Bosco et al. (2015) and Paterson et al. (2016).

A confirmatory factor analysis (CFA) was conducted on the data collected from the second round with the sixteen items loading on their three designated factors and factors free to covary. The overall model fit of this model was moderate, relative to common benchmarks ($\chi^2(101, N=773)=426.0, p<0.000; RMSEA=0.065$ [95% $CI=0.058, 0.071$], $CFI=0.950, SRMR=0.045$). Importantly, however, inspection of modification indices did not reveal any clear unaccounted-for relations between the items; that is, there appeared to be no obvious driver of interrelations among the item set that were not accounted for by the CFA model. Furthermore, we also compared this model to several alternatives, including one with a single factor and three 2-factor models, each with two of the needs' items merged onto a common factor and the third need being standalone. We also conducted a 4-factor CFA which included the three fairness items as a fourth factor and a range of alternative 3-factor models with the fairness items loaded onto the three needs factors. Altogether, the results suggested that our hypothesized 3-factor model offered the best fit-parsimony trade-off, though, consistent with the correlations among the scale scores, the inter-factor correlations were strong, ranging from 0.66 to 0.80

(Bosco et al., 2015). The fit statistics for all models are summarized in Table 2.

Following the factor analyses, Cronbach's alphas were calculated for the needs measures with autonomy ($\alpha=0.90$), competence ($\alpha=0.91$), and relatedness ($\alpha=0.90$). Composite scores were then calculated for each factor as the mean response to the items.

Preliminary Analyses

Descriptive statistics and intercorrelations among all variables included in the analysis are calculated and presented in Table 3. Across all conditions and in absolute terms, the mean responses to the four reaction variables were all very slightly above the neutral point of the response scale, which is consistent with other research examining reactions to personality assessments (Anderson et al., 2010; Hausknecht et al., 2004). Consistent with the CFA, strong intercorrelations were observed between the three need-support variables. Fairness was also strongly positively correlated with the three needs. Small but significant negative correlations were found between years of education and perceptions of autonomy, competence, global fairness, but not with relatedness.

Hypothesis and Exploratory Testing

The mean scores and 95% confidence intervals for participants' perceptions of need-support and fairness across conditions are presented in Fig. 2. Prior to conducting pre-registered planned comparison tests, we first conducted a series of univariate ANOVAs to investigate the impact of four conditions on test-takers' perceptions of autonomy-support, $F(3, 1561)=36.01, p<0.001, \eta_p^2=0.06$; competence-support, $F(3, 1561)=28.93, p<0.001, \eta_p^2=0.05$;

Table 2 Model fit statistics from all alternative confirmatory factor analytic models of needs satisfaction and fairness items

Model title	χ^2	<i>df</i>	<i>RMSEA</i> (95% <i>CI</i>)	<i>CFI</i>	<i>SRMR</i>
Needs items only					
1 Factor: needs combined	1559.0	105	.134 (.128, .140)	.776	.081
2 Factors: relatedness and others combined	1117.3	103	.113 (.107, .119)	.844	.071
2 Factors: competence and others combined	1211.3	104	.117 (.111, .123)	.829	.071
2 Factors: autonomy and others combined	858.4	103	.097 (.091, .103)	.884	.056
3 Factors: three needs	426.1	101	.065 (.058, .071)	.950	.045
Needs and fairness items					
3 Factors: three needs fairness items loaded on autonomy	1328.1	149	.101 (.096, .106)	.869	.062
3 Factors: three needs fairness items loaded on competence	1133.0	149	.092 (.087, .097)	.891	.049
3 Factors: three needs fairness items loaded on relatedness	1238.3	149	.097 (.092, .102)	.879	.055
4 Factors: three needs + fairness	558.1	146	.060 (.055, .066)	.954	.044

$N=773$. Diagrams of all models and all output files are available in the online supplemental materials

RMSEA root mean square of error of approximation, *CFI* comparative fit index, *SMSR* standardized root mean square residual

Table 3 Means, standard deviations, and intercorrelations among demographic and dependent variables

Variable	Mean	SD	1	2	3	4	5	6	7
1. Age	40.97	12.20							
2. Participant is female	0.51	0.50	.08						
3. Years of formal education	16.05	2.96	.06	-.03					
4. Years working full-time	17.22	11.89	.85	.01	.03				
5. Autonomy satisfaction	3.15	1.07	.01	-.03	-.10	.03			
6. Competence satisfaction	3.11	0.99	-.04	-.02	-.13	-.01	.67		
7. Relatedness satisfaction	3.23	0.68	.00	-.01	-.08	-.01	.50	.63	
8. Global fairness perceptions	3.27	1.17	-.03	-.04	-.12	-.01	.69	.78	.63

Due to a small amount of missing demographic information, bivariate *N*s range from 1548 to 1565. For all $|r| \geq .05$, $p < .05$. For all $|r| \geq .07$, $p < .01$

relatedness-support, $F(3, 1561) = 2.62$, $p = 0.050$, $\eta_p^2 < 0.01$ and fairness $F(3, 1561) = 21.01$, $p < 0.001$, $\eta_p^2 = 0.04$. These results show that, as Fig. 2 implies, the format and design of the personality assessment affect perceptions of fairness and autonomy- and competence-support, but not relatedness-support. Furthermore, in absolute terms, the Likert format appeared to elicit somewhat *positive* reactions, with mean responses being somewhat higher than the neutral point of the scale. By contrast, irrespective of the condition, all three forced-choice formats appeared to elicit more neutral to only very slightly negative reactions, on average.

