



Self-Presentation in Selection Settings: the Case of Personality Tests

Bernd Marcus¹ · Judy Goldenberg² · Saul Fine² · Henning Hummert¹ · Anne Traum¹

Published online: 13 July 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Based on theoretical views that applicants express meaningful skills and motivation when presenting themselves in personnel selection settings, we challenge conventional wisdom that self-presentation necessarily impairs the diagnostic value of “fakable” selection devices. Instead, we propose to supplement the traditional psychometric approach to personnel selection with a social perspective that leverages the competitive nature of selection. In order to capture an outcome of self-presentation, we introduce the Ideal Employee Coefficient (IEC) as a supplement to traditional scoring of responses to personality items. Construct and criterion-related validity evidence using the IEC was collected in two studies covering three samples from diverse settings, populations, and measures. The IEC consistently showed incremental criterion-related validity beyond the same tests’ traditional scores, as well as construct-related evidence in line with theoretical underpinnings. Findings imply that traditional personality constructs can be meaningfully aggregated with measures of self-presentation that are cost-effectively derived from the same data sources.

Keywords Personnel selection · Self-presentation · Faking · Personality assessment

A very common concern in the practice of personnel selection is that job applicants may distort the information they are willing to convey about themselves in order to receive a job offer. In fact, research appears to corroborate such concerns. For example, it is well established empirically that work-related settings affect how test takers respond to personality items. Setting effects on mean scores, in the socially desirable direction, have been shown in laboratory faking experiments

(Viswesvaran & Ones, 1999), in real-world applicant settings (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006), and even in low-stakes situations where work-related framing was added to standard personality items (Schmit, Ryan, Stierwalt, & Powell, 1995). Yet, the consequences of such setting effects, for instance for predictive validity, remain subject to controversial debates. Whereas some scholars have argued that setting effects on personality scores could be largely ignored for practical purposes (e.g., Ones, Viswesvaran, & Reiss, 1996), others (e.g., Morgeson et al., 2007; Tett & Christiansen, 2007) remain concerned about detrimental setting effects and may even abandon the use of instruments vulnerable to such effects for personnel selection. Common labels such as “faking” or “response distortion” reflect such more pessimistic views. In the present paper, we prefer to use the broader and less value-laden term “self-presentation” (cf. Goffman, 1959) instead.

Still different positions include identification of potentially valid components of self-presentation (e.g., Johnson & Hogan, 2006; Kleinmann et al., 2011) and weighing possible positive effects against potential detriments due to self-presentation (e.g., Marcus, 2009; Tett & Simonet, 2011). A common theme among those latter perspectives is that applicants’ successful self-presentation to some extent may reflect job-relevant skills and motivation and, thus, predict job

✉ Bernd Marcus
bernd.marcus@uni-rostock.de

Judy Goldenberg
judyg@zahav.net.il

Saul Fine
finesaul@gmail.com

Henning Hummert
henning.hummert@uni-rostock.de

Anne Traum
anne.traum@uni-rostock.de

¹ Institute of Business Administration, University of Rostock, Ulmenstr. 69, 18057 Rostock, Germany

² Behavioral Sciences Department, Israel Defense Forces, Tel HaShomer, Israel

performance. Whereas such potentially valid elements of self-presentation have been used to explain why the predictive validity of fable selection instruments is—accidentally—retained in actual applicant settings, surprisingly little attempt has been made to proactively utilize this potential. As Tett and Christiansen (2007) put it: “It has not been shown [...] that faking predicts future job success.” (p. 985). However, some initial evidence in that direction was recently reported by Ingold, Kleinmann, König, and Melchers (2015), who found faking (as measured by difference scores between simulated applicant and honest conditions) to be positively related to supervisor ratings of job performance, though in a relatively small ($N=92$) sample.

In the present study, we offer theoretical arguments and some empirical evidence for a more balanced view on self-presentation. In the theoretical part, we offer a new perspective on self-presentation as a potential source of validity in selection, which is based on what we consider the ambiguous nature of selection settings from the applicant’s perspective. Whereas applicants are typically told in instructions to act as if they were not applying for a job (i.e., to be “honest”), they actually do apply for a job and are likely to be aware of that fact. We argue that taking those conflicting demands into account is key to utilizing the potential of self-presentation and we offer a social perspective on selection settings as a supplement to the traditional psychometric view on selection. In our empirical part, we then chose personality testing as an exemplary case for exploring the validity potential of the social perspective. For that purpose, we introduce a method aimed at complementing psychometric personality scores by scoring test taker responses with a focus on social expectations. We present and discuss findings from two empirical studies covering three samples aimed at testing the incremental validity of our newly introduced scoring method beyond traditional scoring.

Valid Elements of Self-Presentation

Selection instruments are typically not designed to measure self-presentation but for measuring constructs such as personality traits, abilities, or various kinds of skills and knowledge deemed relevant for a particular job. In fact, job relevance is at the core of professional standards for evaluating selection devices (Society for Industrial and Organizational Psychology [SIOP], 2018). Critics of instruments vulnerable to self-presentation argued that measurement of actual target constructs is impaired in the presence of self-presentation, maybe even at the expense of measuring undesirable constructs such as Machiavellianism or low integrity (e.g., McFarland & Ryan, 2000; Roulin, Krings, & Binggeli, 2016; Tett & Simonet, 2011). If, on the other hand, there is a positive potential in self-presentation, those negative effects would need

to be counterbalanced by valid components being reflected in self-presentation behavior.

Virtually all theories of self-presentation or faking specify that both the actor’s motivation and some sort of skill set (or ability, or capacity) are present in self-presentation behavior and success (in terms of obtaining a job offer) (e.g., Levashina & Campion, 2006; Marcus, 2009; McFarland & Ryan, 2000; Roulin et al., 2016; Tett & Simonet, 2011). In line with this general notion, several authors proposed elements of self-presentation skills or motivation, or both, that may be positively related to job performance. On the skill side, Kleinmann et al. (2011) introduced the construct “ability to identify criteria” (ATIC) as the general ability to correctly read situational cues and to infer constructs being measured in selection, and they present evidence that ATIC is positively related to scores on a range of selection instruments. On the motivation side, Tett and Simonet (2011) highlight ambition as one potential factor underlying differences in self-presentation. These are two examples out of a longer list of stable characteristics that may be activated in high-stakes situations and also be relevant for later achievements on the job (for a recent example, see Pelt, van der Linden, & Born, 2018, on emotional intelligence). More broadly speaking, applicants need to be able and willing to be successful with their self-presentation, and those who are may later be successful in their job to the extent the same factors contribute to performance there. After extracting a general factor from personality items present solely in applicant samples, Schmit and Ryan (1993) were probably the first to coin the term “ideal-employee factor” for a composite of trait terms desirable in almost any job.

Applying socio-analytic theory to self-presentation, Hogan and colleagues (e.g., Johnson & Hogan, 2006) moved perhaps one step further than the aforementioned authors. Socio-analytic theory uses the terms personality and self-presentation almost synonymously. People convey a certain image of themselves in order to achieve the goals of getting ahead and getting along with others (or agency and communion, Bakan, 1966). Although this image partially varies across situations, what people consider desirable depends upon their relatively stable self-image, and they generally tend to strive for consistency across situations. Test taking is considered just one of many opportunities to tell others how one wants to be seen. The retained validity of personality tests in selection settings is essentially explained by stating that “the factors that facilitate skillful self-presentation in everyday life might also apply to competent self-presentation on personality tests” (Johnson & Hogan, 2006, p. 217).

Marcus’ (2009) theory of self-presentation builds on some of the above ideas but extends them by introducing elements of self-presentation more specifically tailored to the context of personnel selection. The theory defines self-presentational skills as the degree to which applicants understand the

prospective employer's expectations. Although similar to the ATIC construct of Kleinmann et al. (2011), this skill set does not imply that applicants translate questions in interviews or tests into abstract personality constructs but rather refers to expectations of actual partners in social interactions. Motivation to self-present is defined as the degree to which applicants are willing to adapt to the expectations they perceive. In addition to socio-analytic theory's focus on cross-situational consistency, Marcus (2009) proposes that the larger the discrepancy between an applicant's self-image and this person's perception of the employer's expectation, the lower the motivation to adapt to those expectations. This reasoning corresponds with findings that, whereas personality scores are partially inflated in actual applicant settings (Birkeland et al., 2006), these effects are less pronounced and less generalizable across traits than in directed faking experiments in the lab (Viswesvaran & Ones, 1999).

