



## Face Validity in the Context of Personality Assessment: An Experimental Approach

Adelina Neagu

University of Bucharest

### ARTICLE INFO

#### Article history:

Received 24-February-2022

Accepted 28-April-2022

Available online 01-May-2022

This article should be cited as: Neagu, A. (2021). Face validity in the context of personality assessment: an experimental approach. *Studia Doctoralia. Psychology and Educational Science*, 13(1), 69-81  
<https://doi.org/10.47040/sdpsych.v13i1.144>

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Corresponding author at: University of Bucharest, Department of Psychology, 90 Panduri Av, Bucharest, RO. Tel.: +40 (0) 31-425.34.45.  
E-mail address: adelina.neagu74@gmail.com

### ABSTRACT

In the online environment, there are free psychological questionnaires available for anyone to complete, but test-taker perceptions have not been studied for these tools. Starting from the distinction between cognitive and visceral credibility and from the model of applicants' reactions in the selection process, this paper has the objective of evaluating whether manipulating the feedback format can predict perceptions related to face validity, accuracy and usefulness. The secondary objective was to test the predictor role of face validity in relation to perceived accuracy, respectively to perceived usefulness. To achieve these objectives, the participants have been randomly assigned in two groups, depending on the feedback format of the IPIP questionnaire. The results indicate that feedback format does not predict any of the three variables, but that face validity significantly predicts the perceptions related to accuracy and those related to usefulness. These results suggest the fact that respondents do not perceive differently the face validity, the accuracy and the usefulness of a test depending on the visual aspect of feedback. Face validity proves, however, to be a relevant predictor even outside of personnel selection.

**Keywords:** *face validity; visceral credibility; feedback; perceived accuracy; perceived usefulness*

### 1. INTRODUCTION

The fact that the internet has affected science and the way the general population relates to it is common sense.

The so-called health online communities have appeared, where people have discussions about health, disease, lifestyle, etc. This type of communities generates an impressive amount of information, which can be analyzed and reduced to well-defined themes and approaches

(Colineau & Paris, 2010), which are consistent among different topics of discussion (Park, Conway, & Chen, 2018). A specific type of discourse is the one oriented around auto-respective heterodiagnosis and around using a diagnosis as an identity aspect (Giles & Newbold, 2011), a fact which can affect the classic relationship between clients and health experts. Therefore, there are observable and relevant

shared characteristics between different online communities of this kind and common patterns of use, rules and specific languages have been formed, making it easy to develop online groups on any topic of discussion.

The ready availability of scientific content for the general population is achieved not only at the level of information, but also regarding assessment. Evaluating one's personality online has been popularized by the creation of the so-called personality questionnaires purposed for entertainment, produced by companies with the goal to distribute them in order to generate the consumption of information or products (Berberick & McAllister, 2016). For the consumers of this type of content, the outcome is represented by the opportunity to share on the internet a readily written information, which is descriptive of (a) their interests, through the nature of the questionnaire (e.g. if they are interested in Disney, maybe they will share a quiz that shows which Disney character they resemble the most); (b) their identity (adapted to the audience, through the fact that respondents choose to only distribute certain results); (c) their point of view, through that fact that they can add comments related to the content they are sharing (Quinn, 2017).

These quizzes or questionnaires are not presented online as being associated with psychology as a science, but their spread has normalized the evaluation and online sharing of certain aspects related to one's own identity. Instead, scientific personality questionnaires are being published online in an interactive format (i.e. with feedback being provided after completion) with the expressed purpose of delivering educational materials and sources regarding personality psychology and of offering a free option to test one's personality (e.g. <http://ipip.ori.org/>). The problem is that there are also non-scientific questionnaires online, which are extremely popular and which claim they are aligned to the scientific rigors (NERIS, n.d.). They can lead to the creation of a false image at the collective level of what personality psychology means, they can contribute to the formation of unrealistic expectations, which can affect the intent to request psychological help (Vogel, Wester, Wei, & Boysen, 2005), or they can even be wrongly used among online communities, with a decisional purpose (Giles & Newbold, 2011). Nevertheless, we could not identify empirical studies which approach the attitudes of internet users towards questionnaires published online and associated (by claim or by nature) with scientific personality assessment. Therefore, the objective of this paper is to evaluate whether the way in which feedback is delivered within these questionnaires has an effect on the way in which they are perceived by test-takers, from the perspective of information credibility and attitudes towards psychological tests.

## **Theoretical framework**

### **Feedback relevance.**

Providing the clients with feedback after completing psychological tests is not always done by psychotherapists (Curry & Hanson, 2010), although psychologists tend to deem this step of evaluation as useful, as it is facilitating communication with the clients and helps involving them in the decisional process at the same time (Mutén, 1991; Smith, Wiggins, & Gorske, 2007). Feedback has a beneficial effect not only in the therapeutic process (Finn & Tonsager, 1997), but also in education (Hattie & Timperley, 2007) and it also has an effect on candidates' perceptions towards the selection process in personnel selection (Gilliland et al., 2001), but in all these scenarios, the effects differ depending on the manner in which the feedback was delivered.

### **Visceral credibility.**

A test-taker's effort to analyze the feedback from online questionnaires entails informational processing in an environment full of unfiltered information and the content of the feedback has to be submitted to a decision related to credibility. Online credibility implies the consumer's evaluation of the degree to which the source, the message, and the web design can be trustworthy and presents an acceptable level of expertise in a given field (Choi & Stvilia, 2015). Evaluating the abilities to make decisions regarding the credibility of online information has revealed that many participants do not rely on a message or on the source, but have a heuristic approach, performing poorly in the tasks unfolded in the studies (McGrew, Breakstone, Ortega, Smith, & Wineburg, 2018). In this regard, the distinction between visceral and cognitive credibility was made, the former being achieved on a subjective, superficial level (Reich, 2011), predominantly based on the visual aspect of the website (Robins & Holmes, 2008). When they have to indicate the perception they have in relation to the credibility of a website, the participants consider more aesthetically pleasing websites to be more credible (Robins & Holmes, 2008; Robins, Holmes, & Stansbury, 2009), while in the comments related to the reasons for perception formation, the biggest incidence is had by the arguments related to the visual aspect of the website (Fogg et al., 2003). Not all internet users rely on visceral credibility equally, so focusing on the visual aspect of a webpage prevails in the case of participants who are not experts in the field the information they are evaluating belongs to (Stanford, Tauber, Fogg, & Marable, 2002, cited in Rieh & Danielson, 2007). Literature which has approached visceral credibility does not present practical guidelines on how a website has to be designed in order to obtain an effect on credibility. Particularities such as the symmetry of webpage elements (Tuch, Bargas-Avila, & Opwis, 2010), using colors and complexity (Reinecke et al., 2013) have been identified in the literature, but neither of them has been systematically studied and proven efficient.