Next, we undertook four pre-registered planned contrasts and several additional unregistered contrasts; the results of

all contrast tests are shown in Table 4. In line with the pre-registration, for our hypothesis testing, we followed the Holm-Bonferroni (Holm, 1979) procedure to adjust alpha for multiple tests. The results, however, were not affected by the corrections, and for parsimony, we present only the uncorrected results here with full corrected results available in the supplement. The pre-registered contrasts compared, for autonomy- and competence-support, each final say format conditions separately, to the control condition. Compared to the control condition, the fixed final say format was associated with significantly higher perceptions of autonomy-support, competence-support, and fairness, and therefore Hypotheses H1a and H2a were supported (Contrast 1 in Table 4). The sizes of the differences (i.e., Cohen's

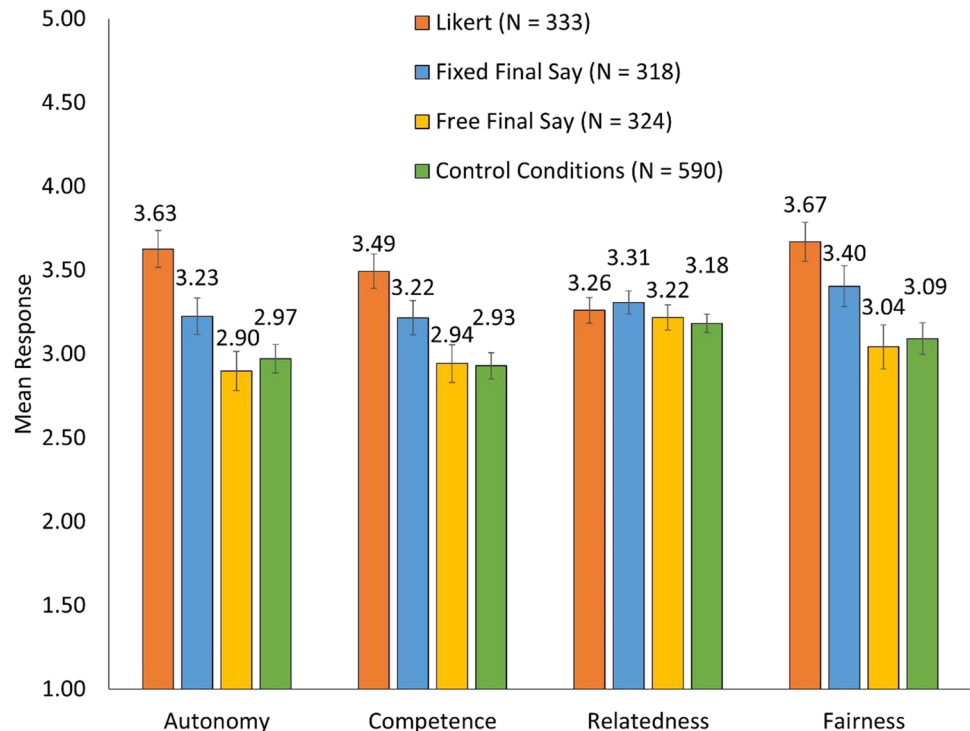
Fig. 2 Mean responses and 95% confidence intervals for dependent variables across experimental conditions

Table 4 Results of planned contrast tests

Contrast	Likert	Fixed	Free	Control	Autonomy support					Competence support				
					M_{Diff}	t	p	d_s	95% CI d_s	M_{Diff}	t	p	d_s	95% CI d_s
1	0	1	0	-1	.254**	3.51	<.001	.245	.109, .382	.287**	4.27	<.001	.303	.166, .440
2	0	0	1	-1	-.073	-1.01	.311	-.069	-.204, .067	.014	0.21	.836	.014	-.122, .150
3	1	0	0	-1	.655**	9.20	<.001	.626	.488, .763	.565**	8.53	<.001	.588	.451, .725
4	1	-1	0	0	.401**	4.92	<.001	.397	.242, .553	.278**	3.66	<.001	.295	.140, .449
	Likert	Fixed	Free	Control	Relatedness support					Fairness				
					M_{Diff}	t	p	d_s	95% CI d_s	M_{Diff}	t	p	d_s	95% CI d_s
1	0	1	0	-1	.125*	2.65	.008	.190	.054, .327	.313**	3.93	<.001	.275	.138, .412
2	0	0	1	-1	.035	0.75	.454	.052	-.084, .187	-.049	-0.62	.533	-.042	-.177, .094
3	1	0	0	-1	.078	1.68	.093	.114	-.020, .248	.578**	7.56	<.001	.510	.374, .646
4	1	-1	0	0	-.047	-0.88	.376	-.070	-.223, .084	.265*	2.95	.003	.241	.087, .395

d_s uses the pooled standard deviation for the conditions being compared as the denominator. df for all contrasts = 1561

* $p < .01$, ** $p < .001$

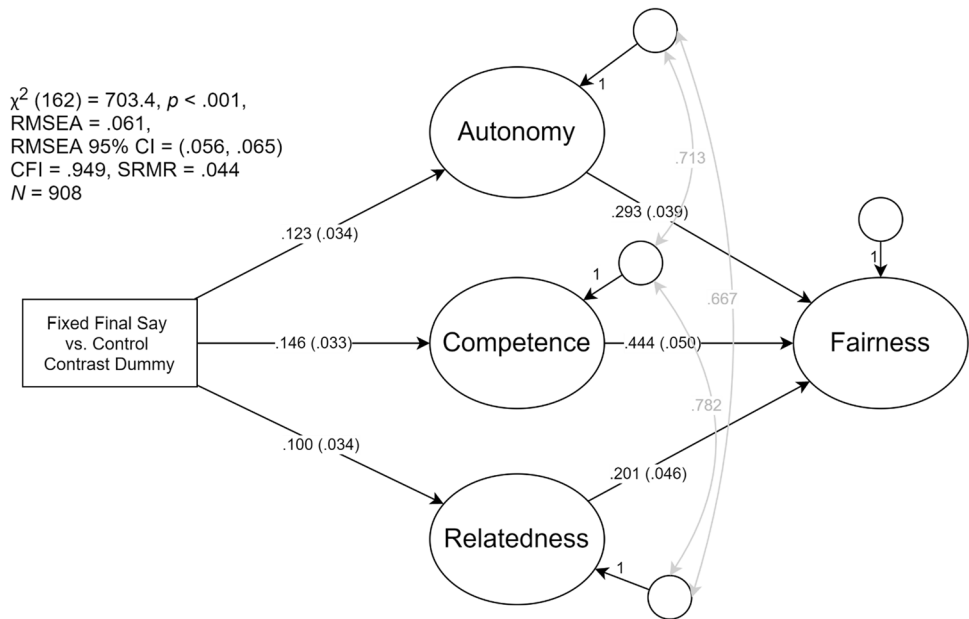
d_s ; Lakens, 2013) were around a quarter to three-tenths of a pooled standard deviation. By contrast, the free final say format was not associated with significantly different perceptions of autonomy- or competence-support than those observed in the control condition, thus H1b and H2b were not supported (Contrast 2).