To summarize, a number of authors have specified elements of skills and motivation that may be needed to perform successfully on selection instruments and may also translate into later performance on the job, albeit through partially different mechanisms. While the potential presence of positive effects of self-presentation on validity might explain why the numerous attempts to improve validity by eliminating self-presentation did not write quite a story of success (for brief reviews, see Burns & Christiansen, 2011; Kuncel & Borneman, 2007), little attempt has been made so far to utilize this potential. In the following section, we propose that one reason for this failure may lie in the fact that the predominant approach to personnel selection is not tailored to cover self-presentation, and we offer a different approach to supplement the traditional paradigm.

Personnel Selection as Psychometrics and as a Social Game

As noted earlier, professional standards of selection (SIOP, 2018) propose to first establish job requirements in terms of job-relevant constructs (i.e., knowledge, skills, abilities, and other characteristics [KSAOs] such as personality traits), and then to design and apply instruments measuring those KSAOs. This traditional approach, hereafter referred to as “psychometric,” has certainly led to the development of a broad range of diverse selection devices that allow for meaningful predictions of job performance (e.g., Schmidt & Hunter, 1998).

Although the common focus of all psychometrically derived KSAOs is on job requirements, there are fundamental differences between the types of instruments designed to measure different kinds of requirements. Whereas knowledge (K) and skills (S) are relatively specific elements of a particular job, abilities (A) and other characteristics (O; i.e., mainly

personality traits) are relatively stable dispositions conceptualized as latent causes of K and S that need to be inferred more indirectly. Once this is done, A and O constructs then can be adopted from general (i.e., nonjob-specific) taxonomies such as the five-factor model of personality or Fleishman and Quaintance's (1984) ability categories. Wernimont and Campbell (1968) referred to this distinction between direct (KS) and indirect (AO) approaches in selection as samples vs. signs of job behavior. “Sample” refers to procedures such as job sample tests or assessment center exercises that simulate elements of the position to be filled, whereas “signs” refer to more abstract constructs typically measured with ability or personality tests designed for broader purposes.

Another psychometrically relevant distinction is between KSA constructs on one hand and O constructs on the other. Whereas instruments designed to measure KSAs are typically administered with “do your best” instructions, personality tests measuring Os regularly include instructions asking for accurate or honest responses (e.g., “There are no right or wrong answers.”). This latter difference aligns with the distinction of maximum and typical performance on the job (Sackett, Zedeck, & Fogli, 1988). Although more than 50 years ago, Wallace (1966) already proposed to cross the line between the concept of personality as how one “really is” and personality as what one is capable to be (i.e., between O and KSA), the practice of personality testing still almost exclusively rests on the former type of concept.

Hence, selection procedures developed through traditional psychometric procedures can be ordered along the continua of job-specific vs. (relatively) context-free and of maximum vs. typical levels of performance. The specific context of the selection setting is typically not an integral part of designing psychometric selection procedures. Even attempts to contextualize originally context-free procedures like personality tests (Schmit et al., 1995; Shaffer & Postlethwaite, 2012) are aimed at shifting signs somewhat toward the sample pole on the dimension of job specificity. Hence, contextualization in that sense refers to the job context, not the selection context. We hold here that the potentially valid elements of self-presentation discussed in the previous section unfold in the specific context of selection. Thus, taking specifics of the selection context into account may be relevant for realizing the validity potential of self-presentation.

With regard to those specifics, several theorists (e.g., Johnson & Hogan, 2006; Marcus, 2009; Roulin et al., 2016) described personnel selection as a social game whose actors include representatives of the prospective employer and (mostly several) applicants who also compete with each other for a job. In this competitive social game, all actors convey certain images of themselves their respective co-actors need to decipher, and vice versa. Marcus (2009) referred to these complementary tasks as selection and attraction, respectively. In general, both employers and applicants have to complete

either task. Although the competition among applicants toward the end of the process tends to shift their focus more on the task of attracting the employer, they still have the option to withdraw in case they perceive the employer as insufficiently attractive.

From a psychometric perspective, obtaining a job offer is the outcome of superior scores on the KSAOs being measured, whereas from a social perspective, job offers result from superior self-presentation skills and motivation. The two perspectives will converge to the extent that measured KSAOs match self-presentation skills and motivation. Yet, if two measures converge perfectly, there is room neither for validity detriments nor increments. We therefore argue that the likelihood of finding potential of incremental validity due to self-presentation is greatest where there currently is a mismatch between specifics of the selection context and underpinnings of the traditional psychometric approach. By contrast, a purely psychometric view on selection would consider any discrepancy between measured constructs and target KSAOs as bias.

Arguably, the perspective on selection as a social game of self-presentation aligns to differing degrees with instruments ordered along the psychometric dimensions of signs vs. sample and of typical vs. maximum performance, respectively. By sampling task elements of the job, performance on instruments designed to measure K and S in selection should translate relatively directly into performing the same K and S constructs on-the-job. Similarly, the maximum performance instructions of ability (A) tests lead us to conclude that no specific skills of self-presentation are needed to understand the employer's expectations with regard to those tests (although test anxiety may restrict performance on ability tests in the selection context, cf. Marcus, 2009).

However, with regard to the O type of requirements (i.e., personality), there is a clear conceptual mismatch between the psychometric and the social perspective on selection. Personality test items and, by implication, also many questions in interviews, are designed to measure typical cognitive, emotional, and behavioral styles that occur across different situations with differing social expectations and over longer periods of time (e.g., McCrae & Costa, 1997). By contrast, personnel selection represents a relatively exceptional situation likely to trigger a mindset more commensurate with maximum performance and adaptation to a specific set of social expectations (cf. Tett & Simonet's, 2011, notion of "faking as performance"). Such a mindset is at odds with intentions of personality test authors, as expressed in instructions to describe typical tendencies and to avoid "socially desirable" self-presentation. Yet, how applicants deal with social demands in a specific situation may reflect skills and motivation that are job relevant, as discussed earlier, but need not match the target constructs test authors had in mind when designing the questions.

Finally, there may be a difference between the psychometric and the social perspective that lies beyond the KSAO distinctions. Regardless of specific category, from a psychometric perspective, selection procedures derive their criterion-related validity from how well an applicant's score reflects her or his standing on the KSAO construct the test is designed to measure. Thus, any indicator the score is based on is seen as reflecting a particular construct. From a social perspective, being successful as an applicant depends on how well the employer's expectations are met overall, which translates into predictive validity to the extent these expectations are job relevant. Any degree of mismatch on the part of the applicant may result from a complex set of causes, including lack of self-presentational skills, or motivation, or an actually undesirable standing on any of the constructs the employer's expectations may be based on, or a mixture of all these factors. Because multiple combinations of logically independent causes may lead to similar degrees of (mis)match, the focal construct of the social perspective has a formative nature. Table 1 summarizes major differences between the psychometric and the social perspective on measuring O constructs in selection.

If, in fact, accidental positive effects of self-presentation on measures of O constructs compensate for impaired measurement of target constructs, as the retained validity of personality tests applied for selection suggests (see, e.g., Tett & Christiansen, 2007, for an overview), there may be room for incremental validity by intentionally scoring responses to questions aimed at O constructs for self-presentation. In the next section, we present a practical proposal aimed at that purpose. We chose personality tests as prototypical measures of O constructs, which are also at the core of most discussions on self-presentation in selection settings. Our empirical studies will cover both a compound trait (i.e., integrity, study 1) and general dimensions of personality (i.e., the Big Five; study 2) as two different types of personality assessment widely used in selection. On the criterion side, both studies will cover supervisor ratings of broadly defined job performance as probably the most generic type of criterion used in validation research. However, different criteria are added in study 2 to provide some initial evidence of generalizability to the broader criterion space.