### **Using human-like avatars.**

One aesthetical particularity of web pages or web environments is the use of images representing human figures. This type of characters is fairly used in the online environment. They have a social influence on the internet users, being used either as social models (in learning or behavioral intervention platforms), or as representations of the person using them (predominantly in online games), having potential effects on motivation and affectivity (Baylor, 2011). When they are used on e-commerce platforms, they influence the perception related to the purchase process, but also the perceptions related to the product and to the informational value of the message communicated by the avatar, as opposed to the contexts in which the product is promoted without using an avatar or through a non-human avatar (Holzwarth, Janiszewski, & Neumann, 2006; Jin & Bolebruch, 2009). As far as credibility is concerned, human avatars are perceived as more credible than non-human ones in a simulated context of online interaction (Nowak & Rauh, 2005; Nowak, Hamilton, & Hammond, 2009) and in the context of promoting products on virtual reality platforms (Mull, Wyss, Moon, & Lee, 2015). When avatars are being used by people who write reviews on e-commerce websites, the perceived credibility of the avatar influences the perceived credibility of the message source (McGloin, Nowak, & Watt, 2014).

### **Tangential information.**

Another way to superficially evaluate the credibility of information is to compare it to what the consumer is actually expecting to learn (McGrew et al., 2018). Non-experts have a reasonable capacity to correctly identify which day to day aspects of life are specific to personality traits (Hall, Schlegel, Castro, & Back, 2019), such that the provision of tangential information (i.e. variables correlated with personality traits) has the potential to increase the perceived credibility at a visceral level. Although, for valid instruments, this type of information can be identified in the scientific literature and is also supplied to the respondents of licensed psychological assessment tools (Costa & McCrae, 2014), they are not accessible to all the internet users who want to evaluate their personality.

### **Respondents' reactions to psychological tests.**

The main evidence related to attitudes towards psychological testing can be found in the domain of personnel selection. The model proposed and meta-analytically tested by Hausknecht, Day and Thomas (2004) indicates the fact that perceptions related to the selection process influence general candidate perceptions (e.g. attitudes towards testing, motivation to test), which, in turn, have behavioral effects relevant for the organization (e.g., intention to recommend the organization). Although they mention the fact that, at a practical level, a clear distinction

between variables which influence perceptions is hard to be done, Ryan and Huth (2008) divide these variables into three categories, namely content, method, and format.

### **Reactions to online psychological questionnaires**

When internet users respond to online personality questionnaires or quizzes, their task is to process information that is available online (i.e. they read, analyze and share the quiz and the feedback), but it is still a form of psychological assessment. Therefore, we will further describe and use in our study three variables which measure reactions to psychological questionnaires, while also taking into account the evidence from the study on online information credibility. In this field, the credibility variable has not been used, but it is conceptually close to the perceptions related to relevance (i.e. face validity), accuracy, and usefulness (Rieh & Danielson, 2007).

### **Face validity.**

The perceived face validity of a questionnaire is the test taker's appreciation of the relevance of the tool, related not to the latent construct that is being measured, but to the global purpose of the evaluation (Nevo, 1985). The theoretical papers related to this concept state the fact that a psychological evaluation process should not be a negative experience for the respondents and that face validity is another step towards the improvement of these processes in order to avoid negative effects motivationally, legally, or at the level of public image (Secolsky, 1987; Holden, 2010). The general nature of the definition of face validity allows its adaptation to any evaluation context. Thus, the empirical evidence related to face validity that we will refer to is extracted from the personnel selection literature, a field in which face validity is defined and measured as a factor of job association (i.e. if the respondent can identify a link between the evaluation and the job they are applying for; Smither, Reilly, Millsap, Pearlman, & Stoffey, 1993), together with the perceived predictive validity.

Face validity is defined as being influenced by the content of the assessment tool (Ryan & Huth, 2008) and the attempts to manipulate face validity have focused on modifying test items so that the test seems more relevant in the context of selection. Modifying certain items in order to have a content which makes a reference to the actual work activity has not had any effect on face validity in the case of a cognitive test battery (Smither et al., 1993), nor in the case of a personality test (Holtz, Ployhart, & Dominguez, 2005). However, a third paper has used the same strategy in order to manipulate reactions to cognitive tests and the results have shown significant differences between the group who completed the classic version of the test and the group who completed the test with modified items ( $p < .01$ ,  $d = .36$ ), the effect being stronger for women than for men ( $d = .42$ , respectively  $d = .22$ ; Grand, Ryan, Schmitt, & Hmurovic,

2010). Another strategy that was used in order to manipulate face validity consists in providing information to the test-taker before the beginning of the assessment, case in which providing information related to the measured construct (e.g. the definition of the construct) does not have a significant impact on face validity (Smither et al., 1993), but announcing the scientific validity of the test has a significant effect on face validity ( $p < .05$ ,  $\omega^2 = 0.2$ ; Holtz et al., 2005), with the mention that these papers have used a scale for job association and not for face validity particularly. Nevertheless, we have to remember that face validity is beneficial within testing, but it can not, under any circumstances, be chosen to the detriment of the psychometric quality of the test and the manipulation of some information before the actual testing seems to affect its precision (Holtz et al., 2005).

