

**On-line personality assessment: Are electronic versions equivalent
to the traditional one?**

Maja Parmač, Zvonimir Galić, Željko Jerneiđ
Department of Psychology
Faculty of Philosophy, Zagreb, Croatia

Suggested running head: On-line personality assessment

Address for correspondence:

Maja Parmač

Department of Psychology

Faculty of Philosophy

Ivana Lučića 3

10 000 Zagreb

mparmac@ffzg.hr.

ABSTRACT

The aim of this study was to test the equivalence of five-factor personality questionnaire IPIP100 between three different item presentation modes. The item presentation modes were: traditional paper-and-pencil, electronic identical to paper-and-pencil one and electronic mode with item-by-item presentation. Our research design included dependant samples of psychology students (N=80) and was done in proctored and nonanonymous settings. The results showed that three different questionnaire versions have equivalent factor structures, equal reliabilities, correlations between the same factors at the level expected from their reliabilities, equal means and variances and identical correlation patterns with external variables. Based on the gained data we can conclude that these are basically identical questionnaire forms, i.e. that electronic questionnaire can be used as a substitute for traditional paper-and-pencil version.

Key words: on-line testing, personality questionnaires, modes of item presentation, equivalence

INTRODUCTION

Nowadays, Internet has a great significance in everyday life. It provides us with the most recent information about events worldwide. Through Internet we do business, search new jobs, products and services; we trade, entertain ourselves, etc. The number of Internet users increases every day. By the end of 2005 more than billion people, i.e. 15.7% of the world population are likely to have Internet access (<http://www.internetworldstats.com>, 2006). Moreover, Internet is increasingly used in psychology (Barak, 1999), especially for psychological testing.

Psychological testing via Internet is not a novelty in psychology. Basically, it is computerized testing with some specific characteristics. According to Bartram (1994) there are several practical advantages of computerized testing over standardized paper-and-pencil tests, for example saving of disposable materials, presenting more standardized test instructions and punctual time keeping. Since the scores are automatically stored, precise results can be calculated very rapidly, with inevitable common human mistakes avoided. Also immediate, objective, expert-based narrative feedback can be given to the test-takers. New data can be automatically and easily added to test database thus adjusting norms and proper use for research purposes. In addition, psychological testing via Internet has other advantages (Barak & English, 2002). First, its main characteristic is lack of limitations related to time and place of testing, e. g. unless it directly interferes with testing purposes, the user can fill in the questionnaire (test) from the place and at the time of his convenience. Second, this type of testing refers to keeping tests updated. When using an on-line test (especially the one on the publisher's web pages), we can be sure that we use the most recent, improved test version. Further, updates of instructions, scoring technique and norms can be easily made to tests available on-line having no need for additional distribution or training and supervision of users (Barak & Buchanan, 2003). Thus, electronic tests are easily designed, updated and distributed (English, 1996).

The development of an electronic version of a well-known psychological instrument is the most common way of Internet testing. However, before using an electronic version, it is necessary to ensure the metric characteristics to be identical to those of traditional test form. According to Bartram (1994) for electronic version to be equivalent to traditional, both forms must have equal reliabilities, intercorrelations at the level expected from their reliabilities,

have comparable correlations with other variables as well as equal means and standard deviations. Besides, there is one more additional criterion that Bartram did not explicitly state, but it is very important, i.e. factor structure of the two forms of an instrument should be identical in order to be considered equivalent.

The equivalence of metric characteristics was the subject of a number of researches in the field of computer/Internet testing in psychology. General conclusions differ depending on the type of instrument involved. Having considered cognitive measures, based on meta-analysis, Mead and Drasgow (1993) concluded that computerized power tests were equivalent to paper-and-pencil tests, while the equivalence of speed tests, probably due to the different motor demands (Danthiir et al., 2005), was questionable. The situation is more complex with “non-cognitive” measures (primarily personality and biographical questionnaires). Even though there is a considerable correspondence in results of electronic and traditional questionnaires, some studies have shown that they are not always equivalent (Buchanan, 2002). Differences are found mainly in factor structures of questionnaires which measure more than one construct (Buchanan et al., 1999; Johnson, 2000; Woolhouse & Myers, 1999).