Scoring Personality Tests for Self-Presentation: the Ideal Employee Coefficient

In terms of the distinction just introduced, design and scoring of personality tests is virtually always based on the psychometric approach. Responses to items are summed up or averaged to reflect the individual standing on a particular trait. If administered in the context of a social game, test takers only

Table 1 Psychometric vs. social perspective on measuring other (O) characteristics in personnel selection

	Psychometric Perspective	Social Perspective
Source of validity	Job requirement (as based on job analysis)	Match between social demands in selection and on-the-job
Target construct	Stable personality trait	Situation-specific skills and motivation
Meaning of score	Standing on job requirement (trait) construct	Match between responses and employer's expectations
Contextualization	None or job context	Selection context
Expected response behavior	Not affected by situation ("honest")	Affected by situation
Meaning of self-presentation	Bias	Target construct
Construct type	Reflective	Formative

have the option to present themselves using the given response format (often a rating scale). Thus, self-presentation cannot manifest itself anywhere but in responses to scales originally developed for psychometric purposes. The present approach is meant to extract the social meaning of these responses in order to supplement traditional psychometric scoring and thereby utilize the potential of the dual nature of personality test responses collected in high-stakes settings. Specifically, we are scoring responses to personality tests for the degree of match with employer's expectations. From a theoretical perspective, we try to disentangle the substantive, yet supposedly context-free, psychometric meaning from the context-driven, yet substantively undefined, social meaning inherent in one and the same set of responses. As we do not collect any new responses, one practically relevant implication is that the social information could be obtained very cost-effectively. Furthermore, we wanted our method to be easily applicable by practitioners who may not always have the resources to apply highly sophisticated psychometrics (e.g., Zickar, Gibby, & Robie, 2004; Ziegler, Maaß, Griffith, & Gammon, 2015).

We label the outcome of our scoring method *ideal employee coefficient* (IEC), which refers to Schmit and Ryan's (1993) early notion of an Ideal Employee Factor (note that we are *not* attempting to measure the factor extracted by Schmit and Ryan but simply wanted to acknowledge their pioneering work). Furthermore, one idea adopted from self-presentation theory (Marcus, 2009) is that laypersons who complete a personality test in an applied setting typically do not ruminate about the constructs being measured but look at test items like at any other single element of assessment that may affect the final outcome. From the test taker's perspective, there are thus "right and wrong" answers to personality items despite frequent test instructions to the contrary. A social approach to selection implies finding those "right" answers (in contrast to giving "true" answers) in much the same sense as performance tests imply finding correct or optimal solutions. This, in turn, requires a standard of good or optimal performance against which to compare the individual performance of the test taker.

For the purpose of developing such a standard, we first ask a group of subject matter experts (SMEs) to complete a personality test with the instruction to respond to each item in the way they think an ideal employee should do. Whereas SME judgments have a long tradition in designing sample-like predictors such as situational judgment tests (e.g., Chan & Schmitt, 2002), the sign-type of approach to personality has barely made use of this social component representing the employer's subjective view. Because ideal responses can be anywhere on the given rating scale, the IEC has an inherent means of handling nonlinear predictor–criterion relations. In order to reflect the information typically available to actual test takers on what is expected from them (cf. Marcus, 2009), SMEs should be familiar with job requirements but naïve with regard to the psychometrics underlying standard personality scores. Then, interrater reliability across SMEs is established using intraclass correlation coefficients ($ICC_{(2)}$). At this stage, it may be necessary to drop single SMEs if their responses turn out to reflect a highly idiosyncratic view on the ideal employee (analogous to eliminating items based on low item–total correlations in test construction). The IEC is supposed to reflect relatively agreed-upon standards of ideal responses. For similar reasons, we next eliminate test items for which no such agreement among experts could be reached (see "Method" sections for criteria used for that purpose in the present research). We then compute, for each single item, the median of the SMEs' ideal employee ratings. Median ratings across all items in the test are labeled *ideal employee profile* (IEP) and provide for the needed standard of best performance on the test. The IEP is then correlated against each test takers' responses to the same items in a work-related setting. This results in a profile correlation coefficient (technically equivalent to $ICC_{(1)}$) for each individual test taker, which represents this individual's final IEC score.

The IEC score thus is an omnibus index of similarity between expert ratings of ideal responses to items on a particular test and individual responses to those same items. As outlined earlier, this match is considered the (formative) outcome of a process of self-presentation involving multiple factors, including elements of self-presentation and the individual's standing on the traits

the test is designed to measure. It is probably unrealistic to expect the IEC to be independent of test content. If a test is composed of items that lack any job relevance, it would be impossible to define a reliable standard of ideal responses needed to compute the IEC and to expect validity for predicting job behavior. However, whereas traditional scores are intended to measure traits, the IEC is aimed at capturing the social meaning of responses. From a psychometric perspective, context effects on traditional scores are considered bias, whereas from a social perspective, trait effects on the IEC can be considered bias in an analogous sense. Traditional scores and the IEC are thus flip sides of the same coin, with reversed foci either on trait constructs or on the social meaning of the same set of responses. The two pieces of information combined should yield a richer account of the full meaning of those responses than either type of information alone.

Although there are theoretical propositions (e.g., Marcus, 2009) on how specific constructs (e.g., mental ability, job experience, etc.) affect the outcome of self-presentation, the IEC shall not be confused with measures of those constructs nor of any other reflective construct. Scoring the IEC is technically and logically independent of the content of the instrument on which it is based. It can be computed in exactly the same way for measures of narrow and homogeneous traits, as well as for compound traits or multidimensional inventories. Conceptually, the IEC is “social” as it necessarily changes if social expectations (i.e., expert ratings comprising the IEP) change, or if test takers’ willingness to adapt to those expectations changes. Because the IEC and traditional scores have different foci, it is not necessary that all items are retained for the IEC score.

Computationally, the IEP to which individual responses are compared is specific for the test at hand. Whether or not it is also job-specific is an empirical question that may in part depend on the content of test items. For some types of personality tests and constructs, generalizable validity has been established (e.g., integrity, Van Iddekinge, Roth, Raymark, & Odle-Dusseau, 2012; or conscientiousness, Barrick, Mount, & Judge, 2001), whereas for other types of tests, ideal responses may depend more on job content.

Study Objectives and Hypotheses

For the present research, we used previously collected data from three different samples to investigate the validity of the IEC. In the first two samples, the same set of personality and ability tests and the same criterion measures had been used in different settings resembling either low- or high-stakes conditions. These two samples were combined for study 1 to establish the IEC’s construct and criterion-related validity. For study 2, a large data set

collected in an actual high-stakes setting with a different personality test and a range of different criteria was used to examine to what extent the IEC’s criterion-related validity generalizes to such real-world applications. Based on our rationale outlined earlier that the IEC adds to a richer account of the meaning of item responses, we expect the following relations of the IEC to measures of job performance.

H1: IEC scores will be positively related to test takers’ job-related performance (criterion-related validity).

H2: The IEC yields positive relations to test takers’ job-related performance over and above validities of the same test takers’ regular personality test scores based on the same responses (incremental validity).

Unlike the traits aimed at in traditional scoring, the skills and motivational components supposedly driving the criterion-related validity of the IEC are believed to be triggered only in high-stakes settings. Sensitivity to setting is essential for establishing construct validity of the IEC, in order to rule out the possibility that it measures context-free characteristics (e.g., person-job fit) rather adaptation to the situation. Notably, we expect setting effects on IEC’s criterion-related validity directly opposite to setting effects on regular psychometric scores typically expected as a consequence of “faking”:

H3a–b: Criterion-related (a) and incremental (b) validities of the IEC are higher in high-stakes than in low-stakes settings (moderator effect of setting).

In addition to criterion-related validities, the saturation of the IEC with components of self-presentation should be reflected in setting effects. First, higher stakes create a motivation to adapt to external expectations, which is not present when stakes are low, a difference that should affect IEC scores (H4). Second, whereas research has shown that most personality tests are largely independent of cognitive ability if stakes are low (e.g., Ackerman & Heggestad, 1997), smarter applicants are held to possess better self-presentation skills (Marcus, 2009), which implies that IEC scores are related to cognitive ability in selection, but not in neutral, settings (H5).