A direction which has not been approached is to test whether superficial characteristics of the test, such as the visual aspect, have an effect on face validity. Additionally, during personnel selection, it happens that the feedback received by candidates is provided only in the format accepted/rejected, so that the impact of feedback (as a source of information related to the test) on face validity has not been studied previously. Starting from this fact, the first two hypotheses have been formulated:

H1a: *The format of the feedback predicts face validity, in the case of a valid personality test (IPIP).*

H1b: *The difference in face validity between a test which is not valid, but is popular online (16 Personalities) and a valid test (IPIP) varies depending on the feedback format of the valid test.*

Perceived accuracy and usefulness. In the study of online information credibility, the credibility judgements are understood in the sense of accepting the information found online as being correct and also using it with a given purpose (e.g. Robins & Holmes, 2008; Robins, Holmes, & Stansbury, 2009). Acceptance of the results has been studied as the evaluation of the degree to which the respondents to questionnaires believe the feedback to be accurate, at content level (Furnham & Varian, 1988). The perceived usefulness has been rarely studied in the field of personnel selection, a possible explanation being the fact that the assessors are the ones responsible for establishing the usefulness of the selection tools, while the respondents are being asked to rate their attitudes towards these decisions, for example, regarding their perceived fairness of the selection process (Chan, Scmitt, Jennings, Clause, & Delbridge, 1998).

In a paper which had the objective to alleviate negative test-taker perceptions towards ipsative tests (Harland, 2003), the results obtained indicate the fact that receiving feedback significantly improves the perceptions related to

the usefulness of the test. In the context of online personality auto-evaluation, the two following hypotheses have been formulated, with the objective of evaluating whether the feedback format, manipulated so it alters credibility in a visceral way (superficially), is a significant predictor for the perceptions related to accuracy and usefulness:

H2: *Feedback format predicts the perceived accuracy of a valid questionnaire (IPIP).*

H3: *Feedback format predicts the perceived usefulness of a valid questionnaire (IPIP).*

A secondary objective of this paper is to test the relationship between the dependent variables which measure the respondents' perceptions. In the study of online credibility, the concepts of relevance, perceived accuracy and perceived usefulness are not clearly differentiated enough and, thus, are often used interchangeably (Rieh & Danielson, 2007). In turn, in regard to the personnel selection, these concepts are treated separately. Hausknecht et al. (2004) have placed face validity in the category of perceptions related to the characteristics of selection procedures, a category with a predictive role towards candidates' perceptions towards testing and, according to the results of the meta-analysis conducted by the authors, face validity is the strongest predictor for attitudes related to tests ( $p = .54$ ). Although useful for organizing the relationships between different respondents' reactions, this last information is limited in view of the fact that it groups attitudes related to tests in a single variable, thus not differentiating the influence of face validity on specific perceptions. Therefore, the following two hypotheses have been formulated, in order to test whether two distinct candidate perceptions, namely accuracy and usefulness, are predicted by face validity:

H4: *Face validity predicts the perceived accuracy of a valid personality test (IPIP).*

H5: *Face validity predicts the perceived usefulness of a valid personality test (IPIP).*

Overall, the hypothesized relationships can be tested using a single mediation model. With respect to the recommendation made by Ryan and Huth (2008), to test how exactly some perceptions are influenced by face validity manipulation, we will propose two hypotheses, with the mention that testing them will depend on the degree to which the other hypotheses are supported by data or not:

H6: *The relationship between the feedback format and perceived accuracy is mediated by face validity, in the case of a valid personality test (IPIP).*

H7: *The relationship between the feedback format and perceived usefulness is mediated by the face validity, in the case of a valid personality test (IPIP).*

## 2. METHODOLOGY

### Experimental manipulation

In order to achieve the first objective of this paper, of evaluating the effect of feedback format manipulation on the respondent's perceptions of online personality tests, two questionnaires were used, one scientifically valid (IPIP) and one non-valid (16 Personalities), whose characteristics are detailed in a following section. The non-valid questionnaire (i.e. 16 Personalities) presents all the aspects of feedback format previously mentioned, which can induce an increased degree of visceral credibility (e.g., using colors, symmetry, images, etc.), so that it was completed by participants in order to build a frame of reference, related to what can be found online concerning personality assessment. Subsequently, the valid questionnaire was completed, in two different formats, the participants being randomly assigned into two groups depending on the format of the valid questionnaire. Because we could not identify any websites available in Romanian, able to offer this valid questionnaire in an interactive format (i.e. to offer feedback upon completion) for free, for both experimental groups we used the platform described further.

To be able to offer feedback to the participants, the Psytoolkit platform (Stoet, 2010; Stoet, 2017; [www.psytoolkit.org](http://www.psytoolkit.org)) was used. In a way similar to other online personality scales (e.g. <https://openpsychometrics.org/tests/IPIP-BFFM/>), the participants were provided with explanations within the feedback as to what each personality trait from the Five Factors Model represents, with descriptions for low and high

scores (McCrae & John, 1992). Moreover, to make the results more easily understandable, a list of terms which describe the low and high scores was included in the feedback, for each trait individually (e.g. for high emotional stability: calm, comfortable, resilient; for low emotional stability: worried, unstable, vulnerable; McCrae & Costa, 2003). Because of some limitations related to the platform, the participants did not receive the percentiles corresponding to their scores, but were divided into three categories: low scores (under  $z = -1$ ), average (between  $z = -1$  and  $z = 1$ ), respectively high scores (higher than  $z = 1$ ), through relating to the means and the standard deviations obtained by Ilescu et al. (2015) for the five traits. For the control group, all these aspects were written in plain text, without adding colors or images. For the experimental groups, the literature suggestions were respected concerning the visceral credibility (i.e., using colors, symmetry and images). The images that represent human-like avatars were selected from a free pack of stickers

offered by a social network ([www.bit.ly/FacebookStickerStore](http://www.bit.ly/FacebookStickerStore)). The tangential information was selected from meta-analytical studies, which can be found in Table 1. The main issue concerning the experimental manipulation is that, in order to not significantly diminish the size of the final sample, no pilot study was done to evaluate the perception of the respondents with the aim of finding potential improvements in feedback format design.