The studies that tested equivalence of electronic and traditional versions of “non-cognitive” measures gave special attention to socially desirable responding. The difference in arithmetic means on personality dimensions between traditional and electronic version (e.g. Joinson, 1999) found by the equivalence studies are interpreted with the different levels of social desirability in responding. Researchers give one of two different hypotheses depending on direction of difference. First, the participants are more honest and express less social desirability in responding when completing electronic version of the instrument. This hypothesis is based on the assumption that people perceive on-line tests more anonymously and privately and, as a result, are more open to reveal private information. The other hypothesis suggests that people are more concerned when they interact with computer because they are afraid that their answers are permanently stored and could be seen by others. As a result, they give more socially desirable responses to electronic versions (Lievens & Harris, 2003). Richman et al. (1999) conducted comprehensive meta-analysis, showing that transforming an instrument into electronic form has no clear impact on social desirability of responding. First, it seems that social desirability of responding depends on the interaction between presentation mode and the participant’s anonymity, i.e. combination of anonymity and on-line testing lead to the lowest level of social desirability in responding (Joinson, 1999).

Moreover, some specific characteristics of the presentation mode could be relevant for it. For example, restricted back-tracking and changing of responses in electronic forms was found both to lead to higher (Richman et al., 1999) and lower (Federico, 1991) levels of socially desirable responding in comparison to traditional versions. Obviously, further research is needed to resolve these equivocal findings.

The aim of our study was to test the equivalence of paper-and-pencil and electronic form of a five-factor personality questionnaire and to find out whether the modality of presentation of items has any influence on responding. Moreover, if we consider that the real testing for practical purposes (e.g. for employee selection or clinical assessment) is done in proctored and non-anonymous settings, we find it important to compare responses of respondents with those conditions held constant across modalities of test presentation. As explained earlier, the sheer equivalence of two psychological instruments should be tested as the first step in doing psychological research via Internet. In this way, we didn't test the specificity of psychological research or assessment via Internet but only the influence of presentation mode on responses on a personality questionnaire with all other conditions held equal. Moreover, electronic presentation of test material introduces some additional possibilities over classical "paper-and-pencil" testing that may influence the results on personality questionnaires. For example, "item-by-item" presentation and restricted changing of responses could lead to changed levels of social desirable responding. In our study we wanted to compare the responses between three modes of item presentation: paper-and-pencil, electronic completely equivalent to paper-and-pencil and electronic with item-by-item presentation with restricted backtracking and changing of results. In addition, our research design includes dependant samples of participants. Considering that most of the studies that found differences were conducted on independent samples, it cannot be excluded that those differences reflected real differences between participants. The dependent design, used in this study is more powerful in terms that such differences do not reflect differences between participants, but can be ascribed to the differences between situations.

METHOD

Participants

The study included 80 participants (70 female and 10 male), students of psychology at the Faculty of Philosophy in Zagreb, ranged between 18 and 47 years ($M = 21.9$; $SD = 3.35$).

Instruments

Two questionnaires were used in this study:

1. IPIP100 personality questionnaire consisted of 100 short statements describing specific behaviour (International Personality Item Pool, Goldberg et al., 2006). The questionnaire measures five broad personality dimensions (Extraversion, Agreeableness, Conscientiousness, Neuroticism and Intellect) with each dimension measured by twenty items. Participants had to estimate, using the rating scale from 1 to 5, how accurately each statement describes them, where 1 meant “very inaccurate” and 5 “very accurate”. The original version of the questionnaire was translated from English by Boris Mlačić and Goran Milas (1999) (<http://www.ipip.ori.org>).
2. Social desirability scale was composed of two measures of social desirability: Croatian adaptation of Paulhus BIDR (Balanced Inventory of Desirable Responding; Milas, 1998) and L-scale from Eysenck Personality Questionnaire. The questionnaire included two scales of Paulhus model of desirable responding (10 items for impression management and 12 items for self-deception) and L-scale from EPQ (21 items), and it contains 43 statements (questions) that participants have to agree or disagree with.