H4: Mean IEC is higher in high-stakes than in low-stakes settings (motivation-related effect of setting).

H5: The IEC is more highly related to general mental ability (GMA) in high- than in low-stakes settings (skill-related effect of setting).

Study 1 is outlined to test all of the above hypotheses, whereas study 2 replicates findings on H1 and H2 in an actual high-stakes setting.

Study 1

In study 1, we wanted to establish criterion- and construct-related validity of the IEC, as specified in H1 to H5 above. For these purposes, we employed two data sets in which the exact same measures had been used in job incumbents either with regular honest instructions or with the instruction to act as applicants.

Method

Samples and Procedures We employed data from two samples on which general information is available in previously published papers. Sample 1 is composed of $N = 174$ job incumbents in two different companies (one retail chain and one manufacturing firm) in Germany who held heterogeneous jobs ranging from semiskilled, to skilled entry-level with and without customer contact, to managerial (see Marcus, Schuler, Quell, & Hümpfner, 2002). All participants took an integrity test with the instruction to act as if it were employed during selection for the very job they currently hold. Compared with instructions typically used in laboratory faking experiments, this instruction was meant to enhance the realism of the setting. Specifically, participants were not referred to some imaginary job of unknown meaning to them but to a selection situation they had experienced in the past. This guaranteed that all participants had a realistic frame-of-reference for their imagination. As the original study was concerned with counterproductive work behavior, only limited information on demographics had been collected to guarantee anonymity (cf. Marcus et al., 2002). Age was measured in three categories: younger than 25 years (16.7%), 25 to 40 (51.7%), and older than 40 years (31.6%). Sex was measured only in the retail subsample ($N = 98$, 53% male), because there had been too few women in the manufacturing sample to warrant anonymity.

Sample 2 is composed of $N = 272$ job apprentices trained within the German dual system, which consists of an academic part in specialized schools and on-the-job training during regular employment. Assessments were taken under honest conditions at school, whereas performance ratings were obtained at work from direct supervisors for a subsample of $N = 170$ apprentices. Participants were working in food production (16%) and food retail (84%) industries. Mean age was 18 years; 41.5% of the sample were women (cf. Marcus & Wagner, 2007).

Measures The same instruments were applied in both samples. The measure of personality was the 115-item German integrity test *Inventar beruflicher Einstellungen und Selbsteinschätzungen* (IBES, Marcus, 2006), which consists of both an overt and a personality-based part. The IBES is modeled after the content of widely used US integrity tests

and was found to closely resemble psychometric properties of those tests (cf. Marcus, Lee, & Ashton, 2007). GMA was measured with a German version of the *Wonderlic Personnel Test* (WPT, Wonderlic, 1996), a speeded 50-item measure of g widely used for personnel selection. Job performance was measured broadly with the mean across six items that overlapped between the two samples. Items were adopted from previous research and tapped into task (two items referring to quantity and quality of work), contextual (three items; e.g., “volunteers for extra work”), and overall job performance (one item) (cf. Marcus et al., 2002).

The IEC was constructed by providing the IBES with the instruction described earlier to eight experienced SMEs who held jobs in general management, marketing, or software engineering. SMEs were selected for their recruiting experience in a range of different jobs. Because it was not possible to obtain expert ratings from the test takers’ employers, the breadth of jobs covered appeared to correspond to the heterogeneity of the test taker samples. As integrity tests are designed to tap nonjob-specific constructs, a generic ideal profile also appeared adequate. None of the experts had any known relationship with the study participants. Analogous to the typical SD value expected for 5-point Likert-type scales, a cutoff of $SD > 1$ across SMEs was used to eliminate items on which insufficient consensus could be reached. This led to the exclusion of 15 items. One original SME had to be dropped due to very low agreement with the remaining experts (a post hoc interview implied that this SME did not follow instructions to portray an ideal applicant). The final IEP was then based on the median across seven SMEs and yielded an $ICC_{(2)}$ of .89, thereby providing some initial evidence of generalizability of social expectations across jobs for the present integrity items. As described earlier, each respondent’s IEC was then scored by correlating that person’s vector of responses to the IBES items with the IEP. Mean IEC was .30 in sample 1 ($SD = .18$; range = $-.34$ to $.68$) and .25 in sample 2 ($SD = .14$; range = $-.14$ to $.56$).

Results

Table 2 shows descriptives and bivariate correlations among variables for both samples. In line with H1, the IEC correlated positively with job performance in both samples. Criterion-related validity of the IEC was slightly higher in the simulated applicant ($r = .28, p < .01$) than in the honest ($r = .19, p < .05$) condition, but this difference in correlations was not statistically significant ($z = .88, p = .19$, one-tailed). Although in the expected direction, it thus does not support H3a. As expected in H4, mean IEC was significantly higher in the simulated applicant setting ($t(320) = 3.18, p < .01, d = .31$). Also expectedly (H5), the IEC correlated with GMA in the simulated applicant ($r = .31, p < .001$), but not in the honest ($r = .04, ns$) condition ($z = 2.58, p < .01$, one-tailed).

Table 2 Study 1 descriptives and intercorrelations

Variables	<i>M</i>	<i>SD</i>	Rel. ¹	(1)	(2)	(3)	(4)	(5)	<i>M</i>	<i>SD</i>	Rel. ¹
(1) Job performance	3.41	.66	.89	–	.19*	.15*	.22*	.15	3.33	.50	.81
(2) IEC	.30	.18	.89	.28**	–	.42**	.19*	.04	.25	.14	.89
(3) Integrity overt	213.70	23.37	.89	.13	.52**	–	.68**	.00	200.15	24.43	.90
(4) Integrity personality--based	188.37	18.17	.85	.21**	.34**	.69**	–	-.01	171.65	17.84	.86
(5) General mental ability	97.94	14.82	–	.25**	.31**	.08	.05	–	96.79	8.57	–

Sample 1 (applicant instruction) estimates are presented below; sample 2 (honest instruction) estimates above the diagonal

¹ Internal consistency reliabilities representing ICC₍₂₎ based on expert ratings for the IEC and Cronbach's alpha for remaining variables; *N* varies from 174 to 189 for sample 1, and from 170 to 272 for sample 2

IEC Ideal Employee Coefficient

p* < .05; *p* < .01

Table 3 displays results of hierarchical regression analyses, in which the two integrity subtests and GMA were entered first, and the IEC was entered at step 2. Despite its correlation with GMA, the IEC was incrementally valid beyond both GMA and the sign-based integrity scores in the sample instructed to act as applicants ($\beta = .26$, $p < .01$, $\Delta R^2 = .044$, $f^2 = .052$), which supports H2. According to Cohen's (1988) conventions, the effect size corresponds to a small- ($f^2 \geq .02$) to medium-sized ($f^2 \geq .15$) effect. By contrast, in the sample instructed to respond honestly, the IEC's incremental validity failed to reach conventional levels of statistical significance despite some small positive effect ($\beta = .15$, $p = .06$, $\Delta R^2 = .020$, $f^2 = .022$). As a more formal test of H3b, we computed the product term of the IEC \times sample (after mean centering both variables) and entered that product in a moderated regression analysis. Although in the expected direction, the moderator effect was not statistically significant ($\beta = -.04$; $SE = .03$, $t = -1.23$, $p = .109$, one-tailed). Hence, H3b received no formal support. Notably, incremental validities of regular integrity scores beyond GMA were similar in both samples and even slightly higher in the honest ($\beta_s = .24$, $-.00$; $p < .01$, $\Delta R^2 = .056$, $f^2 = .057$) than in the simulated applicant ($\beta_s = .22$, $-.03$; $p < .01$, $\Delta R^2 = .042$, $f^2 = .045$) condition.

Discussion

In sum, results from study 1 tended to confirm our hypotheses on the IEC. In contrast to regular scores, the IEC was shown to be sensitive to the different instruction sets in ways predicted in H4 and H5, and it demonstrated evidence of practically and statistically significant incremental criterion-related validity in the applicant simulation. Setting effects on criterion-related validities were nonsignificant, however, although the direction of observed differences corresponded to H3a and H3b.

Overall, these findings seem encouraging for establishing the construct and criterion-related validity of the IEC.

