

## Assessing the Equivalence of Computerized and Conventional Versions of Psychological Tests

**Daria S. Gnedykh**

Saint-Petersburg University, St. Petersburg, Russian Federation

E-mail: [d.gnedykh@spbu.ru](mailto:d.gnedykh@spbu.ru)

ORCID ID: <https://orcid.org/0000-0003-4955-4779>

---

### Abstract

**Introduction.** This paper (a) discusses the theoretical and methodological evidence for the equivalence of computerized and conventional versions of psychological tests, (b) analyzes the studies investigating psychometric parameters of computerized versions of conventional tests, and (c) examines contradictions in approaches to assessing the equivalence of the two test forms. This paper represents a first effort in structuring the main problems in establishing the equivalence of paper-and-pencil and computer-based testing, as well as in finding ways and means to overcome them. Much attention is devoted to minimal and sufficient mathematical and statistical tools for assessing the equivalence of the two test forms.

**Theoretical Basis.** The main problems associated with assessing the equivalence of computerized and conventional tests include the following: (a) the level of cultural and informational competence of respondents, (b) anxiety, (c) social environment, (d) motivation for testing, (e) difficulties in creating the same conditions for paper-and-pencil and computer-based testing (L. N. Babanin, Y. P. Chua, M. Russell, P. Květon et al.). Researchers most often choose the following procedures for assessing the equivalence of the two versions of psychological tests: (a) comparison of the main statistical parameters (mean values, variances, etc.) and (b) assessment of construct validity and reliability of the computer-based versions of psychological tests.

**Results and Discussion.** The analysis of research methodology for assessing the equivalence of computerized and conventional versions of psychological tests focuses on a variety of approaches to (a) the use of mathematical and statistical methods for assessing psychometric parameters of computer-based versions of conventional tests, (b) the choice of research design, (c) considering specific characteristics of the situation of computer-based testing. It is necessary to formulate specific and structured requirements for the procedure for assessing the equivalence of computerized and conventional versions of psychological tests. The author suggests recommendations for the main sections of such requirements related to (a) the procedure for conducting empirical research, (b) mathematical and statistical methods, and (c) control of the factors specific to computer-based testing that may have impact on the results of equivalence assessment.

### Keywords

psychological tests, psychometric parameters, equivalence of tests, mathematical and statistical methods, computerization of tests, computer-based testing, computer psychodiagnostics, research methodology, validity, test reliability

## Highlights

- The lack of the standardized assessment of psychometric equivalence of paper-and-pencil and computer-based tests is a methodological issue.
  - The standardized procedure for assessing the equivalence of paper-and-pencil and computer-based tests should impose uniform requirements for research design, as well as for methods of mathematical and statistical data processing.
  - To consider specific factors of a computer-based testing situation, researchers are advised to use observation cards or questionnaires for identifying the level of computer literacy in respondents, their motivation and attitudes towards computer-based psychodiagnostics.
- 

## For citation

Gnedych, D. S. (2020). Assessing the equivalence of computerized and conventional versions of psychological tests. *Rossiiskii psikhologicheskii zhurnal (Russian Psychological Journal)*, 17(1), 44–59. doi: 10.21702/rpj.2020.1.4

Received: January 28, 2020

Revision received: February 29, 2020

Accepted: March 4, 2020

## Introduction

The increased pace of life, large amount of information, and a wide range of technical possibilities dictate the need for time savings during psychological research. As a result, computer-based testing replaces paper-and-pencil (conventional) formats of psychological tests, which makes it possible to remotely collect information and to automate the processing of data. Various services, including Google forms, online survey services, and specialized programs are used for that purpose. In most cases the replacement of a paper-and-pencil format by a computer-based one is reduced to a simple action – the text of a questionnaire (or another kind of stimulus materials) is presented in an electronic format and sent to all the respondents. In a similar way, the norms used for interpreting paper-and-pencil tests are applied to their computer-based counterparts.

Researches have investigated the issues of the equivalence of computer-based and conventional tests for a long time. Mazzeo, Druesne, Raffeld, Checketts, & Muhlstein (1992) argue that special studies of the comparability of indicators of these formats need to be conducted. Previous studies in this field introduced a number of fundamental contradictions regarding psychometric parameters that should be assessed. Hypothetically, the validity of the two test forms will be equivalent, if the two test forms are proved to be equivalent (George, Lankford, & Wilson, 1992; Ford, Vitelli, & Stuckless, 1996). However, L. M. Honaker made the point that validity indices from a paper-and-pencil version cannot be automatically transferred to a computer-based one (Honaker, 1988). Such cases require additional validity assessments (Russell, Goldberg, & O'connor, 2003). Anastasi and Urbina (2009, p. 93) agree with him saying, '... the reliability and validity of the test can vary depending on the format of presentation'. Baturin and Melnikova (2011a, 2011b) note that the procedure for creating a computer-based version of a conventional test is not a simple process of copying stimulus materials from one format to another, but its *modification*, which implies the process of its processing. A number of other scholars share this opinion, saying 'the computer-based version is a completely independent test, which use is possible only after