Each questionnaire had three different item presentation modes: traditional paper-and-pencil mode, electronic identical to paper-and-pencil mode and electronic with item-by-item presentation. In electronic version identical to paper-and-pencil version the questionnaires were designed to be completely identical (equal number of questions per page, possibility of reviewing all questions before answering, possibility of back-tracking, correcting, etc). In electronic version with item-by-item presentation each item was presented individually for both questionnaires, with no possibility of backtracking or correcting answers.

Procedure

Every respondent participated in three testing situations having the interval of three weeks between two situations. In every situation participants were filling-in both questionnaires (IPIP100 and questionnaire of social desirability) in the same item presentation mode. In one situation participants were filling the questionnaires in paper-and-pencil form, in the other situation they used electronic version identical to paper-and-pencil, and in the third situation they used electronic version with item-by-item presentation.

Testing was conducted in small groups of 20 participants in each group, with a trained test administrator who read the instruction before each questionnaire. The participants were told to fill-in the questionnaires as honestly as possible. Although they had to write down their name, gender, age and year of study after filling the questionnaire, they were told that their responses would be confidential. Participants were randomly assigned to the situations. The order of the presentation modes, as well as the order of filling out the questionnaires in the same mode was rotated.

Electronic versions of the questionnaire were available through web browser on computers in computer classroom at the Faculty of Philosophy. Questionnaires were prepared for by research assistants and the participants responded to by a mouse click on selected response option.

RESULTS

For the purpose of testing equivalence between three versions of IPIP100, five criteria for equivalence between electronic and traditional versions of test have been checked. In order to be considered equivalent versions must:

- have equivalent factor structures;
- have equal reliabilities;
- correlate with each other at the level expected from their reliabilities;
- have equal means and standard deviations;
- have comparable correlations with other variables.

1. Factor structure of personality questionnaires

As the first step of equivalence testing, we tested factor structure of questionnaires between the three situations. The data were analyzed by means of the principal component analysis with five fixed factors extracted and afterwards rotated with varimax procedure. Percentage of total variance explained was rather equal between the three modalities of presentation (49.6% for “paper-and-pencil”, 47.3% for electronic version equivalent to paper-and-pencil and 50.5% for electronic with item-by-item presentation).

Rotated factor solutions for three versions of IPIP100 indicated that factor structures corresponded to theory. Almost all items in three presentation modes had primary projections on factors they initially supposed to measure (in paper-and-pencil questionnaire 96, in electronic version identical to paper-and-pencil version 98 and in electronic version with item-by-item presentation 96 items). Moreover, high correlations of factor scores (.73-.91) between three situations indicated that latent questionnaire structure, regardless of the item presentation mode, consists of the same factors. So, despite the fact that the ratio between the numbers of participants (80) and the number of variables (100) was highly unfavorable for conducting factor analysis, the results support our conclusion about the equivalence of the factor structure between different modes of presentation. Gained stability of factor solutions, taken together with the study by Arrindel and van der Ende (1985) showing that neither the observations-to-variables ratio nor absolute number of observations aren't crucial for factor stability, support the use of factor analysis in our study.

2. Reliability of personality questionnaire scales

Reliabilities of IPIP scales are expressed by Cronbach's alpha coefficient of internal consistency for three questionnaire modes (paper-and-pencil, electronic identical to paper-and-pencil version, electronic version with item-by-item presentation). As noted in the Table 1, their values are very high (.75-.96.). In general, we can notice that coefficients of internal consistency are almost identical for each factor in three questionnaire versions (the largest difference is 0.04).