However, a number of limitations may compromise this preliminary conclusion. First, although the exact same measures were used in both samples, these samples were adopted from previous research and thus differ in features other than instructional sets employed. Job groups were roughly comparable, though, as in both samples the majority consisted of frontline employees in the retail industries, which corresponds to one traditional core field of application for integrity tests (e.g., Ones, Viswesvaran, & Schmidt, 1993). However, there may still have been specifics of jobs or industries affecting how integrity items were perceived, which we were not able to control in the present study due to the heterogeneous composition of both the SME and the test taker samples. Furthermore, participants in the honest condition were homogeneously young, whereas employees in the simulated applicant setting were much more heterogeneous in age. We therefore controlled for age in the latter group, which led to a slight but insubstantial decrease in the IEC's incremental validity ($\beta = .23$, $p < .05$, $\Delta R^2 = .032$, $f^2 = .038$). Finally, the present selection setting was only simulated. Although realism was enhanced by giving applicant instructions to employees who had actually been applicants before, the present results call for replication under real high-stakes conditions, which was our major objective in study 2.

Study 2

The major objectives of study 2 were testing the generalizability of the findings of study 1 on the criterion-related validity of the IEC to (1) an actual applicant setting, (2) with an occupationally more homogeneous sample, (3) to a partially predictive design, (4) with a different type of personality test

Table 3 Study 1 hierarchical regression of job performance on predictors

Sample	1: Applicant instruction	2: Honest instruction
Step 1: regular scores		
Integrity overt	-.03 (-.001; -.006/.005)	-.00 (-.000; -.004/.004)
Integrity personality-based	.22* (.008; .001/.015)	.24* (.007; .001/.012)
General mental ability	.26** (.022; .010/.035)	.17* (.009; .001/.018)
<i>R</i>	.336	.280
<i>R</i> ²	.113***	.080*
Step 2: IEC added		
Integrity overt	-.17 (-.005; -.011/.001)	-.08 (-.002; -.006/.003)
Integrity personality-based	.22* (.008; .001/.015)	.26* (.007; .001/.013)
General mental ability	.20** (.017; .004/.030)	.16* (.009; .000/.017)
IEC	.26** (.953; .299/1.607)	.15 (.532; -.031/1.096)
<i>R</i>	.396	.315
<i>R</i> ²	.157***	.099*
ΔR^2	.044**	.020
Effect size for IEC (<i>f</i> ²)	.052	.022
<i>N</i>	163	165

Entries corresponding to single variables are standardized regression coefficients (*β*s). Unstandardized coefficients *B* and the lower/upper bound of 95% confidence intervals appear in parentheses below corresponding *β*s. Listwise deletion of missing values used

IEC Ideal Employee Coefficient

p* < .05; *p* < .01; ****p* < .001

(measuring a set of personality dimensions rather than a compound trait), and (5) using a broader set of sources of criterion measures. In addition, in study 2, we explored to what extent the IEC predicts meaningful criteria beyond a traditional questionnaire measure of social desirability. Whereas the latter kind of measures tend to confound valid and invalid elements of self-presentation and of target traits (Connelly & Chang, 2015), the IEC is designed to capture criterion-valid elements of self-presentation. In line with established evidence (e.g., Ones et al., 1996), we therefore do not expect questionnaire measures of social desirability to show meaningful relations to job performance or to affect IEC–performance relations. Study 2 does allow for an explicit test of this supposition.

Method

Sample and Procedure Study 2 used a data set collected by the Israel Defense Forces (IDF) in a large sample of regular soldiers (*N* = 15,629) who applied for promotion to officer ranks. Personality testing was part of a larger set of psychological assessments used to determine selection for an intensive officer training course (OTC), which prepares for promotion to officer ranks. Criterion data were obtained for a subsample (*N* = 1904 to 2693) on three criteria: peer ratings of performance during training, overall training performance, and supervisor ratings collected 1 year after the training. Hence, we were able to test H1 and H2 with data based on three different sources (peers, supervisor, objective performance) and two

different time frames (concurrently or shortly after predictor measures and predictively after a 1-year interval). Predictor data were unknown to all raters from whom we obtained criterion measures.

Measures Personality was measured with a 255-item inventory designed for the IDF to measure the Big Five dimensions of personality, namely agreeableness, conscientiousness, extraversion, neuroticism, and openness to experience. Each dimension is measured with 48 items scored on 7-point scales of agreement. In addition, 15 items embedded in the scale are designed to tap social desirability.

Ratings represent means across five (peers), or six (supervisors), items tapping various performance dimensions (e.g., adherence to values, abilities as commander, physical fitness, organization, etc.). Supervisor ratings were obtained from participants’ immediate line supervisors, whereas peer ratings represent averages across a sizable number (often 20 or more) of fellow training participants who rated each other at the end of the training. The OTC score is a weighted average across trainers’ ratings of leader effectiveness and scores on leadership exercises, simulations, and other tests related to military command during training (e.g., navigation, firing range, military maneuvers using live fire, location defense, military history, and use of necessary weapons and military tools). In terms of the number of activities included, objective assessments are given much larger weight (about 87%) in the overall score than subjective ratings, yet it is not possible to

disentangle components of the overall score post hoc, as these were not stored in the database.

IEC scores were based on an independently collected median ideal employee profile across five IDF recruitment experts who did not participate in performance ratings. Because the rating scale of the personality test had changed from five to seven points between studies 1 and 2, an $SD < 2$ now was chosen as cutoff for keeping single items, in order to reflect conventions for the longer scale. This led to the deletion of two items. Interrater agreement was $ICC_{(2)} = .93$. Mean IEC was at $.46$ ($SD = .10$, range = $-.31$ to $.73$).

Results

As displayed in Tables 4 and 5, the IEC again showed some evidence of bivariate ($r = .08, .09, .15$; all $ps < .001$) and of incremental validity ($\beta = .18, \Delta R^2 = .020, f^2 = .021$; $\beta = .13, \Delta R^2 = .010, f^2 = .010$; $\beta = .25, \Delta R^2 = .038, f^2 = .040$; all $ps < .001$) for all three criteria of peer ratings, supervisor ratings, and OTC scores, respectively. These findings consistently support H1 and H2. Furthermore, selected candidates had higher mean IECs than the remainder of the sample ($t(4, 451) = 13.3, p < .001, d = .26$). Finally, the IEC predicted criteria not just beyond traditional trait measures but also beyond our questionnaire measure of social desirability.

Discussion

Results of study 2 largely corroborate findings on the IEC's criterion-related validity observed in study 1. Given the low criterion-related validities observed for regular personality scores in this study, one may have concluded that responses to personality items were of little utility in this context. However, this conclusion would have overlooked effects of self-presentation and must be reconsidered if IEC scores are taken into account. Although observed validities for the IEC were not high, they outperformed regular personality and social desirability score validities, they were all statistically significant, and two of the f^2 values exceeded at least the conventional (Cohen, 1988) cutoff of $.02$ for small effects. Notably, the largest effect was observed for OTC scores, which are predominantly based on objective performance. Although this may seem counterintuitive given the presumably social nature of the IEC, it needs to be stressed that participants were trained for leadership, which is a task of inherently social nature. Moreover, as in study 1 above, there were no feedback processes linking the various subjective sources involved. Hence, it seems implausible that stereotypes or implicit theories of performance transmitted through feedback (Staw, 1975; but see also DeNisi & Pritchard, 1978) explain the observed validities of IEC scores. Finally, successful candidates tended to have higher IECs, which supports interpretation of the IEC as an indicator of test performance (note that not all

participants without criterion data had actually failed selection, which may have attenuated the observed effect).

As the present inventory was multidimensional, it would have also been possible to compute construct-wise IECs (i.e., one per dimension). Although we performed respective analyses based on an anonymous reviewer's suggestion (which led to marginal increases in criterion-related validity), it may be worth mentioning why we refrain from reporting these results in detail. Construct-wise (in terms of reflective constructs) computation of IEC scores runs counter to the very conceptual idea of the IEC. As mentioned earlier, the IEC is not an alternative scoring method for the constructs personality tests are designed to measure. This job is done by traditional scores, though imperfectly. The IEC is an alternative scoring method for capturing the social meaning of responses *regardless* of underlying constructs.