The “practical” benefit of the fixed final say format is arguably best judged by examining the extent to which it can reduce the gap in reactions between the Likert format (arguably the “best case” scenario for personality assessments) and the traditional forced-choice format. To that end, we conducted two additional exploratory contrasts; these were not pre-registered. Contrast 3 (in Table 4) provides a baseline estimate of the differences in perceptions between a classical forced-choice format (the control conditions) and a Likert format; that is, this contrast reveals how large the forced-choice-Likert gap in reactions is. Contrast 4 shows the difference between the fixed final say and Likert formats. Therefore, the differences in effect sizes between Contrasts 3 and 4 estimates the extent to which the fixed final say modification “bridged” the gap between classical forced-choice and Likert formats. We did not examine relatedness support in these analyses because the differences between all formats were non-significant, and we had no hypothesis with respect to this outcome. We found that the fixed final say modification closed 36.6% of the gap between Likert and forced choice in perceptions of autonomy support and nearly half (49.8%) of the gap between perceptions of competence support. In terms of fairness, the fixed final say modification reduced the difference in perceptions between forced choice and Likert formats by over half (52.7%). Overall, while the fixed final say modification appeared to improve reactions, relative to the classical forced-choice format, in absolute terms, the

observed mean scores suggest that reactions to the modified format were still generally close to neutral rather than positive per se.

Finally, we undertook a structural equation model (SEM) to examine the indirect effect on fairness, through need-support, of the fixed final say format, relative to the control format. These analyses allow us to test the mediating role of autonomy (H3a) and competence (H3b) satisfaction but were not pre-registered. We also urge strong caution regarding causal inferences because need-support and fairness were all measured in the same section of the survey. We first created a dummy-coded variable that represented whether participants were in the control condition, or the fixed final say condition, thus excluding participants in other conditions (retained $N = 908$). In Mplus 8.5, using MLR estimation, we then specified a SEM with four factors (three needs and fairness), regressed the fairness factor onto the need factors and the need factors onto the contrast dummy variable. The overall fit for this model was moderate, per conventional criteria ($\chi^2(162) = 703.4$, $p < 0.001$, $RMSEA = 0.061$ (0.056, 0.065), $CFI = 0.949$, $SRMR = 0.044$). The standardized model coefficients are shown in Fig. 3. Indirect effects of the contrast dummy variable on fairness were estimated using 5000 bias-corrected bootstraps (Hayes, 2013). These analyses revealed that the total standardized indirect effect of the contrast dummy on fairness was 0.121 (95% CI = 0.066, 0.175). All three need-support specific indirect effects were significantly greater than zero. The largest was through competence need-support (standardized estimate = 0.065, 95% CI = 0.034, 0.101; 53.7% of the total indirect effect), then autonomy (standardized estimate = 0.036, 95% CI = 0.017, 0.062; 29.8% of the total indirect effect), then relatedness (standardized estimate = 0.020, 95% CI = 0.006, 0.040; 16.5% of the total indirect effect). Thus, H3a and H3b were supported.

Fig. 3 Structural equation model of fairness perceptions indirectly affected by forced-choice questionnaire format via needs support. Items are omitted to reduce clutter. All parameters are standardized and statistically significant at $p < .01$. Standard errors are shown in parentheses. $N = 908$



Discussion

Our aim in this research was to address the dilemma faced by organizations whereby the faking-resistant benefits offered by forced-choice personality measures come at a cost of unfavorable reactions, compared to single-stimulus formats (Dalal et al., 2019). Drawing on BPNT as a novel theoretical perspective, we proposed that forced-choice measures elicit these negative reactions because they (1) limit the freedom of applicants to choose their responses (i.e., undermining autonomy) and (2) limit the ability of applicants to demonstrate their strengths and competencies (i.e., undermining competence). We then designed and experimentally evaluated two modifications to the forced-choice assessment format that sought to improve the satisfaction of these two needs. One modification, “fixed final say”, allowed the respondent to identify personality statements the respondent had ranked first (last) while completing the forced-choice component as being poor (good) descriptors of their personality. The other modification, “free final say”, invited participants to use a free-text response box to describe how the forced-choice format may have led to an inaccurate assessment of their personalities. To our knowledge, this study represents the first attempt to develop test design interventions aimed at improving reactions to forced-choice personality measures using BPNT, or more broadly self-determination theory, as a guiding theory.

The results of this study provided additional evidence that, in a simulated job application scenario, reactions to a forced-choice personality questionnaire were less favorable than those to a single-stimulus format. Furthermore, the study provided evidence for the proposition that autonomy

and competence satisfaction represents mechanisms that drive the more negative applicant reactions to the forced-choice measure. In line with our expectations, participants who received the mechanistic “fixed final say” format reported higher perceptions of autonomy- and competence-support, compared to those who received the unmodified forced-choice measure. However, we found no evidence that the “free final say” format affected any participants’ reactions, relative to those of participants who did not receive any additional instructions following a forced-choice personality assessment. The primary difference between the intervention formats was the way participants were asked to respond; whereas the “fixed” format generated the actual responses of participants and asked them to evaluate each item via a checkbox, the “free” format required participants to *recall* their responses and evaluate their experience via text. On the one hand, the latter approach provides relatively more freedom in how participants may choose to respond, which could elicit greater perceptions of autonomy-support. On the other hand, responding in text form requires additional cognitive processing, compared to the checkbox-style responses and may be perceived as more difficult or effortful, potentially causing some frustration. Indeed, inspection of the text responses revealed that, rather than eliciting “corrections” to their personalities, the opportunity to respond instead prompted some participants to report on their generally negative experience of the questionnaire itself, potentially undermining any positive effects of the intervention. Thus, it may be that the instructions to focus on “correcting the record” were not clear enough to participants.