Table 1. *Studies used for extracting tangential information to be used in the experimental group*

| Source  | Variable related to a personality trait |
|---|---|
| Anglim, Horwood, Smillie, Marrero, & Wood, 2020         | Well-being                              |
| Barańczuk, 2019a  | Coping                                  |
| Barańczuk, 2019b  | Social support                          |
| Bogg & Roberts, 2004                                    | Healthy behaviors                       |
| Buecker, Maes, Denissen, & Luhmann, 2020                | Loneliness                              |
| Fayard, Roberts, Robins, & Watson, 2012                 | Guilt                                   |
| Harari, Reaves, Beane, Laginess, & Viswesvaran, 2018    | Expatriate adjustment                   |
| Kern & Friedman, 2008                                   | Longevity                               |
| Malouff, Thorsteinsson, Schutte, Bhullar, & Rooke, 2010 | Couple satisfaction                     |
| Prinzie, Stams, Deković, Reijntjes, & Belsky, 2009      | Parenting                               |
| Wilmot, Wanberg, Kammeyer-Mueller, & Ones, 2019         | Workplace advantages                    |

## Participants

The sample consisted of 101 participants (35% males, 65% females; Mage = 26.40, SDage = 9.91), 49 in the control group, respectively 52 in the experimental group. The convenience sampling was done through a social network. Before collecting any data, the informed consent for completing the study was obtained, including consent for personal data processing. The participants were randomly assigned to the two experimental groups with the support of an online platform ([www.random.org](http://www.random.org)). All the participants from the sample met all three criteria of inclusion, namely:

(a) to not have been enrolled or to not currently be enrolled in higher education programs related to the psychological field, (b) to not have speciality knowledge on the theories on personality and (c) to be at least 18 years old at the time of study completion. The first two criteria of inclusion are justified by the specificity of the face validity to be measured by people outside of the specialty field specific to the assessment tool (Nevo, 1985).

## Procedure

Through the use of social networks, an invitation to voluntarily participate in the study was sent out. Every interested person contacted the author and the people who met the inclusion criteria were randomly assigned, one by one, in one of the two groups using the website [www.random.org](http://www.random.org). The existence of the two experimental groups was not brought to the attention of the participants until the end of the survey completion so as to not influence the results. Therefore, the purpose of the study, of evaluating perceptions related to two different personality questionnaires, was communicated to them, without including any details related to the experimental manipulation.

Depending on the group they were assigned to, the participants received a link with a Google Forms survey, including the sections we will describe further. In the first section, the informed consent related to the participation and processing of personal data was obtained, as well as the demographic data mentioned previously (age, gender). Section 2 contained a link to the 16 Personalities questionnaire, which the participants had to complete, to read the feedback and then to go back to the initial Google Forms page. Section 3 included the questions which evaluated the face validity, perceived accuracy and perceived usefulness for the 16 Personalities test.

Section 4 contained a link to the IPIP scale, being the only section which differed for each individual group: the experimental group received the link for the IPIP questionnaire which contained (1) tangential information and (2) images and colors used in the feedback, while the control group received a link to the version of the IPIP questionnaire in which the feedback consisted only in reporting the values for the personality traits scores (low,

average of high) and a short description of each score. Section 5 included the items which evaluated the face validity, perceived accuracy and perceived usefulness for the IPIP test, both groups receiving the same questions. The last question from the Google Forms survey had the role of identifying the extent to which the participants already knew about the personality questionnaires they had completed, a variable which was not used in testing the hypotheses, but which can be informative concerning the possibility of results generalization. After the form was completed and the results were registered, the participants were offered detailed information on the purpose of the research and the existence of the two experimental groups (i.e. debriefing).

The average time of completion was estimated at 25 minutes, varying depending on the time allocated by each participant for the analysis of the feedback received upon completion of both personality questionnaires. The data to be analyzed was only collected through the Google Forms survey, the results from the personality questionnaires having not been analyzed, since the per se evaluation of the personality traits/types was not relevant for the objectives of this study.

## Instruments

*Respondents' perceptions.* Although, generally, these perceptions are measured with a scale of two items minimum (Harland, 2003), measurements based on a single item have also been used (e.g. Steiner & Gilliland, 1996; Anderson & Witvliet, 2008). In this paper, the three perceptions related to testing were measured with one item each, created for the purpose of this study.

*Face validity.* The face validity was measured through the question: For the purpose of helping me understand my own personality, this questionnaire is:, with the following response options on the Likert scale: 1 = extremely unfit for its purpose, 2 = slightly unfit for its purpose, 3 = somewhat fit for its purpose, 4 = fit for its purpose and 5 = extremely fit for its purpose (Nevo, 1985). The same question was used for each of the two personality questionnaires.

*Perceived accuracy.* Being evaluated after receiving feedback, the perceived accuracy of the personality questionnaires is related to reality and not to expectations. Therefore, the question was: I think these results reflect my true personality, with a response measure on the Likert scale, where 1 = totally disagree and 5 = totally agree. The question was repeated for both personality questionnaires.

*Perceived usefulness.* As it is suggested in the pre-test information and instructions for the 16 Personalities questionnaire (i.e. the first personality scale completed by the participants during the study), the purpose of this assessment was gaining knowledge about oneself and thus the question was formulated as follows: I believe the results of this questionnaire have offered me useful information for my day to day life, the answers being measure on the Likert

scale, where 1 = totally disagree and 5 = totally agree. The question has been repeated for each of the two personality scales. The items used for measuring the perceived accuracy and the perceived usefulness are adapted from those proposed by Harland (2003).

**Personality.** 16 Personalities Questionnaire (16 Personalities; NERIS, n.d.). This personality questionnaire claims itself to be a blend between the Myers-Briggs Questionnaire (Myers & McCaulley, 1985), adapted to the theory of the Five Factors. The questionnaire contains 60 items, with a Likert scale measurement with 7 answer options where 1 = totally agree and 7 = totally disagree. An item example is: You find it difficult to introduce yourself to other people. Presently, the only data related to the validity and reliability of this questionnaire is reported on the main website, without describing the methodology that was used to calculate the psychometric properties. Moreover, there are no existing publications in peer-reviewed journals regarding the psychometric qualities of this personality scale, and thus this questionnaire can not be considered scientifically valid.

**The IPIP 50 items scale** (<http://iip.ori.org/>). This questionnaire is based on Five Factors Theory, which postulates the fact that personality is based on five dimensions, stable in time and between varied communities, namely extraversion, agreeableness, openness,

conscientiousness and emotional stability (McCrae & Costa, 2003). We used the 50 items IPIP scale which is inspired by the indicators proposed by Goldberg (1992) and translated to Romanian by Iliescu et al. (2015). Since there is a relevant body of data to support the validity and reliability of the Five Factors model (Goldberg, 1992; Costa & McCrae, 2003), as well as the relationship between these personality traits and other relevant variables (see Table 1), and the validity of the IPIP items (Goldberg et al., 2006), the IPIP 50 items scale was used in the study as a valid tool for personality assessment.