Table 1 about here

3. Correlation between different questionnaire forms

Correlation coefficients between different presentation modes of the questionnaire indicate high relationship between them. Values of correlation coefficients between versions for each trait are approximately the same and at the level of reliability for each scale.

Table 2 about here

4. Comparison of means and standard deviations

Table 3 shows means and standard deviations for each trait in three testing situations (paper-and-pencil, electronic identical to paper-and-pencil version, electronic version with item-by-item presentation).

To examine whether individual scores differ for each of the five dimensions between the item presentation modes, five one-way ANOVA-s for dependent samples were calculated. The only statistically significant difference was found for Neuroticism ($F(2,78) = 3.54$; $p = 0,03$), between traditional paper-and-pencil version and electronic version identical to paper-and-pencil form (post hoc-*LSD*, $p = 0,01$). There were no significant differences on other scales, i.e. individual scores for Extraversion, Agreeableness, Conscientiousness and Intellect do not differ regardless of presentation mode (paper-and-pencil, electronic identical to paper-and-pencil version, electronic version with item-by-item presentation).

Moreover, F-test for variance equivalence of dependent samples (Kanji, 1993) only showed the significant difference in variances for Extraversion between two electronic versions of questionnaire ($F = 0.27$; $p < 0.05$). No significant difference was found among other variances in different presentation modes.

Table 3 about here

5. Correlation between personality traits and social desirability measures

For the purpose of this study individual results on paper-and-pencil version of questionnaire for measuring social desirable responding were used as external variable for testing the last criterion of equivalence. Internal consistency reliability estimates for separate social desirability scales were between .63 - .83.

To determine the relationship between personality traits and social desirability measures correlations between each personality trait and social desirability scores were calculated. Due to the low reliability of some social desirability scales, correlation coefficients (Table 4) were corrected for attenuation.

Table 4 about here

As shown in Table 4, personality traits have identical correlation pattern with social desirability measures in three item presentation modes. Extraversion, Neuroticism (negative) and Intellect have significant correlations with self-deception, while Consciousness significantly correlates with scores on L-scale and impression management scale. Agreeableness shows no correlation with social desirability measures.

DISCUSSION

The aim of this study was to test the equivalence of five-factor personality questionnaire IPIP100 between different item presentation modes. The item presentation modes were: traditional paper-and-pencil, electronic identical to paper-and-pencil mode and electronic mode with item-by-item presentation.

Factor analysis confirmed theoretical five-factor structure with minimum deviance, and high correlations of factor scores indicated that three questionnaire version had the same object of measurement. Reliabilities of personality questionnaire scales were almost identical for each factor in three questionnaire versions. Furthermore, correlation coefficients between different questionnaire versions were equal for each trait and at the level of reliability for each scale. The only statistically significant difference was found on Neuroticism scale, between traditional paper-and-pencil version and identical electronic version (participants score somewhat higher in paper-and-pencil version), while there were no significant differences on other scales regardless of the presentation mode. Most of the studies found lower social desirability in responding on computerized or on-line questionnaires when compared to paper-and-pencil form (Davis, 1999; Joinson, 1999; Ployhart et al., 2003). It might be due to the fact that people perceive responding in this situation more anonymously and are more prone to reveal information about themselves. Considering that, we could expect differences between questionnaire versions reflected in lower scores on Conscientiousness and Agreeableness scale, and higher on Neuroticism scale in electronic questionnaire versions. On the other hand, there is a hypothesis that people are more concerned when interacting with computer because they are afraid that their answers are permanently stored and could be seen by others (Lievens & Harris, 2003). According to this hypothesis we would expect higher scores on Conscientiousness and Agreeableness scale, and lower on Neuroticism scale in electronic questionnaire versions. Since our participants, while filling in all questionnaires, wrote down their names (they were not anonymous), and were not afraid that their answers would be permanently stored in computer (participants were students who were familiar with computer bases), we did not expect difference between individual results considering item presentation modes. Our expectations, excluding the difference between paper-and-pencil and identical electronic version on Neuroticism scale, were confirmed. It is necessary to point out that no difference was found between individual results in paper-and-pencil version and electronic version with item-by-item presentation. Therefore, we concluded that difference found