This study also contributes to the growing body of evidence that forced-choice questionnaires elicit more negative

reactions than single-stimulus questionnaires in job application contexts (Converse et al., 2008) and offers some explanations as to why (as per Dalal et al., 2019). Harland (2003) found that forced-choice personality measures, when used in a leadership development context, were rated lower in perceived accuracy, usefulness, test-taker control, and respectfulness than Likert counterparts. In the context of personnel selection, Converse et al. (2008) found that test-takers reacted more positively to Likert measures than forced-choice measures, but found no evidence that the differences in reactions were due to differences in perceptions of workload, test-taking expectancy, or belief in tests. The findings of this study show that test-takers perceived all variations of the forced-choice questionnaire, regardless of modification, to be less autonomy-supportive, less competence-supportive, and less fair than a Likert counterpart. However, in absolute terms, we must acknowledge that the reactions, overall, to the forced-choice format are perhaps best summarized as neutral rather than negative per se. Indeed, contrary to the research summarizing applicant reactions to personality questionnaires generally (e.g., Hausknecht et al., 2004), we found that reactions to the Likert format, in absolute terms, were generally quite favorable.

The “fixed final say” intervention appeared to bridge some of the gap between the Likert and forced-choice format with respect to perceptions of fairness, autonomy, and competency support, yielding some insights into why these formats are disliked. Specifically, it appears that participants’ negative reactions can be attributed to a combination of respondents feeling as though they were “forced” to respond in certain ways, and that these constraints limited their capacity to “sell” their strengths. In absolute terms, however, the intervention that aimed to loosen these constraints appears not to have led participants to experience the forced-choice format in a generally positive way overall, suggesting there may still be opportunities to further improve the forced-choice design.

When it comes to determining overall fairness perceptions (Gilliland, 1993), consistent with past research (Hausknecht et al., 2004; McCarthy et al., 2017a, 2017b), the satisfaction of the need for competence (analogous to the “chance to perform” justice rule) was a strong driver. Our indirect effect analyses showed, however, that the satisfaction of the needs for autonomy and relatedness — both of which are relatively underrepresented in the more dominant justice framework — may also promote stronger overall fairness perceptions. We must be cautious about making causal claims, given the cross-sectional nature of the mediator-outcome relation, but we recommend that future assessment design research consider these two variables in the development of new interventions.

Another contribution of this research is the development of a new measure that can be used to study need satisfaction

in a job application assessment context. Indeed, while many instruments for measuring the satisfaction of basic psychological needs do exist (Chen et al., 2015; Van den Broeck et al., 2010), none could be cleanly adapted to fit the requirements of this study. As such, we chose to develop and evaluate a measure of the degree to which elements of test design are supportive of the basic psychological needs of autonomy, competence, and relatedness. We were encouraged by the exploratory factor analytic results, validated by confirmatory analysis in an independent sample, which suggested that participants were able to distinguish between these items and revealed that the factor structure was as anticipated. Furthermore, we were also encouraged by the finding that the need-supporting interventions appeared not to affect the responses to the need for relatedness, giving some initial insight into the discriminant validity of the three scales. Nonetheless, we must recognize that the scales have not been validated elsewhere, and that further adaptations may be necessary for the items to fit with all assessment types or contexts. Thus, we encourage future researchers to investigate the scales’ validity in contexts beyond this study. Indeed, it is relatively common for measures of the three needs to be highly correlated as they were in this study (e.g., Chen et al., 2015; Van den Broeck et al., 2010), and so we suggest an appropriate construct validation strategy might be to demonstrate that changes on a single need-satisfaction scale are associated with an intervention that is designed to support that specific need only. Furthermore, we suggest that need satisfaction measures be validated against criteria in real-applicant settings (e.g., job acceptance, recommendation, or online review behaviors).

Theoretical Implications

This study also provides some novel theoretical insights into what drives negative applicant reactions to assessment procedures. In articulating these, we must first recognize there are clear conceptual overlaps between some of the content captured by the basic psychological needs and procedural justice rules. For example, the “chance to perform” rule, operationalized in the selection procedural justice scale (Bauer et al., 2006), is functionally very similar to the satisfaction of the competence need in the context of assessing applicant reactions. Other justice rules also appear to have some conceptual overlap with the other needs. The *two-way communication*, *respect*, and *openness* rules show some correspondence with the concept of the satisfaction of the need for relatedness. We note that for many assessment tools or formats (e.g., online cognitive testing or video interviewing), these rules may simply not be applicable, or it is unclear how the rules might lead to reaction-improving interventions. By contrast, we anticipate that there might be opportunities to better satisfy

the need for relatedness when completing these types of assessments through interventions that do not necessarily focus on improving two-way communication, respect, or openness.

Similarly, the satisfaction of the need for autonomy shares some overlap with Gilliland's (1993) procedural justice rules. For example, justice theory has informed a class of interventions that involve providing information or explanations about assessments to candidates (McCarthy et al., 2018, 2017a, 2017b; Truxillo et al., 2002). In BPNT terms, providing people with a rationale is known to be an effective autonomy-supportive intervention (Steingut et al., 2017). We see, however, other opportunities to design autonomy-supportive interventions that do not have clear counterparts within the justice framework (Gagné et al., 2022). For example, here, we demonstrated how reactions can be improved by increasing autonomy by reducing some constraints in the forced-choice format, suggesting the restriction of freedom is part of the mechanism which drives negative reactions to forced-choice measures. We also note recent research that has demonstrated how promoting autonomy-supportive features of asynchronous video interviews has improved perceptions of these assessments (Basch & Melchers, 2019).

Adopting the BPNT lens to understand reactions to personality assessments may also promote further insights into the effects of other features of forced-choice assessment design. For example, one important variable feature is the nature of the items that are blocked together when the questionnaire is constructed. Typically, test designers aim to select items for a common block that share similar levels of social desirability, such that there are no "stand-out" items that a faker would obviously rank first or last (Feldman & Corah, 1960; Pavlov et al., 2021). One could anticipate, however, that having to face many blocks of items that are very closely matched on desirability could undermine candidates' autonomy and competence and that negative reactions could be buffered by including several relatively poorly-matched item blocks. Similarly, blocks can also vary with respect to the trait content of the items therein, and perhaps the extent that certain traits are mixed together (i.e., forcing respondents to sacrifice one to benefit the other) or kept apart from one another when blocks are constructed could have implications for the satisfaction of the need for competence.