### Experimental design

The study conducted was an inter-subject experiment with two groups, the inter-subject variable being a categorical one, namely the feedback format. Face validity is a dependent variable in relation to the feedback format, and it represents an independent variable in relation to perceived accuracy and perceived usefulness. If these relations prove to be significant, we will further integrate all of these relations into a single mediation model, with face validity being the mediator variable.

## 3. RESULTS

### Descriptive statistics

Table 2 shows the means, standard deviations and the correlations between the study's variables. For the entire sample, 36.6% of the participants were previously aware of the 16 Personalities questionnaire, 1% of the participants were aware of the IPIP questionnaire, 3% did know both

questionnaires and 59.4% were not aware of either of them. Post hoc power analysis results were calculated using the G\*Power v3.1 software (Faul, Erdfelder, Lang, & Buchner, 2007). Those results are presented below for each relationship between study variables separately.

Table 2. Means, standard deviations and correlation coefficients for all study variables

|                                    | <i>M</i> | <i>SD</i> | 1     | 2     | 3     | 4     | 5     | 6 |
|------------------------------------|----------|-----------|-------|-------|-------|-------|-------|---|
| 1.Face validity (16 Personalities) | 4.36     | .67       | -     |       |       |       |       |   |
| 2.Accuracy (16 Personalities)      | 4.20     | .80       | .54** | -     |       |       |       |   |
| 3.Usefulness (16 Personalities)    | 4.12     | 1.17      | .59** | .41** | -     |       |       |   |
| 4.Face validity (IPIP)             | 4.25     | .82       | .44** | .12   | .34** | -     |       |   |
| 5.Accuracy (IPIP)                  | 4.14     | .86       | .28** | .25*  | .26** | .58** | -     |   |
| 6.Usefulness (IPIP)                | 4.06     | 1.01      | .44** | .17   | .58** | .59** | .46** | - |

\*\* .p < .01, \*p < .05

The results of a linear regression analysis (Table 3) yielded an insignificant regression equation ( $F[1, 99] = .045$ ,  $p = .833$ ),  $R^2 = .00$ . The results indicate that feedback format did not significantly predict face validity of the IPIP questionnaire ( $\beta = -.02$ ), and thus Hypothesis 1a is not supported by the data. The results of the power analysis revealed a statistical power of 5% for  $B = -.035$  and a sample size of 101 participants. In order to test whether the difference in face validity between the two questionnaires (i.e. Face validity [16

Personalities] - Face validity [IPIP]) is significantly different between the two experimental conditions, a  $t$  test for independent samples was used. According to the results, there is no significant difference between the control group ( $M = .04$ ,  $SD = .67$ ) and the experimental group ( $M = .17$ ,  $SD = .90$ );  $t(99) = -.83$ ,  $d = .16$ ,  $p = .480$ . Thus Hypothesis 1b is not supported by the data. Power analysis revealed a statistical power of 12% for an effect size of  $d = .16$  and sample sizes of 49 and 52 participants respectively.

Table 3. Results of the regression analysis with face validity (IPIP) as the dependent variable

| Predictor | <i>R</i> | <i>R</i> <sup>2</sup> | <i>B</i> | <i>SE</i> | 95%LLCI | 95%ULCI |
|-----------|----------|-----------------------|----------|-----------|---------|---------|
| Group     | .02      | .00                   | -.04     | .82       | -.36    | .29     |

Note: the categorical predictor was set as a dummy variable, with the value 0 for the control group and 1 for the experimental group.

In order to test the effect of feedback on perceived accuracy and usefulness, two linear regression analyses were performed. An insignificant regression equation ( $F[1, 99] = 1.458$ ,  $p = .230$ ),  $R^2 = .01$  was obtained for the relationship between feedback format and accuracy (Table 4). Feedback format does not significantly predict perceived accuracy ( $\beta = -.12$ ), thus indicating that Hypothesis 2 is not supported by

the data. Power analysis showed a statistical power of 22%, for  $B = -.206$  and a sample size of 101 participants. Another insignificant equation ( $F[1, 99] = .594$ ,  $p = .443$ ),  $R^2 = .00$  was obtained and it showed that feedback format did not significantly predict ( $\beta = -.08$ ) perceived usefulness (Table 5) and thus Hypothesis 3 is not supported by the data. Power analysis showed a statistical power of 8%, for  $B = .115$  and a sample size of 101 participants..

Table 4. Results of the regression analysis with perceived accuracy (IPIP) as the dependent variable

| Predictor | <i>R</i> | <i>R</i> <sup>2</sup> | <i>B</i> | <i>SE</i> | 95%LLCI | 95%ULCI |
|-----------|----------|-----------------------|----------|-----------|---------|---------|
| Group     | .12      | .01                   | -.21     | .86       | -.55    | .13     |

Note: the categorical predictor was set as a dummy variable, with the value 0 for the control group and 1 for the experimental group

Table 5. Results of the regression analysis with perceived usefulness (IPIP) as the dependent variable

| Predictor | <i>R</i> | <i>R</i> <sup>2</sup> | <i>B</i> | <i>SE</i> | 95%LLCI | 95%ULCI |
|-----------|----------|-----------------------|----------|-----------|---------|---------|
| Group     | .08      | .00                   | .12      | 1.01      | -.24    | .55     |

Note: the categorical predictor was set as a dummy variable, with the value 0 for the control group and 1 for the experimental group.

The effect of face validity on perceived accuracy and usefulness was tested using two separate linear regression analyses. A significant regression equation ( $F[1, 99] = 49.217$ ,  $p < .001$ ),  $R^2 = .33$  was revealed. Face validity significantly predicts ( $\beta = .57$ ) perceived accuracy (Table 6). Power analysis showed a statistical power of 99%, for  $B = .607$  and a sample size of 101 participants. Moreover,

another significant equation ( $F[1, 99] = .52.513$ ,  $p < .001$ ),  $R^2 = .34$  was obtained and it showed that face validity significantly predicts ( $\beta = .58$ ) perceived usefulness (Table 7). Power analysis showed a statistical power of 99%, for  $B = .726$  and a sample size of 101 participants. Thus, Hypothesis 4 and Hypothesis 5 are supported by the data..