An unexpected finding is that, in some instances, trait–criterion (especially conscientiousness–OTC) relations opposite to what is usually reported in the literature turned even stronger after entering the IEC. Apparently, the IEC at times accounts for valid portions of those traits—a finding that merits further attention but lies somewhat beyond the scope of the present paper.

General Discussion

Summary of Results and Theoretical Implications

In the present article, we proposed to look at test takers' self-presentation behavior in high-stakes settings as a hardly avoidable, though unwanted, outcome of social expectations in those settings. We further proposed to supplement a purely psychometric and, by implication, largely context-free view on assessment with a social view that takes context effects into account. Whereas the psychometric perspective leads to viewing self-presentation (or “faking”) as bias, from a social perspective, self-presentation may be viewed as revealing additional, and potentially useful, pieces of information about the test taker. We then introduced the IEC as a practical approach to extracting this information from regular personality test scores. In three samples covering a range of different contexts, instruments, jobs, organizations, and criteria, the IEC consistently displayed evidence in line with expectations on sensitivity to context and on relations to other constructs and performance criteria. This occurred despite the facts that (a) original psychometric scores' validity of the present personality scales varied considerably across samples and criteria, (b) correlations between IEC and the original scores were substantial, (c) simulating a high-stakes setting in study 1 led to a significant association between IEC and cognitive ability, and (d) study 2 covered a criterion which was only marginally affected by subjective ratings. Hence, the IEC appears to capture

Table 4 Study 2 descriptives and intercorrelations

Variables	<i>M</i>	<i>SD</i>	<i>N</i>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Peer performance rating	52.78	28.68	2693	(.89)									
(2) Supervisor performance rating	5.66	.90	1904	.24**	(.86)								
(3) Overall training score	81.59	5.30	2641	.42**	.19**	–							
(4) IEC	.46	.10	15,629	.08**	.09**	.15**	(.93)						
(5) Agreeableness	5.39	1.82	2675	.08**	.00	.04*	.13**	(.86)					
(6) Conscientiousness	5.21	1.68	2675	–.03	.02	.00	.56**	.53	(.93)				
(7) Extraversion	5.31	1.81	2675	.01	.01	.01	.36**	.31	.51**	(.92)			
(8) Neuroticism	4.59	1.75	2675	–.01	.02	–.02	–.44**	–.44**	–.66**	–.45**	(.94)		
(9) Openness to experience	4.69	1.89	2675	.00	–.04	.01	–.03	.15**	.05**	.32**	–.03	(.88)	
(10) Social desirability	5.20	1.81	2675	.05**	.02	.02	.38**	.67**	.68**	.34**	–.69**	.03	(.82)

Cronbach’s alpha coefficients are entered in the main diagonal, except for IEC, where entry represents ICC₍₂₎ based on expert ratings

IEC Ideal Employee Coefficient

p* < .05; *p* < .01

valid, job-related variance inherent in personality scores obtained in work settings. Whereas Tett and Christiansen’s (2007) dictum was that there is no evidence for the predictive validity of self-presentation in selection held for the questionnaire measure of social desirability in study 2, it cannot be upheld for the elements of self-presentation measured by the IEC.

Beyond questionnaire measures, the first common component across items of differing content is often held to capture socially desirable responding in high-stakes situations (e.g., Schmit & Ryan, 1993). In order to rule out that this component accounted for criterion-related validities observed for the IEC, we extracted the first component from personality items in the present studies and regressed our criteria on respective factor scores vis-à-vis the IEC. Despite substantial relations between those factor scores and the IEC in the range of *r* = .40 to .60, IEC criterion-related validities remained largely intact for all criteria in study 2 (with βs ranging from .10 to .20) and in the simulated applicant setting in study 1 (β = .29), whereas factor score validities were close to zero or even slightly negative. In line with our notion that IEC validity depends on high-stakes settings, IEC and factor scores both yielded similar and statistically insignificant validities around β = .13 in the honest condition in study 1 (detailed results of this set of analyses are available upon request from the first author). Taken together, these findings also provide some evidence that a general factor of personality found to predict job performance in research settings (Van der Linden, te Nijenhuis, & Bakker, 2010) does not account for the IEC’s criterion-related validity in high-stakes setting.

These findings correspond to theoretical views that self-presentation in personality assessment reflects more than mere faking or response distortion. According to Marcus’ (2009) theory, for example, self-presentation on personality tests in selection settings is a double-sided coin. On one hand, self-presentation affects the construct validity of personality tests

(in their traditional psychometric understanding as trait measures) negatively. On the other hand, self-presentation carries diagnostically useful information on applicants’ skills and motivation, which is commensurate with the social perspective on assessment presented here. In that sense, self-presentation may in fact correspond to McCrae and Costa’s (1983) classic notion of “more substance than style,” yet the substance may be of a different quality than the traits personality tests are originally designed to measure. From a practical point of view, it may therefore not be wise to try eliminating the specific substance of self-presentation in an attempt to distill the “pure” substance of traits, as is evident from the limited success of such attempts in terms of improving criterion-related validities. The present approach points to a different direction: making use of both kinds of substances. According to our results, this route appears more promising.

Theoretically, this leads to the question what kind of substance is measured by the IEC. It is tempting to approach this question from the perspective of a nomological net of relations with well-defined and stable constructs, as scholars of psychological assessment are typically trained to do. However, such attempts would overlook that the IEC is designed to capture the match between social expectations and the test taker’s adaptation to those expectations in a particular situation. It is not assumed that individual IEC scores generalize to different situations where this individual’s skills and motivation to adapt were different. Hence, rather than correlating IEC scores with a range of constructs to uncover a nomological net, in study 1, we tested the sensitivity of the IEC to changes in the social context. Although we emphasized criterion-related validity in the empirical parts of our research (i.e., tests of H1 and H2), we believe that the part of study 1 in which we tested setting effects (i.e., H3, H4, and H5) was appropriate for an initial test of the IEC’s construct validity and that our findings were encouraging in this respect.

Table 5 Study 2 hierarchical regression of job performance on predictors

Criterion	Peer rating	Supervisor rating	Training score (OTC)
Step 1: regular scores			
Agreeableness	.10*** (1.63; .82/2.44)	-.02 (-.01; -.04/.02)	.05 (.15; -.00/.30)
Conscientiousness	-.15*** (-2.58; -3.59/-1.58)	.03 (.02; -.02/.05)	-.03 (-.09; -.28/.10)
Extraversion	.04 (.67; -.09/1.43)	.03 (.01; -.02/.04)	-.01 (-.02; -.06/.13)
Neuroticism	.01 (.21; -.74/1.15)	.08* (.04; .01/.08)	-.02 (-.07; -.25/.11)
Openness to experience	-.02 (-.32; -.84/.30)	-.05 (-.02; -.05/.00)	.01 (.02; -.09/.14)
Social desirability	.08* (1.28; .25/2.31)	.07 (.03; -.01/.07)	-.01 (-.03; -.23/.16)
<i>R</i>	.131	.074	.046
<i>R</i> ²	.017***	.005	.002
Step 2: IEC added			
Agreeableness	.15*** (2.44; 1.61/3.27)	.01 (.01; -.02/.04)	.12*** (.36; .20/.51)
Conscientiousness	-.24*** (-4.11; -5.19/-3.04)	-.03 (-.02; -.06/.02)	-.15*** (-.48; -.68/-.28)
Extraversion	.02 (.32; -.44/1.07)	.01 (.00; -.03/.03)	-.04 (-.11; -.25/-.05)
Neuroticism	.03 (.49; -.45/1.43)	.10** (.05; .02/.09)	.00 (.00; -.17/.18)
Openness to experience	-.01 (-.16; -.78/.45)	-.04 (-.02; -.04/.00)	.02 (.06; -.05/.18)
Social desirability	.06 (.93; -.10/1.95)	.06 (.03; -.01/.07)	-.04 (-.12; -.31/-.07)
IEC	.18*** (58.12; 42.73/73.51)	.13*** (1.31; .73/1.90)	.25*** (14.85; 11.98/17.71)
<i>R</i>	.191	.125	.200
<i>R</i> ²	.037***	.016***	.040***
ΔR^2	.020***	.010***	.038***
Effect size for IEC (<i>f</i> ²)	.021	.010	.040
<i>N</i>	2675	1890	2623