The BPNT lens may also be useful for understanding how the forced-choice format might trigger different reactions among fakers when compared to non-fakers. It seems plausible that those with stronger intentions to fake would feel more strongly that their autonomy is being undermined by the forced-choice format than those who do not intend to fake because it is the fakers who most desire the freedom to endorse all of the most desirable items they encounter. By contrast, it also seems plausible that non-fakers would feel most strongly that their competence is being undermined by

the forced-choice format because the format prevents these respondents from demonstrating their (legitimate) strengths. In both cases, the fixed final say modification may provide an avenue to better understand these mechanisms, as fakers may tend to flag as many socially desirable items as they can, whereas non-fakers may be relatively more focused on flagging items that describe personality-based skills.

Practical Implications

From a practical perspective, our findings suggest that an avenue to improve the experience of completing a forced-choice personality assessment may be to incorporate the fixed final say modification. The observed improvements in reactions were modest in absolute terms. Thus, on a small scale (e.g., for small to medium employers or boutique test developers), it may be that the improvements in reactions borne from the modifications tested here are practically trivial. By contrast, for larger employers or test developers who routinely assess high volumes of candidates, even small improvements to reactions to forced-choice measures are likely very welcome, especially if the costs of implementation are small. Indeed, as Funder & Ozer (2019) explain, small effects combined with scale, can quickly accumulate into large and meaningful consequences, thus we feel there is reason to be optimistic about modest interventions to assessment design that yield promising effects.

Nonetheless, there remain practical and ethical considerations that require further thought. First, from a practical perspective, the fixed final say format added approximately 3 min to the assessment time on average. While Speer et al. (2016) found evidence that testing duration appears not to have a major influence on test reactions, recruiters may nonetheless be concerned about the increased test length and its impact on the candidate experience.

Second, both practically and ethically, there remains the question of what should be done with a candidate's "final say" responses. Indeed, candidates will likely expect that their final-say responses would be incorporated in the selection decision making, and yet the psychometrics of the forced-choice format do not suggest any obvious way to incorporate those responses into trait scores. We suggest that one possible way for decision makers to use the final say data is to regard it as applicant feedback or commentary that accompanies the trait scores rather than influences them. For example, a list of the statements marked by the candidate could appear in the report that contains the applicants' personality scores, ensuring that selection decision makers receive this additional feedback from the candidates, while preserving the trait scores. While such an approach would ensure the candidates' final say responses are not "wasted," it introduces concerns

about how decision makers might evaluate the final say responses, and whether such evaluations can be undertaken objectively. In principle, with large sample sizes, it may be possible to determine whether the responses to the fixed final say component provide diagnostic or psychometric information that could be used to systematically adjust trait scores, but until this is established, sharing the final say responses may be too risky to implement in practice.

Finally, one anonymous reviewer noted that an applicant who has indicated that the test did not describe them (e.g., checking a lot of, or all, the boxes in the fixed final say activity), and is not hired may feel there are grounds to complain about the assessment that apparently did not describe them well. Thus, overall, the fixed final say modification may require further development before being implementable in practice.

Limitations and Directions for Future Research

We must be careful to put the findings of this investigation into context of its limitations. We reiterate that the use of a novel measure carries risks with respect to construct and criterion validity. Indeed, in designing the study, we closely examined alternative commonly used measures of applicant reactions and discovered that very few of them can be sensibly adapted to an experiment involving a hypothetical job application, thus it was not possible to find useful criteria against which to validate the measure. We suggest that future research could consider adopting a lightly-adapted version of the measure to study reactions to other types of assessment situations, or in relation to other interventions that target different needs.

We also must recognize the difference between a hypothetical experiment and a true job application that may have affected the pattern of results we observed. Because the need-satisfaction measures prompted participants to reflect on experience of completing the assessment itself, rather than reflecting on the wider (hypothetical) job application itself, we are optimistic that it may function equivalently in the field. The lack of a job at the end of the “selection” process, however, makes it difficult to measure outcome variables which are important to organizations (e.g., organizational attractiveness, recommendation intentions, and job acceptance behaviors; Chapman et al., 2005). Indeed, Hausknecht (2013) noted that actual applicants have more invested in the hiring process and may therefore be more sensitive to particular features of selection and testing. Thus, it remains unclear from this study whether intervening at the level of a specific assessment’s design is a fruitful way to reduce the chances that a candidate’s negative experience leads to a reluctance to accept a job offer or reapply to the same organization in the future. Nonetheless, we view

improved assessment design as an area in which test developers can contribute to optimizing the experiences of candidates, potentially offering a competitive advantage.

A final limitation of the study was that the personality assessment was neither designed to be psychometrically rigorous nor the item blocks were matched on social desirability to the same extent, as one would expect from a purpose-built assessment. It therefore remains possible that the effects observed here reflect idiosyncratic features of the questionnaire we used. Accordingly, we encourage test developers wishing to assess the impact of response format on applicant reactions to their own assessment to consider the research design presented here.

Conclusion

This paper provides additional evidence that forced-choice personality measures elicit unfavorable reactions in comparison to single-stimulus measures in simulated job application settings. Preliminary psychometric evidence for a new basic psychological need scale provides a foundation for future research and may prove to be a useful tool for understanding the hierarchy of reactions in assessment formats. Further research with self-determination theory and alternative theoretical approaches is still needed to fully understand the mechanism which drives this hierarchy of reactions. Broadly speaking, this research is valuable as it could improve the selection process for both candidates and hiring organizations.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10869-023-09876-w>.

Acknowledgements The authors acknowledge Angus Hughes for his assistance with the pre-registration.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

Data Availability Raw data are available for download from the project's OSF website: <https://osf.io/j8wdy/>.