Table 6. Results of the regression analysis with perceived accuracy (IPIP) as the dependent variable

| Predictor            | <i>R</i> | <i>R</i> <sup>2</sup> | <i>B</i> | <i>SE</i> | 95%LLCI | 95%ULCI |
|----------------------|----------|-----------------------|----------|-----------|---------|---------|
| Face validity (IPIP) | .58      | .33                   | .61**    | .71       | .44     | .78     |

\*\*p &lt; .001

Table 7. Results of the regression analysis with perceived usefulness (IPIP) as the dependent variable

| Predictor            | <i>R</i> | <i>R</i> <sup>2</sup> | <i>B</i> | <i>SE</i> | 95%LLCI | 95%ULCI |
|----------------------|----------|-----------------------|----------|-----------|---------|---------|
| Face validity (IPIP) | .59      | .34                   | .73**    | .82       | .53     | .93     |

\*\*p &lt; .001

Considering the fact that Hypotheses 1a, 1b, 2 and 3 are not supported by the data, we did not further test the mediation relationships stated in Hypotheses 6 and 7.

#### 4. DISCUSSION

The first objective of this paper has been to manipulate the feedback format with the aim of influencing (a) face validity (Hypotheses 1a and 1b), respectively (b) general perceptions related to the test, namely accuracy (Hypothesis 2) and usefulness (Hypothesis 3). The analyses that we conducted have identified effects that are statistically insignificant and low in magnitude. These results and, additionally, the high correlation between the perceptions of the 2 personality tests are counterintuitive. Although we were not able to identify studies which compare the attitudes towards the two personality questionnaires (or alternative forms for any of them) used in this study, the 16 Personalities questionnaire shows up frequently in online discussions, the personality types which it measures are used as descriptors to differentiate people, being known online as a friendly way to understand personality, while the scales that measure the Five Factors are promoted as valid methods and used predominantly in research (e.g. <https://openpsychometrics.org/>), but they are not as popular online as the 16 Personalities questionnaire.

These results can suggest the fact that feedback is not used by respondents as a source of information in the formation of perceptions related to a personality scale. Previous studies have defined the content of the items and the instructions received by the respondents as sources of informational content (Ryan & Huth, 2008), but the observed relationships are not consistent. On the other hand, the results can be understood in the sense that the respondents have taken into account the information from the feedback, but that they have analyzed this information cognitively and not superficially or viscerally (contrary to our expectations). People have a reasonable understating of what personality means (Hall et al., 2019), and the visualization of the feedback has not been restricted to a specific amount of time and the participants were able to spend as much time as

they wanted to analyze the results, unlike in other studies where specific instructions have been given to take a decision in the shortest amount of time possible from the exposure to the stimuli (e.g. Robins & Holmes, 2008). While considering these two possible explanations for our results, the reason for the discrepancy in popularity between the two personality questionnaires remains unclear.

Starting from the results of the meta-analysis realized by Hausknecht et al. (2004), we have analyzed the predictor role of face validity in relation to perceived accuracy (Hypothesis 4) and usefulness (Hypothesis 5), both perceptions regarding only the IPIP questionnaire. These 2 hypotheses are supported by data, thus contributing to the understanding of the importance that face validity has besides the domain of personnel selection.

#### Theoretical and practical implications

In the literature related to the manipulation of face validity, there have been multiple approaches, without a clear model provided in this regard (Ryan & Huth, 2008). We have not identified any paper which had manipulated the feedback format in order to achieve this goal, the obtained results in this paper indicating the fact that this strategy is not efficient and that other methods must be identified and studied. Furthermore, the high correlations between perceptions related to the two personality scales can suggest the fact that classifying types of psychological tests based on respondent reactions (Steiner & Gilliland, 1996; Anderson & Witvliet, 2008) is relevant, these reactions being relatively unitary between the two personality scales. The link between face validity and test-taker perceptions outside personnel selection highlights the usefulness of the initial conceptualization of face validity, as being a perception related to the test in any given context (Nevo, 1985).

At the practical level, this paper proposed the importance of studying the perceptions related to the questionnaires available online for the wide public to use, the improvement of these perceptions being essential in popularizing scientific tools, and raising awareness about the non-scientific ones.

### Limitations

This paper has a number of limitations and interpretation of the results must be done while taking them into account. Firstly, for objective reasons, the actual results of the IPIP questionnaire were not analyzed. Since we can not obtain data related to the psychometric characteristics of this questionnaire for any of the experimental groups, it is possible for the perceived accuracy to be influenced by a poor real accuracy.

Secondly, in the case of tangential information, the delivery of information was not customized depending on the results. All the participants from the experimental group received the same information related to the variables in relation to the personality traits. Dineen, Ling, Ash & DelVecchio (2007) have obtained results which indicate a significant effect had by customizing information on processing, in terms of viewing duration, respectively of memorizing the information. Therefore, the presentation of tangential information in a general form and unspecific to the results obtained in the personality questionnaire has to be taken into account in the interpretation of results. Also, the differential effect of manipulating the visual aspect against the manipulation of tangential information in the IPIP questionnaire has not been analyzed. Although both aspects are superficial and related to visceral perceptions rather than cognitive ones, they are differentiated in literature (Ryan & Huth, 2008) and their effects have to be approached separately as well.

Lastly, the results of the power analysis indicated a low power for the majority of the relationships analyzed in the study. An increase in the sample size might determine the identification of significant effects. For example, (Holtz et al.,

2005) have identified a low, but significant effect following pre-test information manipulation, in order to influence face validity ( $n = 345$ ,  $p < .05$ ,  $\omega = .02$ ).

### Future research directions

Within the theoretical framework of the respondents' reaction to testing, there is a well-defined body of research conducted in the personnel recruitment and selection context (Hausknecht et al., 2004; Ryan & Huth, 2008).

This paper had as an objective the analysis of these variables in the online auto-evaluation context. A direction for future studies could be the replication of results in the same context, but also outside of it, for example in the educational field (Nevo, & Sfez, 1985), or regarding respondents' perceptions towards participating in research (Loosveldt & Storms, 2008).