the assessment of group norms and other psychometric characteristics' (Vasserman, Iovlev, & Chervinskaya, 2010, p. 23). Moreover, some researchers suggest that the coincidence of scores from paper-and-pencil and computer-based testing does not mean that both forms measure the same psychological construct, as supported by a number of personality tests (Meade, Michels, & Lautenschlager, 2004). Despite the fact that this issue has drawn attention at the end of the 20th century, not all researchers, as we will show, carry out a complete assessment of the psychometric parameters of computer-based tests. The mathematical and statistical methods for analyzing their equivalence are quite diverse. The choice of these methods is often not explained by authors.

Diverse opinions regarding psychometric parameters of computerized tests that require verification are partly explained by the lack of clear standardized instructions for the procedure for establishing the equivalence between paper-and-pencil and computer-based versions of psychological tests. Certain recommendations regarding the development of psychodiagnostic tools make it possible to structure the information accumulated in this area and facilitate the work of psychologists and other specialists in this field. A series of articles by Baturin and Melnikova (2009–2011); American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999) (AERA, APA, NCME) seem to be a good example. Assessing the equivalence of paper-and-pencil and computerized tests is a specific area of psychodiagnostics, which requires certainty and standards.

Obviously, the computerization of paper-and-pencil tests will be a massive process in the future. At the same time, researchers and scholars are growing concerned about the validity of computer-based versions of paper-and-pencil tests. Thus, there is a need to standardize and control the process of converting paper-and-pencil tests to computerized formats.

This paper discusses the experience of assessing psychometric parameters of computerized versions of conventional psychological tests, concentrates on the minimum set of mathematical and statistical methods for assessing the equivalence of the two test forms, and suggests the need to formalize the procedure for establishing specific conditions of computer-based testing (directly during the assessment of equivalence) and their subsequent accounting when interpreting the results.

### **Theoretical Basis**

Researchers and scholars actively discuss the influence of various factors on the process and procedure for computer-based testing. Pointing to factors that can affect the equivalence of the two test forms (experience with various computer-based applications, anxiety, social environment, etc.) Babanin (2010) concludes that the level of general cultural and information competence (level of mastering information technology) is the main one. However, the computer literacy of respondents still does not solve the problem of differences in data when assessing the equivalence of computer-based and paper-and-pencil testing (Russell et al., 2003). When converting paper-and-pencil tests to computer-based formats, the conditions for computerized tests should fully correspond to those for their paper-and-pencil counterparts (time for presenting stimulus materials, possibility of correcting answers, strict/non-strict order of answers, etc.). This is one way to increase the equivalence of these formats.

The choice of mathematical and statistical data analysis methods is the next point that requires special attention. What measures can be considered necessary and sufficient to verify the psychometric equivalence of computerized and paper-and-pencil testing?

The Guidelines for Computer-based Tests and Interpretations (APA, 1986) contains the main statistical methods and indicators that determine the psychometric equivalence of computer-based and conventional test forms, including descriptive statistics (mean values, variances, distributions, and rank orders), construct validity, and reliability. Nevertheless, when assessing equivalence the authors do not always observe all these points and use additional types of analysis in certain cases (i.e. Bartram, 1994; van de Looij-Jansen, Goldschmeding, & Jan de Wilde, 2006; Květon, Jelínek, Vobořil, & Klimusová, 2007; Chua, 2012). So what determines the choice of methods for assessing psychometric characteristics of computer-based versions of conventional psychological tests? What is the minimum list of methods that are sufficient to prove the equivalence of paper-and-pencil and computer-based test formats, or the independence of the computer-based format as a tool?

Table 1 provides an overview of studies of the equivalence of computer-based and paper-and-pencil testing over the past 20 years. The analysis criteria were as follows: (a) research design, (b) use of certain statistical methods to process the results, (c) explanations for their choice or purposes of their use, and (d) specific characteristics of computer-based testing procedure (observation of respondents during computer-based testing, additional questionnaires to identify the attitudes of study participants to computer-based testing if it is carried out via the Internet, etc.).