Declarations

Conflict of Interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ackerman, P. L., & Kanfer, R. (2009). Test length and cognitive fatigue: An empirical examination of effects on performance and test-taker reactions. *Journal of Experimental Psychology: Applied*, *15*(2), 163–181. <https://doi.org/10.1037/a0015719>
- Anderson, N., Salgado, J. F., & Hulsheger, U. R. (2010). Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment*, *18*(3), 291–304. <https://doi.org/10.1111/j.1468-2389.2010.00512.x>
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*(1), 1–26. <https://doi.org/10.1111/j.1744-6570.1991.tb00688.x>
- Bartholomew, K. J., Ntoumanis, N., Ryan, R. M., & Thøgersen-Ntoumani, C. (2011). Psychological need thwarting in the sport context: Assessing the darker side of athletic experience. *Journal of Sport and Exercise Science*, *33*(1), 75–102. <https://doi.org/10.1123/jsep.33.1.75>
- Basch, J. M., & Melchers, K. G. (2019). Fair and flexible?! Explanations can improve applicant reactions toward asynchronous video interviews. *Personnel Assessment and Decisions*, *5*(3), 2. <https://doi.org/10.25035/pad.2019.03.002>
- Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). *Personnel Psychology*, *54*(2), 388–420. <https://doi.org/10.1111/j.1744-6570.2001.tb00097.x>
- Berkshire, J. R. (1958). Comparisons of five forced-choice indices. *Educational and Psychological Measurement*, *18*(3), 553–561. <https://doi.org/10.1177/001316445801800309>
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, *14*(4), 317–335. <https://doi.org/10.1111/j.1468-2389.2006.00354.x>
- Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. *Journal of Applied Psychology*, *100*, 431–449. <https://doi.org/10.1037/a0038047>
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research*, *21*(2), 230–258. <https://doi.org/10.1177/0049124192021002005>
- Buil, I., Catalán, S., & Martínez, E. (2020). Understanding applicants' reactions to gamified recruitment. *Journal of Business Research*, *110*, 41–50. <https://doi.org/10.1016/j.jbusres.2019.12.041>
- Cao, M., & Drasgow, F. (2019). Does forcing reduce faking? A meta-analytic review of forced-choice personality measures in high-stakes situations. *Journal of Applied Psychology*, *104*(11), 1347–1368. <https://doi.org/10.1037/apl0000414>
- Chapman, D. S., Uggerslev, K. L., Carroll, S. A., Piasentin, K. A., & Jones, D. A. (2005). Applicant attraction to organizations and job choice: A meta-analytic review of the correlates of recruiting outcomes. *Journal of Applied Psychology*, *90*(5), 928–944. <https://doi.org/10.1037/0021-9010.90.5.928>
- Chen, B., Vansteenkiste, M., Beyers, W., Boone, L., Deci, E. L., Van der Kaap-Deeder, J., Duriez, B., Lens, W., Matos, L., Mouratidis, A., Ryan, R. M., Sheldon, K. M., Soenens, B., Van Petegem, S., & Verstuyf, J. (2015). Basic psychological need satisfaction, need frustration, and need strength across four cultures. *Motivation and Emotion*, *39*(2), 216–236. <https://doi.org/10.1007/s11031-014-9450-1>
- Converse, P. D., Oswald, F. L., Imus, A., Hedricks, C., Roy, R., & Butera, H. (2008). Comparing personality test formats and warnings: Effects on criterion-related validity and test-taker reactions. *International Journal of Selection and Assessment*, *16*(2), 155–169. <https://doi.org/10.1111/j.1468-2389.2008.00420.x>
- Dalal, D. K., Zhu, X., Rangel, B., Boyce, A. S., & Lobene, E. (2019). Improving applicant reactions to forced-choice personality measurement: Interventions to reduce threats to test takers' self-concepts. *Journal of Business and Psychology*, *36*(1), 55–70. <https://doi.org/10.1007/s10869-019-09655-6>
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01
- Donovan, J. J., Dwight, S. A., & Schneider, D. (2014). The impact of applicant faking on selection measures, hiring decisions, and employee performance. *Journal of Business and Psychology*, *29*(3), 479–493. <https://doi.org/10.1007/s10869-013-9318-5>
- Dunlop, P. D., Holtrop, D., Ashby, L. M., Bharadwaj, A., & Donovan, J. J. (2022). Valence, instrumentality, expectancy, and ability as determinants of faking, and the effects of faking on criterion-related validity. *Journal of Business and Psychology*, *37*, 1215–1233. <https://doi.org/10.1007/s10869-022-09797-0>
- Feldman, M. J., & Corah, N. L. (1960). Social desirability and the forced choice method. *Journal of Consulting Psychology*, *24*(6), 480–482. <https://doi.org/10.1037/h0042687>
- Funster, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. <https://doi.org/10.1177/2515245919847202>
- Gagné, M., Parker, S. K., Griffin, M. A., Dunlop, P. D., Knight, C., Klonek, F. E., & Parent-Rocheleau, X. (2022). Understanding and shaping the future of work with self-determination theory. *Nature Reviews Psychology*. <https://doi.org/10.1038/s44159-022-00056-w>
- Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *The Academy of Management Review*, *18*(4), 694–734. <https://doi.org/10.2307/258595>
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R. T., Ashton, M. C., Cloninger, C., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*(1), 84–96. <https://doi.org/10.1016/j.jrp.2005.08.007>
- Gordon, L. V. (1951). Validities of the forced-choice and questionnaire methods of personality measurement. *Journal of Applied Psychology*, *35*(6), 407–412. <https://doi.org/10.1037/h0058853>
- Gunnell, K. E., Crocker, P. R. E., Wilson, P. M., Mack, D. E., & Zumbo, B. D. (2013). Psychological need satisfaction and thwarting: A test of Basic Psychological Needs Theory in physical activity contexts. *Psychology of Sport and Exercise*, *14*(5), 599–607. <https://doi.org/10.1016/j.psychsport.2013.03.007>
- Harland, L. K. (2003). Using personality tests in leadership development: Test format effects and the mitigating impact of explanations and feedback. *Human Resource Development Quarterly*, *14*(3), 285–301. <https://doi.org/10.1002/hrdq.1067>
- Hausknecht, J. P. (2013). Applicant reactions. In K. Y. T. Yu & D. M. Cable (Eds.), *The Oxford handbook of recruitment* (pp. 35–46). Oxford University Press.
- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, *57*(3), 639–683. <https://doi.org/10.1111/j.1744-6570.2004.00003.x>

- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. Guildford Press.
- Hogan, R. T. (2005). In defense of personality measurement: New wine for old whiners. *Human Performance*, 18(4), 331–341. https://doi.org/10.1207/s15327043hup1804_1
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.
- Hough, L. M., & Oswald, F. L. (2008). Personality testing and industrial organizational psychology: Reflections, progress, and prospects. *Industrial and Organizational Psychology*, 1(3), 272–290. <https://doi.org/10.1111/j.1754-9434.2008.00048.x>
- Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hu, J., & Connelly, B. S. (2021). Faking by actual applicants on personality tests: A meta-analysis of within-subjects studies. *International Journal of Selection and Assessment*, 29(3–4), 412–426. <https://doi.org/10.1111/ijasa.12338>
- Huber, C. R., Kuncel, N. R., Huber, K. B., & Boyce, A. S. (2021). Faking and the validity of personality tests: An experimental investigation using modern forced choice measures. *Personnel Assessment and Decisions*, 7(1), 3. <https://doi.org/10.25035/pad.2021.01.003>
- Hughes, A. W., Dunlop, P. D., Holtrop, D., & Wee, S. (2021). Spotting the “ideal” personality response: Effects of item matching in forced choice measures for personnel selection. *Journal of Personnel Psychology*, 20(1), 17–26. <https://doi.org/10.1027/1866-5888/a000267>
- Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology*, 85(6), 869–879. <https://doi.org/10.1037/0021-9010.85.6.869>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Litman, L., Robinson, J., & Abberbock, T. (2016). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior research methods*, 49(2), 433–442. <https://doi.org/10.3758/s13428-016-0727-z>
- Longo, Y., Gunz, A., Curtis, G. J., & Farsides, T. (2016). Measuring need satisfaction and frustration in educational and work contexts: The Need Satisfaction and Frustration Scale (NSFS). *Journal of Happiness Studies*, 17(1), 295–317. <https://doi.org/10.1007/s10902-014-9595-3>
- Marcus, B. (2009). ‘Faking’ from the applicant’s perspective: A theory of self-presentation in personnel selection settings. *International Journal of Selection and Assessment*, 17(4), 417–430. <https://doi.org/10.1111/j.1468-2389.2009.00483.x>
- Marcus, B. (2022). “Faking” is neither good nor bad, it is a misleading concept: A reply to Tett and Simonet (2021). *Personnel Assessment and Decisions*, 8(1), 4.
- McCarthy, J. M., Bauer, T. N., Truxillo, D. M., Anderson, N. R., Costa, A. C., & Ahmed, S. M. (2017). Applicant perspectives during selection: A review addressing “So what?”, “What’s new?”, and “Where to next?” *Journal of Management*, 43(6), 1693–1725. <https://doi.org/10.1177/0149206316681846>
- McCarthy, J. M., Bauer, T. N., Truxillo, D. M., Campion, M. C., Van Iddekinge, C. H., & Campion, M. A. (2017). Using pre-test explanations to improve test-taker reactions: Testing a set of “wise” interventions. *Organizational Behavior and Human Decision Processes*, 141, 43–56. <https://doi.org/10.1016/j.obhdp.2017.04.002>
- McCarthy, J. M., Bauer, T. N., Truxillo, D. M., Campion, M. C., Van Iddekinge, C. H., & Campion, M. A. (2018). Improving the candidate experience: Tips for developing ‘Wise’ organizational hiring interventions. *Organizational Dynamics*, 47(3), 147–154. <https://doi.org/10.1016/j.orgdyn.2018.05.004>
- McFarland, L. A. (2013). Applicant reactions to personality tests: Why do applicants hate them? In N. D. Christiansen & R. P. Tett (Eds.), *Handbook of personality at work* (pp. 281–298). Routledge. <https://doi.org/10.4324/9780203526910.ch13>
- Michaels, E., Handfield-Jones, H., & Axelrod, B. (2001). *The war for talent*. Harvard Business Press.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, 60(3), 683–729. <https://doi.org/10.1111/j.1744-6570.2007.00089.x>
- Nikolaou, I., & Judge, T. A. (2007). Fairness reactions to personnel selection techniques in Greece: The role of core self-evaluations. *International Journal of Selection and Assessment*, 15(2), 206–219. <https://doi.org/10.1111/j.1468-2389.2007.00382.x>
- Nikolaou, I., Bauer, T. N., & Truxillo, D. M. (2015). Applicant reactions to selection methods An overview of recent research and suggestions for the future. In I. Nikolaou & J. K. Oostrom (Eds.), *Employee recruitment, selection, and assessment Employee recruitment, selection, and assessment* (pp. 92–108). Psychology Press.
- Ntoumanis, N., Ng, J. Y. Y., Prestwich, A., Quested, E., Hancox, J. E., Thøgersen-Ntoumani, C., Deci, E. L., Ryan, R. M., Lonsdale, C., & Williams, G. C. (2021). A meta-analysis of self-determination theory-informed intervention studies in the health domain: Effects on motivation, health behavior, physical, and psychological health. *Health Psychology Review*, 15(2), 214–244. <https://doi.org/10.1080/17437199.2020.1718529>
- Patall, E. A., Cooper, H., & Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: A meta-analysis of research findings. *Psychological Bulletin*, 134(2), 270–300. <https://doi.org/10.1037/0033-2909.134.2.270>
- Paterson, T. A., Harms, P. D., Steel, P., & Credé, M. (2016). An assessment of the magnitude of effect sizes: Evidence from 30 years of meta-analysis in management. *Journal of Leadership and Organizational Studies*, 23(1), 66–81. <https://doi.org/10.1177/1548051815614321>
- Pavlov, G., Shi, D., Maydeu-Olivares, A., & Fairchild, A. (2021). Item desirability matching in forced-choice test construction. *Personality and Individual Differences*, 183, 111114. <https://doi.org/10.1016/j.paid.2021.111114>
- Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, 46(4), 1023–1031. <https://doi.org/10.3758/s13428-013-0434-y>
- Peterson, M. H., Griffith, R. L., Isaacson, J. A., O’Connell, M. S., & Mangos, P. M. (2011). Applicant faking, social desirability, and the prediction of counterproductive work behaviors. *Human Performance*, 24(3), 270–290. <https://doi.org/10.1080/08959285.2011.580808>
- Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, 43(3), 450–461. <https://doi.org/10.1037/0022-3514.43.3.450>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Publications.
- Ryan, A. M., & Huth, M. (2008). Not much more than platitudes? A critical look at the utility of applicant reactions research. *Human*