Keeping the context of online auto-evaluation, future studies can analyze the variables we proposed in relation to behavioral criteria, such as the intention to retake the test or to recommend the assessment, similar to the model proposed in the context of personnel selection (Hausknecht et al., 2004). Moreover, studying the effect of individual differences on respondents' perceptions is important, since there have been identified individual differences in forming test-related perceptions (e.g. job association; Oostrom, Born, Serlie, & Van Der Molen, 2010), as well as in forming perceptions related to web page design (symmetry of elements; Tuch et al., 2010).

In a general sense, the cause for the noticeable preference for the 16 Personalities questionnaire instead of the IPIP (or parallel versions of them), in the online environment, remains unknown. There is not yet any empirical data to evaluate this difference, nor the potential consequences of the popularity of an invalid personality questionnaire that is publicly associated with the image of the scientific evaluation of personality.

### REFERENCES

- Anderson, N., & Witvliet, C. (2008). Fairness reactions to personnel selection methods: An international comparison between the Netherlands, the United States, France, Spain, Portugal, and Singapore. *International Journal of Selection and Assessment*, 16(1), 1-13.
- Anglim, J., Horwood, S., Smillie, L. D., Marrero, R. J., & Wood, J. K. (2020). Predicting psychological and subjective well-being from personality: A meta-analysis. *Psychological Bulletin*, 146(4), 279-323.
- Barańczuk, U. (2019a). The five factor model of personality and sense of coherence: A meta-analysis. *Journal of Health Psychology*, 1-14.
- Barańczuk, U. (2019b). The Five Factor Model of personality and social support: A meta-analysis. *Journal of Research in Personality*, 81, 38-46.
- Baylor, A. L. (2011). The design of motivational agents and avatars. *Educational Technology Research and Development*, 59(2), 291-300.
- Berberick, S. N., & McAllister, M. P. (2016). Online Quizzes as Viral, Consumption-Based Identities. *International Journal of Communication*, 10, 3423-3441.
- Bogg, T., & Roberts, B. W. (2004). Conscientiousness and health-related behaviors: a meta-analysis of the leading behavioral contributors to mortality. *Psychological Bulletin*, 130(6), 887-919.

- Buecker, S., Maes, M., Denissen, J. J., & Luhmann, M. (2020). Loneliness and the Big Five Personality Traits: A Meta-analysis. *European Journal of Personality*, 34(1), 8-28.
- Chan, D., Schmitt, N., Jennings, D., Clause, C. S., & Delbridge, K. (1998). Applicant Perceptions of Test Fairness: Integrating Justice and Self-Serving Bias Perspectives. *International Journal of Selection and Assessment*, 6(4), 232-239.
- Choi, W., & Stvilia, B. (2015). Web credibility assessment: Conceptualization, operationalization, variability, and models. *Journal of the Association for Information Science and Technology*, 66(12), 2399-2414.
- Colineau, N., & Paris, C. (2010). Talking about your health to strangers: understanding the use of online social networks by patients. *New Review of Hypermedia and Multimedia*, 16(1-2), 141-160.
- Costa, P. T., Jr., & McCrae, R. R. (2014). The NEO Inventories. In R. P. Archer & S. R. Smith (Eds.), *Personality assessment* (p. 229-260). Routledge/Taylor & Francis Group.
- Curry, K. T., & Hanson, W. E. (2010). National Survey of Psychologists' Test Feedback Training, Supervision, and Practice: A Mixed Methods Study. *Journal of Personality Assessment*, 92(4), 327-336.
- Dineen, B. R., Ling, J., Ash, S. R., & DelVecchio, D. (2007). Aesthetic properties and message customization: Navigating the dark side of web recruitment. *Journal of Applied Psychology*, 92(2), 356-372.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.
- Fayard, J. V., Roberts, B. W., Robins, R. W., & Watson, D. (2012). Uncovering the affective core of conscientiousness: The role of self-conscious emotions. *Journal of Personality*, 80(1), 1-32.
- Finn, S. E., & Tonsager, M. E. (1997). Information-gathering and therapeutic models of assessment: Complementary paradigms. *Psychological Assessment*, 9(4), 374-385.
- Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., & Tauber, E. R. (2003). *How do users evaluate the credibility of Web sites?* Proceedings of the 2003 Conference on Designing for User Experiences - DUX '03, 1-15.
- Furnham, A., & Varian, C. (1988). Predicting and accepting personality test scores. *Personality and Individual Differences*, 9(4), 735-748.
- Giles, D. C., & Newbold, J. (2011). Self-and other-diagnosis in user-led mental health online communities. *Qualitative Health Research*, 21(3), 419-428.
- Gilliland, S. W., Groth, M., Baker, R. C., Dew, A. F., Polly, L. M., & Langdon, J. C. (2001). Improving Applicants' Reactions To Rejection Letters: An Application of Fairness Theory. *Personnel Psychology*, 54(3), 669-703.
- Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4(1), 26-42.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84-96.
- Grand, J. A., Ryan, A. M., Schmitt, N., & Hmurovic, J. (2010). How far does stereotype threat reach? The potential detriment of face validity in cognitive ability testing. *Human Performance*, 24(1), 1-28.
- Hall, J. A., Schlegel, K., Castro, V. L., & Back, M. (2019). What laypeople think the Big Five trait labels mean. *Journal of Research in Personality*, 78, 268-285.
- Harari, M. B., Reaves, A. C., Beane, D. A., Laginess, A. J., & Viswesvaran, C. (2018). Personality and expatriate adjustment: A meta-analysis. *Journal of Occupational and Organizational Psychology*, 91(3), 486-517.
- 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.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81-112.
- 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.
- Holden, R. B. (2010). Face validity. In I. B. Weiner & W. E. Craighead (Eds.), *The Corsini Encyclopedia of Psychology* (4th ed., pp. 637-638). Hoboken, NJ: Wiley.
- Holtz, B. C., Ployhart, R. E., & Dominguez, A. (2005). Testing the rules of justice: The effects of frame-of-reference and pre-test validity information on personality test responses and test perceptions. *International Journal of Selection and Assessment*, 13(1), 75-86.
- Holtz, B. C., Ployhart, R. E., & Dominguez, A. (2005). Testing the rules of justice: The effects of frame-of-reference and pre-test validity information on personality test responses and test perceptions. *International Journal of Selection and Assessment*, 13(1), 75-86.
- Holzwarth, M., Janiszewski, C., & Neumann, M. M. (2006). The influence of avatars on online consumer shopping behavior. *Journal of Marketing*, 70(4), 19-36.
- Iliescu, D., Popa, M., & Dimache, R. (2015). Adaptarea românească a Setului International de Itemi de Personalitate: IPIP-Ro [The Romanian adaptation of the International Personality Item Pool: IPIP-Ro]. *Psihologia Resurselor Umane*, 13(1), 83-112.
- International Personality Item Pool: A Scientific Collaboratory for the Development of Advanced Measures*

Jin, S. A. A., & Bolebruch, J. (2009). Avatar-based advertising in Second Life: The role of presence and attractiveness of virtual spokespersons. *Journal of Interactive Advertising*, 10(1), 51-60.