## Results and Discussion

As can be seen from Table 1, when comparing computer-based and paper-and-pencil formats of the same test, researchers consider different aspects. By accepting the fact that a computer-based test is valid, some of them only assess its reliability. The others consider it sufficient to carry out correlation analysis and comparison of means to prove equivalence. Still others prefer to immediately assess all the psychometric parameters of a computer-based test form. Can the choice of statistical methods depend on specific characteristics of stimulus materials? It seems that if the technique consists of questions (statements) and answer options (like a questionnaire), then converting a paper-and-pencil test to a computer-based format will not entail serious changes. Some researchers have come to this conclusion. Thus, reliability of multiscale tests is preserved when they were converted from a paper-and-pencil form to a computerized one (Romek & Satin, 2000). On the other hand, if stimulus materials represent a creative task, or the technique involves working with pictures, graphic symbols, and is time-limited, etc. (i.e. some cognitive tests), psychometric parameters of computer-based formats of such tests require a more responsible processing. The table shows that this logic does not always underlie the choice of statistical methods for assessing the equivalence of the two test form.

Moreover, there is a variety of approaches to research design when assessing the equivalence of the two test form (computer-based and conventional). Thus, Romek and Satin (2000), van de Looij-Jansen et al. (2006) use intergroup comparisons (when one group takes only a paper-and-pencil test and another – only a computer-based one) (see also Andersson, Kaldo-Sandström, Ström, & Strömgren, 2003; Vecchione, Alessandri, & Barbaranelli, 2012). Ermakov (2016) and Kibalchenko, Ustinov, & Shapovalov (2004) use intragroup comparisons, when representatives of the same group take both paper-and-pencil and computer-based tests after a certain period of time (also see Hays & McCallum, 2005). Both intergroup and intragroup comparisons are also used (Květon et al., 2007; Kononova & Nakhaeva, 2013). Each of the assessment options described above has its advantages and disadvantages.

Table 1	Comparing computer-based and paper-and-pencil (conventional) psychodiagnostic tests: An analysis of existing studies					
Author, year	<u>Psycho-diagnostic test, brief description of stimulus materials*</u>	<u>Research design**</u>	<u>Mathematical and statistical data processing methods</u>	<u>Aim of mathematical statistical analysis</u>	<u>The procedure for considering specific characteristics of computer-based testing (observation or survey)</u>	<u>The main conclusions concerning the equivalence of the two test forms</u>
V. G. Romek, D. K. Satin, 2000	Eysenck Personality Questionnaire (Q; TUL; VERB; PERS) Self-confidence Test (Q; TUL; VERB; PERS)	Intergroup comparison  Remotely (via the Internet); in the presence of a specialist	Comparing the factorial structures of the test; reliability index (Cronbach's a); comparing means and variances; test-retest reliability	Comparing reliability; assessing the difference between the norms	Data on the procedure for considering specific characteristics of a situation of computer-based testing are not provided	Psychometric characteristics of tests are preserved; reliability is not reduced
I. A. Kibalchenko et al., 2004	Assessment of Interests in Primary School Children (Q; TUL; VERB; PERS) Map of Giftedness (Q; TUL; VERB; PERS)	Intragroup comparison	Spearman correlation coefficient	Assessing test-retest reliability	Data on the procedure for considering specific characteristics of a situation of computer-based testing are not provided	High-level representative reliability of computer-based versions; computer-based versions can be used instead of paper-and-pencil ones

**Table 1**  
Comparing computer-based and paper-and-pencil (conventional) psychodiagnostic tests: An analysis of existing studies

<u>Author, year</u>	<u>Psycho-diagnostic test, brief description of stimulus materials*</u>	<u>Research design**</u>	<u>Mathematical and statistical data processing methods</u>	<u>Aim of mathematical statistical analysis</u>	<u>The procedure for considering specific characteristics of computer-based testing (observation or survey)</u>	<u>The main conclusions concerning the equivalence of the two test forms</u>
B. V. Iovlev et al., 2006 (Iovlev, Novozhilova, Chervinskaya, & Shchelkova, 2006)	Neurotic Personality Questionnaire (Q; TUL; VERB; PERS)	Intragroup and intergroup comparison	Ranking of average means of questionnaire scales in each group; correlation analysis	Not provided	Data on specific characteristics of a situation of computer-based testing are not provided	Paper-and-pencil and computer-based versions are equivalent
P. M. van de Looij-Jansen et al., 2006	Youth Health Monitor Rotterdam, YMR (Q; TUL; VERB; PERS)	Intergroup comparison; in the presence of a teacher	Two-way analysis of covariance (ANCOVAs)	Comparing the results of paper-and-pencil and computer-based tests	Data on specific characteristics of a situation of computer-based testing are not provided	The majority of test scores have no significant differences