between paper-and-pencil and identical electronic version was not the result of participants' fear that their answers would be permanently stored. No difference was found among variances in different presentation modes, with the exception of that between two electronic versions of the questionnaire on Extraversion scale. The obtained difference was significant at the 5% level and it was the result of high correlation ($r = .94$) between scales stated above (for scales with high correlations even small difference is sufficient to be significant). Concerning correlations of personality traits and external variables, identical correlation patterns were obtained in three item presentation modes.

There are some additional factors (besides social desirability in responding) that could have an impact on the equivalence of different questionnaire versions. Since the computer provides different item presentation modes, it enables various testing possibilities. As version identical to paper-and-pencil mode, even though conducted on computer, enables backtracking, correcting and changing answers, electronic version with item-by-item presentation does not provide that possibility. Considering that limitation, there are two hypotheses about the impact of backtracking and correcting in electronic questionnaires on the level of social desirable responding. The first is that enabled backtracking and correcting lead to the highest level of socially desirable responding (according to Richman et al., 1999), while, according to the second one, there is a possibility of additional lowering in distortion towards socially desirable responding (Federico, 1991). Our results did not confirm any of the two hypotheses because there were no differences in results between versions that enabled backtracking and correcting (paper-and-pencil version and electronic version identical to paper-and-pencil mode) and electronic version with item-by-item presentation, where this was not possible. Further more, the type of responding (mouse click vs. mark with pencil) had no impact on equivalence of different questionnaire versions. However, there was only slight possibility that self-estimations of typical behavior would be affected by changes in item presentation (Bartram & Bayliss, 1984), which was the only difference between situations in this study.

Other factors that could influence differences between results in different questionnaire versions were low level of PC skills and/or aversion towards computers. However, since filling in electronic version of questionnaire requires only basic computer skills, which our participants possessed, there were no such problems that could affect equivalence of different questionnaire versions.

In last few decades, many studies have been conducted for the purpose of comparing results from computer/on-line and traditional paper-and-pencil methods. The majority of findings show equivalence of these two questionnaire forms when talking about non-cognitive tests (Richman et al., 1999). Some of the often cited studies that showed high correspondence in results between on-line and traditional questionnaires for non-cognitive measures are studies of Buchanan and Smith (1999) and Pettit (1999). Buchanan and Smith (1999) compared traditional paper-and-pencil version of self-monitoring scale to its Internet version. Conducted factor analysis showed similar three-factor structure, similarly loaded by items, in both versions. Reliability of electronic questionnaire version was slightly higher than for the traditional version, and there were no differences in means and standard deviations between those two versions.

Pettit (1999) compared results on traditional paper-and-pencil and electronic version of computer-anxiety scale. She also found similar psychometric characteristics of different scale versions (comparable internal consistency coefficients and correlations with external variables) and concluded that on-line collected data could be compared to those collected in traditional way.

There are few limitations of this study, such as a rather small number of participants, students of psychology that limits the possibility of generalization. Furthermore, this study had dependent design, which implied certain limitations. Three weeks interval between filling out different questionnaire versions is short enough to prevent maturity effects, but some serial effects could emerge. Participants could remember their responses and try to repeat them in subsequent situations to show consistency. That kind of responding could favor the equivalence of different questionnaire versions.

In spite of stated limitations, based on the gained data we can conclude that these are basically identical questionnaire forms i.e. that electronic questionnaire version can be used as a substitute for the traditional paper-and-pencil version. Depending on the purpose of testing, participants' attitudes, space limitations and other characteristics of test situation we can choose which form of questionnaire to apply.