Kern, M. L., & Friedman, H. S. (2008). Do conscientious individuals live longer? A quantitative review. *Health Psychology*, 27(5), 505-512.

Malouff, J. M., Thorsteinsson, E. B., Schutte, N. S., Bhullar, N., & Rooke, S. E. (2010). The five-factor model of personality and relationship satisfaction of intimate partners: A meta-analysis. *Journal of Research in Personality*, 44(1), 124-127.

McCrae, R. R., & Costa, P. T. (2003). *Personality In Adulthood: A Five-Factor Theory Perspective* (Second Edition). Hoboken: Guilford Publications.

McCrae, R. R., & John, O. P. (1992). An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality*, 60(2), 175-215.

McGloin, R., Nowak, K. L., & Watt, J. (2014). Avatars and Expectations: Influencing Perceptions of Trustworthiness in an Online Consumer Setting. *PsychNology Journal*, 1(1-2), 7-28.

McGrew, S., Breakstone, J., Ortega, T., Smith, M., & Wineburg, S. (2018). Can Students Evaluate Online Sources? Learning from Assessments of Civic Online Reasoning. *Theory & Research in Social Education*, 46(2), 165-193.

Mull, I., Wyss, J., Moon, E., & Lee, S.-E. (2015). An exploratory study of using 3D avatars as online salespeople. *Journal of Fashion Marketing and Management: An International Journal*, 19(2), 154-168.

Mutén, E. (1991). Self-reports, spouse ratings, and psychophysiological assessment in a behavioral medicine program: An application of the five-factor model. *Journal of Personality Assessment*, 57(3), 449-464.

Myers, I., & McCaulley, M. (1985). *Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator*. Palo Alto, CA: Consulting Psychologists.

NERIS (n.d.). 16 Personalities. Retrieved from: <https://www.16personalities.com/>

Nevo, B. (1985). Face Validity Revisited. *Journal of Educational Measurement*, 22(4), 287-293.

Nevo, B., & Sfez, J. (1985). Examinees' Feedback Questionnaires. *Assessment and Evaluation in Higher Education*, 10(3), 236-248.

Nowak, K. L., & Rauh, C. (2005). The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. *Journal of Computer-Mediated Communication*, 11(1), 153-178.

Nowak, K. L., Hamilton, M. A., & Hammond, C. C. (2009). The effect of image features on judgments of homophily,

credibility, and intention to use as avatars in future interactions. *Media Psychology*, 12(1), 50-76.

Oostrom, J. K., Born, M. P., Serlie, A. W., & Van Der Molen, H. T. (2010). Effects of individual differences on the perceived job relatedness of a cognitive ability test and a multimedia situational judgment test. *International Journal of Selection and Assessment*, 18(4), 394-406.

Park, A., Conway, M., & Chen, A. T. (2018). Examining thematic similarity, difference, and membership in three online mental health communities from Reddit: a text mining and visualization approach. *Computers in Human Behavior*, 78, 98-112.

Prinzie, P., Stams, G. J. J., Deković, M., Reijntjes, A. H., & Belsky, J. (2009). The relations between parents' Big Five personality factors and parenting: A meta-analytic review. *Journal of Personality and Social Psychology*, 97(2), 351-362.

Quinn, S. M. (2017). *Feeding our Identities: BuzzFeed Quizzes as a Tool for Personal Identification in the Social Digital Age* (Unpublished master's dissertation). Roger Williams University, Bristol, Rhode Island.

Reich, Z. (2011). Source credibility and journalism: Between visceral and discretionary judgment. *Journalism Practice*, 5(1), 51-67.

Reinecke, K., Yeh, T., Miratrix, L., Mardiko, R., Zhao, Y., Liu, J., & Gajos, K. Z. (2013). Predicting users' first impressions of website aesthetics with a quantification of perceived visual complexity and colorfulness. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 2049-2058.

Rieh, S. Y., & Danielson, D. R. (2007). Credibility: A multidisciplinary framework. *Annual Review of Information Science and Technology*, 41(1), 307-364.

Robins, D., & Holmes, J. (2008). Aesthetics and credibility in web site design. *Information Processing & Management*, 44(1), 386-399.

Robins, D., Holmes, J., & Stansbury, M. (2009). Consumer health information on the Web: The relationship of visual design and perceptions of credibility. *Journal of the American Society for Information Science and Technology*, 61(1), 13-29.

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.

Secolsky, C. (1987). On the direct measurement of face validity: A comment on Nevo. *Journal of Educational Measurement*, 24(1), 82-83.

Smith, S. R., Wiggins, C. M., & Gorske, T. T. (2007). A survey of psychological assessment feedback practices. *Assessment*, 14(3), 310-319.

Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K., & Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel Psychology*, 46(1), 49-76.

- Steiner, D. D., & Gilliland, S. W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology, 81*(2), 134-141.
- Stoet, G. (2010). PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42*(4), 1096-1104.
- Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology, 44*(1), 24-31.
- Tuch, A. N., Vargas-Avila, J. A., & Opwis, K. (2010). Symmetry and aesthetics in website design: It's a man's business. *Computers in Human Behavior, 26*(6), 1831-1837.
- Vogel, D. L., Wester, S. R., Wei, M., & Boysen, G. A. (2005). The Role of Outcome Expectations and Attitudes on Decisions to Seek Professional Help. *Journal of Counseling Psychology, 52*(4), 459-470.
- Wilmot, M. P., Wanberg, C. R., Kammeyer-Mueller, J. D., & Ones, D. S. (2019). Extraversion advantages at work: A quantitative review and synthesis of the meta-analytic evidence. *Journal of Applied Psychology, 104*(12), 1447-1470.