Table 1	Comparing computer-based and paper-and-pencil (conventional) psychodiagnostic tests: An analysis of existing studies					
Author, year	Psychodiagnostic test, brief description of stimulus materials*	Research design**	Mathematical and statistical data processing methods	Aim of mathematical statistical analysis	The procedure for considering specific characteristics of computer-based testing (observation or survey)	The main conclusions concerning the equivalence of the two test forms
P. Květon et al., 2007	Bourdon Test (GS; TL; NVERB; COG)	Intergroup comparison	One-way analysis of variance; Tukey test for pairwise comparisons; Cronbach's $\alpha$	Comparing the results of paper-and-pencil and computer-based tests; comparing reliability	To consider the influence of conditions of different designs of computer-based tests were used	Paper-and-pencil and computer-based tests substantially differ from each other
P. Květon et al., 2007	Test of Concentration of Attention, TCA (GS; TL; NVERB; COG)	Intragroup and intergroup comparison; in the presence of a specialist	General Linear Model (GLM); Test of Between-Subjects Effects from the Multivariate GLM; Tukey test for pairwise comparisons; correlation analysis	Assessing the influence of testing order (C-P; P-C; P-P; C-C)*** and its form on the results; assessing reliability	Data on specific characteristics of a situation of computer-based testing are not provided	Test forms are not equivalent. Specific effects of the form and order of testing were found to affect test results

**Table 1**  
 Comparing computer-based and paper-and-pencil (conventional) psychodiagnostic tests: An analysis of existing studies

Author, year	Psychodiagnostic test, brief description of stimulus materials*	Research design**	Mathematical and statistical data processing methods	Aim of mathematical statistical analysis	The procedure for considering specific characteristics of computer-based testing (observation or survey)	The main conclusions concerning the equivalence of the two test forms
Y. P. Chua, 2012	Yanpiaw Creative-Critical Styles test, YBRAINS (G; TUL; VERB; COG)	Intragroup and intergroup comparison	Independent samples t-test; split-plot ANOVA; test-retest reliability (Pearson product moment coefficients); Cronbach's a	Impact of the effect of testing and the test form on results; test-retest reliability; comparing reliability	Data on specific characteristics of a situation of computer-based testing are not provided	High validity of both test forms; the paper-and-pencil version turned out to be more sensitive to testing time and motivation for testing

<p>Table 1 Comparing computer-based and paper-and-pencil (conventional) psychodiagnostic tests: An analysis of existing studies</p>	<p><u>Psycho-diagnostic test, brief description of stimulus materials*</u></p> <p><u>Research design**</u></p> <p><u>Mathematical and statistical data processing methods</u></p> <p><u>Aim of mathematical statistical analysis</u></p> <p><u>The procedure for considering specific characteristics of computer-based testing (observation or survey)</u></p> <p><u>The main conclusions concerning the equivalence of the two test forms</u></p>
<p>P. Žitný et al., 2012 (Žitný, Halama, Jelínek, &amp; Květon, 2012)</p> <p>Test of Intellect Potential, TIP (GS; TL; NVERB; COG)</p> <p>Vienna Matrices Test (GS; TL; NVERB; COG)</p>	<p>Correlation analysis; Fisher z-transformation; Cohen's d effect size on the basis of independent samples t-test</p> <p>Intragroup and intergroup comparison</p> <p>Comparing means of test scores depending on gender and place of residence; comparing criterion and construct validity</p> <p>Data on specific characteristics of a situation of computer-based testing are not provided</p> <p>Construct validity of all test forms is proved; the results of the computer-based version of the test are comparable to those of a conventional one</p>
<p>V. N. Kononova, I. V. Nakhaeva, 2013</p> <p>Color Mirror (PM; MT; TUL; NVERB; PERS)</p>	<p>Phi correlation coefficient, chi-square; correlation analysis</p> <p>Intergroup and intragroup comparison; in the presence of a specialist</p> <p>Analysis of the ratio of test scores; test-retest reliability</p> <p>Individual interaction with a psychologist during testing</p> <p>Testing forms are not fully equivalent</p>

**Table 1**  
Comparing computer-based and paper-and-pencil (conventional) psychodiagnostic tests: An analysis of existing studies