Entries corresponding to single variables are standardized regression coefficients (β s). Unstandardized coefficients *B* and the lower/upper bound of 95% confidence intervals appear in parentheses below corresponding β s. Listwise deletion of missing values used

IEC Ideal Employee Coefficient

* $p < .05$; ** $p < .01$; *** $p < .001$

Implications for the Practice of Personnel Selection

As originally intended, the present results also show that our approach to measuring personality for selection may best supplement, rather than replace, traditional approaches to personality testing. One practically relevant feature of the IEC is that it provides for a cost-effective and simple supplement to traditional personality scores. The IEC is cost-effective as it does not require additional assessments apart from a small-scale expert study to compute IEPs. It is simple as all necessary computations can be performed with any spreadsheet program and interpreted by practitioners with just a basic understanding of statistics. Psychometricians developed highly sophisticated methods for separating true score trait variance from self-presentation (e.g., Ziegler et al., 2015), for dealing with nonlinearities in personality item responses (e.g., Stark, Chernyshenko, Drasgow, & Williams, 2006), or for accounting for the congruence between profiles (e.g., Edwards, 1994). The IEC is not meant to replace these kinds of high-end statistics. Rather, it is meant to supplement the simple traditional mean or sum scores with an equally simple additional score based on statistics the average practitioner is able to handle.

We believe that simplicity is a value in its own right when scholarly solutions are supposed to transfer to practice.

Unlike previous attempts to quantify skills associated with self-presentation, such as ATIC (Kleinmann et al., 2011), the IEC is conceptually independent of the constructs being measured with regular personality scores. A theoretically unlimited number of items could be listed in a single row to construct the IEP in the first step, regardless of whether these items measure one or several constructs, the nature of the constructs, and whether the job relevance of the target constructs could be established. Yet, of course, item content overlaps with that used for regular scores. If validities of regular scores are low (study 2), effect sizes of the IEC tend to diminish as well. It should be stressed that the IEC is not a magic tool that would generate validity without any base in the content of the items it is computed from. The difference between the IEC and traditional scoring rather is that traditional scoring primarily taps into the nomothetic meaning of items as indicators of specific constructs, whereas the IEC is concerned with the idiographic way a specific test taker interprets and reacts to this content in the situation at hand.

Notably, the present approach may appear similar to profile matching techniques often used to derive optimal scores on personality tests for a particular position (Kulas, 2013). One thing the IEC does have in common with those approaches is that both do not assume linear predictor–criterion relations. This assumption underlies the practice of validating traditional summed scores based on Pearson correlations but is at odds with findings of nonlinear personality–performance relations (e.g., Le et al., 2010). In fact, if SMEs agree that a maximum score on each item is optimal, the IEC cannot be computed due to a lack of variance in the underlying IEP (note that there was no indication that this happened in the present data). In contrast to the IEC, however, all methods of profile matching described by Kulas (2013) are based on profiles across constructs rather than items. Moreover, whereas Kulas used empirically keyed optima based on the actual performance of job incumbents who took the test, computation of the IEC is entirely independent of the actual performance of test takers later on-the-job. We added nonlinear (quadratic) terms of regular scores to the regression analyses presented in the “Results” section to provide for an estimate of the degree to which incremental validities observed for the IEC could be attributed to nonlinearity alone. As measured by changes in effect sizes for IECs (Δf^2), these additions had minimal effects ranging from $-.004$ to $+.002$. Future research may still investigate the IEC vis-à-vis linear and nonlinear methods of traditional scoring.

Another approach recently shown to improve the criterion-related validity of personality tests is the empirical keying of single items (Cucina et al., 2018), similar to what has been a long tradition with biodata inventories. Yet, unlike empirical item keys, the IEC does not require criterion data for developing scores nor large-scale studies for reliable cross-validation. These resources are unlikely to be available for most, and especially for small, companies. By contrast, IEC scores could be based on ratings obtained from a small number of experts. On the other hand, reliance on subjective expert ratings implies the danger that stereotypes in selection simply translate into stereotypes later on the job. Fortunately for the IEC, results from study 2 on training criteria and on peers as a different group of stakeholders provide evidence counter to that suspicion.

Limitations and Future Research Directions

As discussed earlier, a general limitation of the present research is that the substantive meaning of IEC scores, by their very nature, is not as readily interpreted as the meaning of traditional scores. Theoretically, the IEC is meant to capture the outcome of a complex process involving heterogeneous skills and motivation that interact with a range of situational factors. Hence, the IEC is held to change with situations, which received initial support in study 1. However, in this

study, we were not able to randomize assignment to conditions, which naturally leaves room for alternative explanations and points to controlled experiments as one route for future research.

Furthermore, only a limited number of tests of construct validity were possible with the present data. We considered it our foremost responsibility to establish the criterion-related validity of our approach, as it would be pointless to explore the meaning of a selection device that lacks practical utility. A positive side to the fact that our data had already been collected is that it can hardly be argued that our studies were designed or tailored to support our hypotheses, which seems of value in its own right given contemporary discussions on the credibility of empirical research. Although the IEC is not designed as a reflective construct measure, it is by no means atheoretical, as theoretical underpinnings (e.g., Marcus, 2009) suggest a whole range of links to outside variables. Another avenue for future research thus is designing studies that would allow for separating effects of skills and motivation present in IEC scores and for further exploring how these scores interact with features of persons and situations. Establishing evidence of such interactions is also best possible with complex experimental designs. Yet, randomization is typically hard to realize in actual applicant settings.

Another issue that merits attention in future research is the degree to which the validity of IEC scores may be attributed to stereotypes shared by test takers and raters of performance. Although we cannot rule out this possibility with confidence, features of the present research make this concern at least appear less likely. On the predictor side, the IEC is, by means of construction, based on two independent sources of ratings (test takers and SMEs generating the IEP). On the criterion side, we obtained ratings from peers and supervisors and results of objective tests. Hence, five independent sources of data were combined for the present research, all of which yielded largely consistent results. We consider this not a particularly weak basis for the initial validation of a new method.

The focus of the empirical part of our research was on the practical issue of incremental validity. Although not all effect sizes we found may seem impressive, we believe that the evidence presented in this paper consistently supports the assumed positive potential of self-presentation in these settings, given the range of variations in data sources, design, instruments, and cultural and occupational contexts covered by the present research. We hope that these findings contribute to a shift in perspective on the phenomenon of self-presentation. Instead of defining applicants’ self-presentation as bias, or even as morally wrong, scholars of personnel selection may be better advised to look at this as a phenomenon inadvertently triggered by the situation that we need to understand and deal with in intelligent ways.