REFERENCES

Arrindel, J. S., & van der Ende, J. (1985). An empirical test of the utility of the observations-to-variables ratio in factor and component analysis. *Applied Psychological Measurement*, 9, 165-178.

Barak, A. (1999). Psychological applications on the Internet: A discipline on the threshold of a new millennium. *Applied and Preventive Psychology*, 8, 231-246.

Barak, A., & Buchanan, T. (2003). Internet-based psychological testing and assessment. In R. Kraus, G. Stricker, & J. Zack (Eds.). *Online Counseling: A Handbook for Mental Health Professionals* (pp. 217-239). San Diego, CA: Elsevier Academic Press.

Barak, A., & English, N. (2002). Prospects and limitations of psychological testing on the Internet. *Journal of Technology in Human Services*, 19, 65-89.

Bartram, D. (1994). Computer-based assessment. In C.L. Cooper & I.T. Robertson (Eds.), *International Review of Industrial and Organizational Psychology*, 9 (pp.31-69). Chichester, England: John Wiley.

Bartram, D., & Bayliss, R. (1984). Automated testing: Past, present and future. *Journal of Occupational Psychology*, 57, 221-237.

Buchanan, T. (2002). Online assessment: Desirable or dangerous? *Professional Psychology: Research and Practice*, 33, 148-154.

Buchanan, T., & Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology*, 90, 125-144.

Buchanan, T., Goldberg, L. R., & Johnson, J. A. (1999, November). WWW personality assessment: Evaluation of an online Five Factor Inventory. Paper presented at the meeting of the Society for Computers in Psychology, Los Angeles, CA.

Danthiir, V., Wilhelm, O., Schulze, R., & Roberts, R. D. (2005). Factor structure and validity of paper-and-pencil measures of mental speed: Evidence for a higher-order model? *Intelligence*, 33, 491-514.

Davis, R. N. (1999). Web-based administration of a personality questionnaire: Comparison with traditional methods. *Behavior Research Methods, Instruments and Computers*, 31, 572-577.

English, N. (1996, October). Web exams and surveys: Using the Web to gather and disseminate information. Paper presented at meeting of Association for Computing Machinery/Special Interest Group on University and College Computing Services User Services Conference (SIGUCCS), Chicago.

Federico, P. A. (1991). Measuring recognition performance using computer-based and paper-based methods. *Behavior Research Methods, Instruments, and Computers*, 23, 341-347.

Goldberg, L.R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloinger, C.R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84-96.

International Personality Item Pool (2006) URL: <http://www.ipip.ori.org>

Internet World Stats (2006) URL:<http://www.internetworldstats.com>

Johnson, J. A. (2000, March). Web-based personality assessment. Paper presented at the 71st Annual Meeting of the Eastern Psychological Association, Baltimore, MD.

Joinson, A. (1999). Social desirability, anonymity, and Internet-based questionnaires. *Behavior Research Methods, Instruments, & Computers*, 31, 433-438.

Kanji, G. K. (1993). *100 statistical tests*. London: Sage Publications.

Lievens, F., & Harris, M. M. (2003). Research on Internet recruiting and testing: Current status and future directions. In C.L. Cooper & I.T. Robertson (Eds.) *International Review of Industrial and Organizational Psychology*, 18 (pp. 131-165). Chichester, England: John Wiley.

Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114, 449-458.

Milas, G. (1998). Korelacijska i eksperimentalna studija odnosa temeljnih dimenzija društvenih stavova, ličnosti i društvene poželjnosti [Correlational and experimental study of relationships between basic dimensions of social attitudes, personality and social desirability]. Doctoral dissertation. Zagreb: Faculty of Philosophy, Department of Psychology.

Pettit, F. A. (1999). Exploring the use of the World Wide Web as a psychology data collection tool. *Computers in Human Behavior*, 15, 67-71.