Author, year	Psycho-diagnostic test, brief description of stimulus materials*	Research design**	Mathematical and statistical data processing methods	Aim of mathematical statistical analysis	The procedure for considering specific characteristics of computer-based testing (observation or survey)	The main conclusions concerning the equivalence of the two test forms
O. V. Mitina, V. V. Sorokina, 2015 (Mitina & Sorokina, 2015)	Schwartz Value Inventory, SVI (Q; TUL; VERB; PERS)	Intergroup comparison; in the presence of a psychologist	Cronbach's $\alpha$ ; Student's t-test; Mann-Whitney U test; Levene's test of equality of variances; correlation analysis	Assessing reliability; comparing the results in subsamples by gender; assessing the validity of the computer-based version	Data on specific characteristics of a situation of computer-based testing are not provided	Psychometric reliability and validity of the computer-based version is proved
S. S. Ermakov, 2016	Raven's Standard Progressive Matrices Plus, SPM+ (GS; TL; NVERB; COG)	Intragroup comparison	Correlation coefficient; Student's t-test	Assessing test-retest reliability; assessing differences between groups	Data on specific characteristics of a situation of computer-based testing are not provided	Computer-based version of the test can be considered a counterpart of its paper-and-pencil version

<p>Table 1 Comparing computer-based and paper-and-pencil (conventional) psychodiagnostic tests: An analysis of existing studies</p>	<p>Author, year</p>	<p>Psychodiagnostic test, brief description of stimulus materials*</p>	<p>Research design**</p>	<p>Mathematical and statistical data processing methods</p>	<p>Aim of mathematical statistical analysis</p>	<p>The procedure for considering specific characteristics of computer-based testing (observation or survey)</p>	<p>The main conclusions concerning the equivalence of the two test forms</p>
	<p>K. V. Sugaev et al., 2018</p>	<p>SIT-30 test, a shortened and modified version of the Short Indicative Test (Q; GS; TL; VERB; NVERB; COG)</p>	<p>Intergroup comparison; Internet version; in the presence of a specialist (both tests)</p>	<p>Correlation and factorial analysis, comparing of means, measures of variability, and instantaneous reliability; comparing item discrimination and difficulty coefficients, their factor loadings on the general factor</p>	<p>Comparing testing results from a controlled and uncontrolled Internet formats.</p>	<p>Data on specific characteristics of a situation of computer-based testing are not provided</p>	<p>Psychometric characteristics of test performance via the Internet are not inferior to those calculated under controlled testing</p>

Note: \* Q – questionnaire, GS – graphic symbols, signs, figures; PM – projective technique; MT – manipulation technique; TL – time-limited testing; TUL – time-unlimited testing; VERB – verbal test; NVERB – non-verbal test; COG – cognitive characteristics; PERS – personal characteristics;

\*\* research design by the following parameters: intragroup/intergroup comparison; remote testing/in the presence of a psychologist (if the article provides the data);

\*\*\* C – computer-based version of the technique; P – paper-and-pencil version of the technique.

Moreover, in our opinion, intragroup comparisons make it possible to control the influence of a respondent's personality when he/she takes both paper-and-pencil and computer-based tests. This kind of influence can manifest itself in an intergroup design situation, when the difference in results can be explained not by the test form but by differences in personal characteristics of two groups of respondents (i.e. if one group comprises more anxious individual participants than another).

When taking computer-based tests additional factors also need to be considered for a thorough assessment of equivalence (Chua, 2012). In this regard, the table included such a criterion for analyzing the content of articles as 'the procedure for taking into account specific characteristics of computer-based testing (observation or survey)'. Even a relatively minor modification in a computer-based test design (i.e. changing a color scheme) can substantially affect the results (Květon et al., 2007). It is also believed that compared to paper-and-pencil tests, taking the computer-based ones is associated with a quite different cognitive load, which leads to 'stratification of the sample, when the worst show even worse results and the best demonstrate even better results' (Ermakov, 2016, p. 203). Moreover, the pre-test and post-test designs may lead to the effect of testing (the experience of passing the pre-test affects the results of the post-test); overlooking this fact researchers may conclude that external conditions affected the results of the post-test (i.e. a computer, if it was the second). To avoid incorrect conclusions related to the effect of testing, Chua (2012) suggested R. L. Solomon's experimental plan for four randomized groups (for details see Solomon, 1949; Campbell & Stanley, 1963; et al.). Another problem associated with taking computer-based tests remotely (via the Internet) is falsification of the results (Sugonyaev, Radchenko, & Sokolov, 2018). It is recommended to recheck such results under controlled conditions (The International Testing Commission, 2006).

A more thorough research design – choosing a number of respondent groups, the sequence of testing series, etc. – requires consideration of all these points. We believe that it will be useful to organize observation of respondents' behaviors when they take computer-based tests. This will help to eliminate the results of respondents demonstrating a high level of anxiety during testing, a low level of motivation or their incompetence when interacting with computer interface (information incompetence). Such an observation can help modify the procedure for computer-based testing or the presentation of stimulus material, if most respondents show difficulties during it. All this may be organized at the stage of programming a computer-based version of a conventional test in order to exclude the influence of some factors on the results of equivalence of its two forms.