References

- Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin*, *121*, 219–245. <https://doi.org/10.1037/0033-2909.121.2.219>.
- Bakan, D. (1966). *The duality of human existence: An essay on psychology and religion*. Oxford, UK: Rand McNally.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, *9*, 9–30. <https://doi.org/10.1111/1468-2389.00160>.
- 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*, 317–335. <https://doi.org/10.1111/j.1468-2389.2006.00354.x>View.
- Burns, G. N., & Christiansen, N. D. (2011). Methods of measuring faking behavior. *Human Performance*, *24*, 358–372. <https://doi.org/10.1080/08959285.2011.597473>.
- Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance*, *15*, 233–254. https://doi.org/10.1207/S15327043HUP1503_01.
- Cohen, J. (1988). *Statistical power analysis for the social sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum.
- Connelly, B. S., & Chang, L. (2015). A meta-analytic multitrait multirater separation of substance and style in social desirability scales. *Journal of Personality*, *84*, 319–334. <https://doi.org/10.1111/jopy.12161>.
- Cucina, J. M., Vasilopoulos, N. L., Su, C., Busciglio, H. H., Cozma, I., DeCostanza, A. H., Martin, N. R., & Shaw, M. N. (2018). The effects of empirical keying of personality measures on faking and criterion-related validity. *Journal of Business and Psychology* (published online May-07-2018). <https://doi.org/10.1007/s10869-018-9544-y>.
- DeNisi, A. S., & Pritchard, R. D. (1978). Implicit theories of performance as artifacts in survey research: A replication and extension. *Organizational Behavior and Human Decision Processes*, *21*, 358–366. [https://doi.org/10.1016/0030-5073\(78\)90059-4](https://doi.org/10.1016/0030-5073(78)90059-4).
- Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes*, *58*, 51–110. <https://doi.org/10.1006/obhd.1994.1029>.
- Fleishman, E. A., & Quaintance, M. K. (1984). *Taxonomies of human performance*. New York: Academic.
- Goffman, E. (1959). *The presentation of self in everyday life*. New York: Doubleday.
- Ingold, P. V., Kleinmann, M., König, C. J., & Melchers, K. G. (2015). Shall we continue or stop disapproving of self-presentation? Evidence on impression management and faking in a selection context and their relation to job performance. *European Journal of Work and Organizational Psychology*, *24*, 420–432. <https://doi.org/10.1080/1359432X.2014.915215>.
- Johnson, J. A., & Hogan, R. (2006). A socioanalytic view of faking. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 209–231). Greenwich, CT: Information Age.
- Kleinmann, M., Ingold, P. V., Lievens, F., Jansen, A., Melchers, K. G., & König, C. J. (2011). A different look at why selection procedures work: The role of candidates' ability to identify criteria. *Organizational Psychology Review*, *1*, 128–146. <https://doi.org/10.1177/2041386610387000>.
- Kulas, J. T. (2013). Personality-based profile matching in personnel selection: Estimates of method prevalence and criterion-related validity. *Applied Psychology: An International Review*, *62*, 519–542. <https://doi.org/10.1111/j.1464-0597.2012.00491.x>.
- Kuncel, N. R., & Borneman, M. J. (2007). Toward a new method of detecting deliberately faked personality tests: The use of idiosyncratic item responses. *International Journal of Selection and Assessment*, *15*, 220–231. <https://doi.org/10.1111/j.1468-2389.2007.00383.x>.
- Le, H., Oh, I., Robbins, S. B., Ilies, R., Holland, E., & Westrick, P. (2010). Too much of a good thing: Curvilinear relationships between personality traits and job performance. *Journal of Applied Psychology*, *96*, 113–133. <https://doi.org/10.1037/a0021016>.
- Levashina, J., & Campion, M. A. (2006). A model of faking likelihood in the employment interview. *International Journal of Selection and Assessment*, *14*, 299–316. <https://doi.org/10.1111/j.1468-2389.2006.00353.x>.
- Marcus, B. (2006). *Inventar berufsbezogener Einstellungen und Selbsteinschätzungen [Job-related attitudes and self-evaluations inventory]*. Test manual. Göttingen, Germany: Hogrefe.
- 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*, 417–430. <https://doi.org/10.1111/j.1468-2389.2009.00483.x>.
- Marcus, B., & Wagner, U. (2007). Combining dispositions and evaluations of vocation and job to account for counterproductive work behavior in adolescent job apprentices. *Journal of Occupational Health Psychology*, *12*, 161–176. <https://doi.org/10.1037/1076-8998.12.2.161>.
- Marcus, B., Schuler, H., Quell, P., & Hümpfner, G. (2002). Measuring counterproductivity: Development and initial validation of a German self-report questionnaire. *International Journal of Selection and Assessment*, *10*, 18–35. <https://doi.org/10.1111/1468-2389.00191>.
- Marcus, B., Lee, K., & Ashton, M. C. (2007). Personality dimensions explaining relationships between integrity tests and counterproductive behavior: Big Five, or one in addition? *Personnel Psychology*, *60*, 1–34. <https://doi.org/10.1111/j.1744-6570.2007.00063.x>.
- McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting Psychology*, *51*, 882–888. <https://doi.org/10.1037/0022-006X.51.6.882>.
- McCrae, R. R., & Costa, P. T. (1997). Personality trait structure as a human universal. *American Psychologist*, *52*, 509–516. <https://doi.org/10.1037/0003-066X.52.5.509>.
- McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology*, *85*, 869–879. <https://doi.org/10.1037/0021-9010.85.5.812>.
- 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*, 683–729. <https://doi.org/10.1111/j.1744-6570.2007.00089.x>.
- Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology*, *78*, 679–703. <https://doi.org/10.1037/0021-9010.78.4.679>.
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, *81*, 660–679. <https://doi.org/10.1037/0021-9010.81.6.660>.
- Pelt, D. H. M., Van der Linden, D., & Born, M. P. (2018). How emotional intelligence might get you the job: The relationship between trait emotional intelligence and faking on personality tests. *Human Performance*, *31*, 33–54. <https://doi.org/10.1080/08959285.2017.1407320>.
- Roulin, N., Krings, F., & Binggeli, S. (2016). A dynamic model of applicant faking. *Organizational Psychology Review*, *6*, 145–170. <https://doi.org/10.1177/2041386615580875>.

- Sackett, P. R., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology, 73*, 482–486. <https://doi.org/10.1037/0021-9010.73.3.482>.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of personnel selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274. <https://doi.org/10.1037/0033-2909.124.2.262>.
- Schmit, M. J., & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structure in applicant and non-applicant populations. *Journal of Applied Psychology, 78*, 966–974. <https://doi.org/10.1037/0021-9010.78.6.966>.
- Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology, 80*, 607–620. <https://doi.org/10.1037/0021-9010.78.4.679>.
- Shaffer, J. E., & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology, 65*, 445–494. <https://doi.org/10.1111/j.1744-6570.2012.01250.x>.
- Society for Industrial and Organizational Psychology. (2018). *Principles for the validation and use of personnel selection procedures* (5th ed.). Bowling Green, OH: SIOP.
- Stark, S., Chernyshenko, O. S., Drasgow, F., & Williams, B. A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology, 91*, 25–39. <https://doi.org/10.1037/0021-9010.91.1.25>.
- Staw, B. M. (1975). Attribution of the “causes” of performance: A general alternative interpretation of cross-sectional research on organizations. *Organizational Behavior and Human Performance, 13*, 414–432. [https://doi.org/10.1016/0030-5073\(75\)90060-4](https://doi.org/10.1016/0030-5073(75)90060-4).
- Tett, R. P., & Christiansen, N. D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007). *Personnel Psychology, 60*, 967–993. <https://doi.org/10.1111/j.1744-6570.2007.00098.x>.
- Tett, R. P., & Simonet, D. V. (2011). Faking in personality assessment: A “multisaturation” perspective on faking as performance. *Human Performance, 24*, 302–321. <https://doi.org/10.1037/0021-9010.88.3.500>.
- Van der Linden, D., te Nijenhuis, J., & Bakker, A. B. (2010). The general factor of personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study. *Journal of Research in Personality, 44*, 315–327. <https://doi.org/10.1016/j.jrp.2010.03.003>.
- Van Iddekinge, C. H., Roth, P. L., Raymark, P. H., & Odle-Dusseau, H. N. (2012). The criterion-related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology, 97*, 499–530. <https://doi.org/10.1037/a0021196>.
- Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*, 197–210. <https://doi.org/10.1177/00131649921969802>.
- Wallace, J. (1966). An abilities conception of personality: Some implications for personality measurement. *American Psychologist, 21*, 132–138. <https://doi.org/10.1037/h0023298>.
- Wernimont, P., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology, 52*, 372–376. <https://doi.org/10.1037/h0026244>.
- Wonderlic, I. (1996). *Wonderlic Personal Test (WPT—German version, Form A and B)*. Libertyville, IL: Wonderlic Personnel Test, Inc..
- Zickar, M. J., Gibby, R. E., & Robie, C. (2004). Uncovering faking samples in applicant, incumbent, and experimental data sets: An application of mixed-model item response theory. *Organizational Research Methods, 7*, 168–190. <https://doi.org/10.1177/1094428104263674>.
- Ziegler, M., Maaß, U., Griffith, R., & Gammon, A. (2015). What is the nature of faking? Modeling distinct response patterns and quantitative differences in faking at the same time. *Organizational Research Methods, 18*, 679–703. <https://doi.org/10.1177/1094428115574518>.

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