- Resource Management Review*, 18(3), 119–132. <https://doi.org/10.1016/j.hrmr.2008.07.004>
- Ryan, A. M., & Ployhart, R. E. (2000). Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management*, 26(3), 565–606. <https://doi.org/10.1177/014920630002600308>
- Rynes, S. L., & Connerley, M. L. (1993). Applicant reactions to alternative selection procedures. *Journal of Business and Psychology*, 7(3), 261–277. <https://doi.org/10.1007/bf01015754>
- Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2021). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology*, 107(11), 2040–2068. <https://doi.org/10.1037/apl0000994>
- Salgado, J. F., Anderson, N., & Tauriz, G. (2015). The validity of ipsative and quasi-ipsative forced-choice personality inventories for different occupational groups: A comprehensive meta-analysis. *Journal of Occupational and Organizational Psychology*, 88(4), 797–834. <https://doi.org/10.1111/joop.12098>
- Sass, R., Frick, S., Reips, U.-D., & Wetzel, E. (2018). Taking the test taker's perspective: Response process and test motivation in multidimensional forced-choice versus rating scale instruments. *Assessment*, 27(3), 572–584. <https://doi.org/10.1177/1073191118762049>
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262–274. <https://doi.org/10.1037/0033-2909.124.2.262>
- Serrat, O. (2017). *A primer on talent management*. Springer. https://doi.org/10.1007/978-981-10-0983-9_44
- Seybert, J., & Becker, D. (2019). Examination of the test–retest reliability of a forced-choice personality measure. *ETS Research Report Series*, 2019(1), 1–17. <https://doi.org/10.1002/ets2.12273>
- Soto, C. J., Napolitano, C. M., & Roberts, B. W. (2021). Taking skills seriously: Toward an integrative model and agenda for social, emotional, and behavioral skills. *Current Directions in Psychological Science*, 30(1), 26–33. <https://doi.org/10.1177/0963721420978613>
- Speer, A. B., King, B. S., & Grossenbacher, M. (2016). Applicant reactions as a function of test length: Is there reason to fret over using longer tests? *Journal of Personnel Psychology*, 15(1), 15–24. <https://doi.org/10.1027/1866-5888/a000145>
- Steingut, R. R., Patall, E. A., & Trimble, S. S. (2017). The effect of rationale provision on motivation and performance outcomes: A meta-analysis. *Motivation Science*, 3(1), 19–50. <https://doi.org/10.1037/mot0000039>
- Su, Y.-L., & Reeve, J. (2011). A meta-analysis of the effectiveness of intervention programs designed to support autonomy. *Educational Psychology Review*, 23(1), 159–188. <https://doi.org/10.1007/s10648-010-9142-7>
- Tett, R. P., & Simonet, D. V. (2021). Applicant faking on personality tests: Good or bad and why should we care? *Personnel Assessment and Decisions*, 7(1), 2. <https://doi.org/10.25035/pad.2021.01.002>
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, 44(4), 703–742. <https://doi.org/10.1111/j.1744-6570.1991.tb00696.x>
- Tett, R. P., Simonet, D. V., & Christiansen, N. D. (2022). Faking is as faking does: A rejoinder to Marcus (2021). *Personnel Assessment and Decisions*, 8(1), 5. <https://doi.org/10.25035/pad.2022.01.005>
- Truxillo, D. M., Bauer, T. N., Campion, M. A., & Paronto, M. E. (2002). Selection fairness information and applicant reactions: A longitudinal field study. *Journal of Applied Psychology*, 87(6), 1020–1031. <https://doi.org/10.1037/0021-9010.87.6.1020>
- Van den Broeck, A., Vansteenkiste, M., De Witte, H., Soenens, B., & Lens, W. (2010). Capturing autonomy, competence, and relatedness at work: Construction and initial validation of the Work-related Basic Need Satisfaction scale. *Journal of Occupational and Organizational Psychology*, 83(4), 981–1002. <https://doi.org/10.1348/096317909X481382>
- Vasconcellos, D., Parker, P. D., Hilland, T., Cinelli, R., Owen, K. B., Kapsal, N., Lee, J., Antczak, D., Ntoumanis, N., Ryan, R. M., & Lonsdale, C. (2020). Self-determination theory applied to physical education: A systematic review and meta-analysis. *Journal of Educational Psychology*, 112(7), 1444–1469. <https://doi.org/10.1037/edu0000420>
- Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 59(2), 197–210. <https://doi.org/10.1177/00131649921969802>
- Wetzel, E., Frick, S., & Brown, A. (2021). Does multidimensional forced-choice prevent faking? Comparing the susceptibility of the multidimensional forced-choice format and the rating scale format to faking. *Psychological Assessment*, 33(2), 156–170. <https://doi.org/10.1037/pas0000971>
- Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., & Anderson, N. R. (2019). Personnel selection in the digital age: A review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology*, 29(1), 64–77. <https://doi.org/10.1080/1359432X.2019.1681401>
- Zhang, B., Sun, T., Drasgow, F., Chernyshenko, O. S., Nye, C. D., Stark, S., & White, L. A. (2020). Though forced, still valid: Psychometric equivalence of forced-choice and single-statement measures. *Organizational Research Methods*, 23(3), 569–590. <https://doi.org/10.1177/1094428119836486>
- Ziegler, M., MacCann, C., & Roberts, R. D. (2011). Faking: Knowns, unknowns, and points of contention. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 3–17). Oxford University Press Inc.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.