Considering the problem of psychometric properties of computer-based versions of conventional tests, K. V. Sugonyaev et al. provides a specific procedure for statistical analysis; we believe that it may help establish the equivalence of the two test forms. The researchers argue that comparing 'measures of central tendency, variability, and simultaneous reliability of integral test scores' (Sugonyaev et al., 2018, p. 8) is insufficient for assessing the comparability of test results. The analysis of test performance at the item level – similarity/difference in the 'distribution of coefficients of item difficulty and discrimination, as well as in the patterns of item factor loadings on the general factor' (Sugonyaev et al., 2018, p. 18) – may be a method designed to prove the equivalence of the two test form. Thus, the authors attempted to explain and prove the necessary minimum of statistical methods for assessing the equivalence between the computerized and paper-and-pencil tests. In the remaining articles (Table 1), as a rule, the researchers do not explain the choice of methods of mathematical and statistical analysis (as

well as the research design (intergroup/intragroup)) from the point of view of their sufficiency for assessing equivalence. And yet, as the conditions for taking a computer-based test differ from those for a paper-and-pencil one, and taking into account all the factors listed above that can influence the performance of a computer-based test, we suggest not abandoning the idea of assessing the validity and reliability of computerized versions of conventional tests.

The analysis of studies related to assessing the equivalence of computer-based and pencil-and-paper tests allows us to conclude that most of them are published by foreign experts. Despite numerous programs for computer-based psychodiagnostics (Mel'nichuk & Sergeev, n.d.; Budko, Mishin, & Tregubova, 2007), few Russian studies address the procedure for assessing psychometric properties of computer-based versions of conventional tests. To make matters worse, a small number of articles examine the quality of assessing the equivalence of computerized versions and their conventional paper-and-pencil counterparts.

### **Conclusion**

Without a doubt, psychometric parameters of computer-based versions of conventional tests should be assessed. Therefore, *distinct requirements for the procedure for such assessments and standardized instructions for specialists involved in programming computer-based versions of conventional tests represent an important aspect of research.*

Obviously, brief instructions presented in the Guidelines for Computer-based Tests and Interpretations (APA, 1986), or the detailed however inconcrete ones from Standards for educational and psychological testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) are insufficient.

We assume that these recommendations should contain distinct instructions for each of the following sections:

1. Research design requirements necessary for a thorough assessment of the equivalence of paper-and-pencil and computer-based versions of psychodiagnostic tests (number of groups, minimum sample size, sequence of various test forms, time interval between repeated measurements, methods for assessing construct validity, etc.).
2. The list of necessary and sufficient methods of mathematical and statistical data processing that help establish the equivalence between paper-and-pencil and computer-based tests.
3. Description of factors that may affect the results of computer-based and paper-and-pencil tests and recommendations for certain conditions for assessing equivalence. The methods to ensure these factors include observation cards for respondents' behaviors during testing and standardized questionnaires that assess motivation, attitudes towards computer diagnostics, the level of computer literacy, etc. (Testing Motivation Questionnaire (Chua, 2012), Computer Familiarity Questionnaire (Mazzeo et al., 1992)). This will help eliminate the influence of additional variables on the test results and obtain more reliable results when assessing the equivalence of conventional and computer-based psychodiagnostic tests.

### **References**

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Anastazi, A., & Urbina, S. (2009). *Psychological testing*. St. Petersburg: Piter. (in Russ.).