Ployhart, R. E., Weekley, J. A., Holz, B. C., & Kemp, C. F. (2003). Web-based and paper-and-pencil testing of applicants in proctored setting: Are personality, biodata, and situational judgment tests comparable? *Personnel Psychology*, 56, 733-752.

Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology*, 84, 754-775.

Woolhouse, L., & Myers, S. (1999, September). Factors affecting sample make-up: Results from an Internet-based personality questionnaire. Paper presented at the British Psychological Society: Social Psychology Section Conference, Lancaster, UK.

Table 1. Cronbach's alpha coefficient of internal consistency for five personality scales in three different item presentation modes

Scale	α (p-p)	α (e-pp)	α (e-ibi)
Extraversion	.93	.92	.94
Agreeableness	.89	.90	.89
Conscientiousness	.92	.91	.92
Neuroticism	.96	.94	.95
Intellect	.79	.77	.75

Note:

p-p – paper-and-pencil questionnaire

e-p-p – electronic version of questionnaire identical to paper-and-pencil version

e-ibi – electronic version of questionnaire with item-by-item presentation

Table 2. Pearsons correlation coefficients between different questionnaire versions for the five personality dimensions

Scale		e-p-p	e-ibi
Extraversion	p-p	.92	.92
	e-p-p	1	.94
	e-ibi		1
Agreeableness	p-p	.92	.92
	e-p-p	1	.89
	e-ibi		1
Conscientiousness	p-p	.85	.86
	e-p-p	1	.92
	e-ibi		1
Neuroticism	p-p	.92	.91
	e-p-p	1	.89
	e-ibi		1
Intellect	p-p	.84	.83
	e-p-p	1	.88
	e-ibi		1

Note: All correlations are significant at $p < 0.01$;

p-p – paper-and-pencil questionnaire

e-p-p – electronic version of questionnaire identical to paper-and-pencil version

e-ibi – electronic version of questionnaire with item-by-item presentation

Table 3. Means, standard deviations and difference significance between each questionnaire mode for five traits

Scale	p-p		e-pp		e-ibi		F	p
	M	SD	M	SD	M	SD		
Extraversion	73.8	12.16	74.4	11.28	73.9	12.39	0.87	.42
Agreeableness	80.9	8.52	80.7	8.64	81.5	8.64	1.82	.17
Conscientiousness	68.0	13.07	69.3	12.49	68.9	12.71	1.41	.25
Neuroticism	56.3	14.95	54.6	13.76	55.8	15.16	3.54	.03
Intellect	73.1	6.96	72.9	6.60	73.2	6.29	0.50	.61

Note:

p-p – paper-and-pencil questionnaire

e-p-p – electronic version of questionnaire identical to paper-and-pencil version

e-ibi – electronic version of questionnaire with item-by-item presentation

Table 4. Correlations between different questionnaire versions for three social desirability measures and five personality traits

Scale		Social desirability measures		
		L-scale	Self-deception	Impression-management
Extraversion	p-p	.02	.41**	-.01
	e-p-p	.07	.45**	.01
	e-ibi	.00	.35**	-.09
Agreeableness	p-p	.14	.00	.13
	e-p-p	.16	.05	.15
	e-ibi	.19	.01	.20
Conscientiousness	p-p	.51**	.16	.42**
	e-p-p	.41**	.15	.33**
	e-ibi	.40**	.10	.38**
Neuroticism	p-p	.07	-.27*	.17
	e-p-p	.03	-.36**	.06
	e-ibi	.02	-.30**	.10
Intellect	p-p	.12	.41**	.07
	e-p-p	-.01	.27*	-.09
	e-ibi	.05	.25*	.00

Note: ** Correlation significant at $p < 0.01$; *correlation significant at $p < 0.05$

p-p – paper-and-pencil questionnaire

e-p-p – electronic version of questionnaire identical to paper-and-pencil version

e-ibi – electronic version of questionnaire with item-by-item presentation