- Andersson, G., Kaldo-Sandström, V., Ström, L., & Strömberg, T. (2003). Internet administration of the Hospital Anxiety and Depression Scale in a sample of tinnitus patients. *Journal of Psychosomatic Research*, 55(3), 259–262. doi: [10.1016/S0022-3999\(02\)00575-5](https://doi.org/10.1016/S0022-3999(02)00575-5)
- APA (1986). *Guidelines for computer-based tests and interpretations*. Washington, DC: Author.
- Babanin, L. N. (2010). Equivalence of conventional and computerized psychological techniques. *Experimental psychology in Russia: Traditions and prospects*. Retrieved from [http://psyjournals.ru/exp\\_collection/issue/32974.shtml](http://psyjournals.ru/exp_collection/issue/32974.shtml) (in Russ.).
- Bartram, D. (1994). Computer-based assessment. In C. L. Cooper (Ed.), *International review of industrial and organizational psychology* (pp. 31–69). London: Wiley.
- Baturin, N. A., & Mel'nikova, N. N. (2009). Technology of test development: Part I. *Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta. Seriya: Psikhologiya (Bulletin of the South Ural State University. Series: Psychology)*, 6(30), 4–14. (in Russ.).
- Baturin, N. A., & Mel'nikova, N. N. (2009). Technology of test development: Part II. *Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta. Seriya: Psikhologiya (Bulletin of the South Ural State University. Series: Psychology)*, 7(42), 11–25. (in Russ.).
- Baturin, N. A., & Mel'nikova, N. N. (2010). Technology of test development: Part III. *Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta. Seriya: Psikhologiya (Bulletin of the South Ural State University. Series: Psychology)*, 8(4), 4–18. (in Russ.).
- Baturin, N. A., & Mel'nikova, N. N. (2010). Technology of test development: Part IV. *Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta. Seriya: Psikhologiya (Bulletin of the South Ural State University. Series: Psychology)*, 11(40), 13–28. (in Russ.).
- Baturin, N. A., & Mel'nikova, N. N. (2011a). Technology of test development: Part V. *Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta. Seriya: Psikhologiya (Bulletin of the South Ural State University. Series: Psychology)*, 12(5), 4–14. (in Russ.).
- Baturin, N. A., & Mel'nikova, N. N. (2011b). Technology of test development: Part VI. *Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta. Seriya: Psikhologiya (Bulletin of the South Ural State University. Series: Psychology)*, 13(18), 48–59. (in Russ.).
- Budko, V. N., Mishin, D. Yu., & Tregubova, T. A. (2007). Testing software package for readiness of students and schoolchildren for studying at university. *Infokommunikatsionnye tekhnologii (Infocommunication Technologies)*, 5(3), 167–169. (in Russ.).
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally & Company.
- Chua, Y. P. (2012). Effects of computer-based testing on test performance and testing motivation. *Computers in Human Behavior*, 28(5), 1580–1586. doi: [10.1016/j.chb.2012.03.020](https://doi.org/10.1016/j.chb.2012.03.020)
- Ermakov, S. S. (2016). Computer diagnostics of intellectual abilities among 4th to 9th grades school students. *Psikhologicheskaya nauka i obrazovanie (Psychological Science and Education)*, 8(4), 199–207. doi: [10.17759/psyedu.2016080419](https://doi.org/10.17759/psyedu.2016080419) (in Russ.).
- Ford, B. D., Vitelli, R., & Stuckless, N. (1996). The effects of computer versus paper-and-pencil administration on measures of anger and revenge with an inmate population. *Computers in Human Behavior*, 12(1), 159–166. doi: [10.1016/0747-5632\(95\)00026-7](https://doi.org/10.1016/0747-5632(95)00026-7)
- George, C. E., Lankford, J. S., & Wilson, S. E. (1992). The effects of computerized versus paper-and-pencil administration on measures of negative affect. *Computers in Human Behavior*, 8(2–3), 203–209. doi: [10.1016/0747-5632\(92\)90004-X](https://doi.org/10.1016/0747-5632(92)90004-X)
- Hays, S., & Mccallum, R. S. (2005). A comparison of the pencil-and-paper and computer-administered

- Minnesota Multiphasic Personality Inventory–Adolescent. *Psychology in the Schools*, 42(6), 605–613. doi: [10.1002/pits.20106](https://doi.org/10.1002/pits.20106)
- Honaker, L. M. (1988). The equivalency of computerized and conventional MMPI administration: A critical review. *Clinical Psychology Review*, 8(6), 561–577. doi: [10.1016/0272-7358\(88\)90081-5](https://doi.org/10.1016/0272-7358(88)90081-5)
- Iovlev, B. V., Novozhilova, M. Yu., Chervinskaya, K. R., & Shchelkova, O. Yu. (2006). Studying the effectiveness of computer psychodiagnostics: methodological perspectives. *Vestnik Sankt-Peterburgskogo universiteta. Ser. 6. Filosofiya, politologiya, sotsiologiya, psikhologiya, pravo, mezhdunarodnye otnosheniya (Vestnik of Saint-Petersburg University, Series 6: Philosophy. Political Science. Law. International Relations)*, 2, 115–124. (in Russ.).
- Kibal'chenko, I. A., Ustinov, D. A., & Shapovalov, S. N. (2004). Computer psychodiagnostics as a condition for the development of giftedness. *Izvestiya TRTU (Proceedings of TRTU)*, 6, 301–303. (in Russ.).
- Kononova, V. N., & Nakhaeva, I. V. (2013). Psychometric equivalence of conventional (Color Mirror) and computerized (Colorimeter of Personality) versions of the projective psychodiagnostic technique. *Vestnik Moskovskogo universiteta. Seriya 14. Psikhologiya (Moscow University Psychology Bulletin. Series 14: Psychology)*, 1, 136–151. (in Russ.).
- Květon, P., Jelínek, M., Vobořil, D., & Klimusová, H. (2007). Computer-based tests: The impact of test design and problem of equivalency. *Computers in Human Behavior*, 23(1), 32–51.
- Mazzeo, J., Druesne, B., Raffeld, P. C., Checketts, K. T., & Muhlstein, A. (1992). Compatibility of computer and paper-and-pencil scores for two CLEP® general examinations. *ETS Research Report Series*, 1. doi: [10.1002/j.2333-8504.1992.tb01446.x](https://doi.org/10.1002/j.2333-8504.1992.tb01446.x)
- Meade, A. W., Michels, L. C., & Lautenschlager, G. J. (2004, April). Are online and paper-and-pencil personality tests truly comparable? *Symposium presented at the 19th Annual Conference of the Society for Industrial and Organizational Psychology*. Chicago, IL.
- Mel'nichuk, A., & Sergeev, V. (n.d.). Modern computer systems for effective personnel assessment. *Psikhologiya i biznes (Psychology and Business)*. Retrieved from <https://psycho.ru/library/93> (in Russ.).
- Mitina, O. V., & Sorokina, V. V. (2015). Values of senior pupils: Development of computer diagnostic tools. *Vestnik Moskovskogo universiteta. Seriya 14. Psikhologiya (Moscow University Psychology Bulletin. Series 14: Psychology)*, 1, 42–59. (in Russ.).
- Romek, V. G., & Satin, D. K. (2000). Maintaining reliability of multifactor tests when using them via the Internet. *Psikhologicheskii Zhurnal*, 21(2), 70–75. (in Russ.).
- Russell, M., Goldberg, A., & O'connor, K. (2003). Computer-based Testing and Validity: A look back and into the future. *Assessment in Education: Principles, Policy & Practice*, 10(3), 279–293. doi: [10.1080/0969594032000148145](https://doi.org/10.1080/0969594032000148145)
- Solomon, R. L. (1949). An extension of control group design. *Psychological Bulletin*, 46(2), 137–150. doi: [10.1037/h0062958](https://doi.org/10.1037/h0062958)
- Sugonyaev, K. V., Radchenko, Yu. I., & Sokolov, A. A. (2018). Voluntary Internet-based testing as a source of valid group psychometric intelligence scores. *Sibirskii psikhologicheskii zhurnal (Siberian Journal of Psychology)*, 69, 6–32. doi: [10.17223/17267080/69/1](https://doi.org/10.17223/17267080/69/1) (in Russ.).
- The International Testing Commission. (2006). International guidelines on computer-based and Internet-delivered testing. *International Journal of Testing*, 6(2), 143–171. doi: [10.1207/s15327574ijt0602\\_4](https://doi.org/10.1207/s15327574ijt0602_4)
- van de Looij-Jansen, P. M., Goldschmeding, J. E. J., & Jan de Wilde, E. (2006). Comparison of anonymous versus confidential survey procedures: Effects on health indicators in Dutch adolescents. *Journal of Youth and Adolescence*, 35, 652–658. doi: [10.1007/s10964-005-9027-0](https://doi.org/10.1007/s10964-005-9027-0)

Gnedykh

Assessing the Equivalence of Computerized and Conventional Versions of Psychological Tests

**RUSSIAN PSYCHOLOGICAL JOURNAL**, 2020, Vol. 17, No. 1, 44–59. doi: 10.21702/rpj.2020.1.4

GENERAL PSYCHOLOGY, PSYCHOLOGY OF PERSONALITY, HISTORY OF PSYCHOLOGY

---

Vasserman, L. I., Iovlev, B. V., & Chervinskaya, K. R. (2010). Computer psychodiagnostics in the theory and practice of medical psychology: Stages and development prospects. *Sibirskii psikhologicheskii zhurnal (Siberian Journal of Psychology)*, 35, 20–24. (in Russ.).

Vecchione, M., Alessandri, G., & Barbaranelli, C. (2012). Paper-and-pencil and web-based testing: The measurement invariance of the Big Five Personality Tests in applied settings. *Assessment*, 19(2), 243–246. doi: [10.1177/1073191111419091](https://doi.org/10.1177/1073191111419091)

Žitný, P., Halama, P., Jelínek, M., & Květon, P. (2012). Validity of cognitive ability tests – comparison of computerized adaptive testing with paper and pencil and computer-based forms of administrations. *Studia Psychologica*, 54(3), 181–194.

**No conflict of interest**