

PEHPP 2019**Pedagogical Education: History, Present Time, Perspectives****THE SPECIFICS OF COMPUTER-BASED PSYCHODIAGNOSTIC
SITUATION IN EDUCATION**

D. S. Gnedykh (a)*, A. M. Krasilnikov (b), A. D. Shchur (c)

*Corresponding author

(a) Saint-Petersburg University, St. Petersburg, Russia, d.gnedyh@spbu.ru

(b) St. Petersburg gymnasium «Alma Mater», St. Petersburg, Russia, Saint-Petersburg University, St. Petersburg, Russia, a.krasilnikov@spbu.ru

(c) St. Petersburg gymnasium «Alma Mater», St. Petersburg, Russia, njoki@yandex.ru

Abstract

The article is dedicated to the problem of influence of test administration mode on the results of assessment in psychodiagnostics. Sometimes it is difficult to develop identical procedures of paper-and-pencil and computer-based test performance, especially for cognitive abilities assessment. Assessment results can be sensitive to the psychodiagnostic situation and procedure. Nevertheless, under the influence of Information and Communication Technologies (ICT) development in psychodiagnostics and education, the creation of computer-based tests is becoming more popular. On the example of computer-based test on concentration attention, the importance of taking into account the specifics of computer-based psychodiagnostic situation in education is discussed. For tasks with a time limit, it was demonstrated that look and feel of the «computer» test can be a factor deflecting subjects' attention away from the task, and the time they spend on the program interface exploring should not be considered in results analysis. The necessity to observe subjects' behaviour in tasks performance in equivalence studies of paper-and-pencil and computer-based test forms is emphasized. Moreover, the results are demonstrated the need of creation of the normative database scores for «computer» test form.

2357-1330 © 2020 Published by European Publisher.

Keywords: Digital cognitive testing, computer-based psychodiagnostics, psychological support of the educational process.

1. Introduction

Modern educational process implies the use of computer technologies in classes or in distance learning. Over the last decades, ICT has also actively used in psychological support of the educational process, especially for psychological assessment and data analysis (Hansgen & Perrez, 2001; Kveton & Klimusova, 2002; Matarazzo, 1983; Moore & Miller, 2018; Vecchione, Alessandri, & Barbaranelli, 2012). Despite the available experience in computer-based psychodiagnostics to date, there are certain problems related to the impact of different factors on assessment results. These problems are specific for computer-assisted administration of psychodiagnostic techniques versus paper-and-pencil one. For instance, whether a task (or set of questions) fit entirely on a computer screen or scrolling is required; the possibility to return to the previous task; the presence of a time counter on a screen; stimuli and background colors, etc. (Bailey, Neigel, Dhanani, & Sims, 2018; Feenstra, Vermeulen, Murre, & Schagen, 2017; Květon, Jelínek, Vobořil, & Klimusová, 2007). The review of characteristics influencing on computer-based assessment outcomes allowed Webster and Compeau to propose the list of «Sources of format differences between computer and paper modes of administration» (Webster & Compeau, 1996, p. 575). The difference in format of computer and paper modes of tests administration is a cause of conflicting data in research of their equivalence.

Thus, an important direction in research of psychological assessment process is the identification of the equivalence between paper-and-pencil and computer-based tests (PPT and CBT) taking into consideration capabilities and limitations of psychodiagnostic situation using computer. The evidence of equivalence includes: comparable reliabilities, comparable means and standard deviations, correlations with each other and with other tests and external criteria, the same rankings of individual examinees (American Psychological Association Committee on Professional Standards and Committee on Psychological Tests and Assessment, 1986; The International Test Commission, 2006). To reach these options, minimizing test format differences in the modes of administration is needed (Webster & Compeau, 1996). It is easier to create identical «computer» and «paper» formats for a questionnaire than for psychodiagnostic techniques revealing cognitive abilities. Tasks variability and time limitation factor sometimes require modification of procedure according to the specifics of working on the computer. In this case, the observation over subject's behavior while performing the assessment tasks on the computer is extremely important when studying equivalence of computer-based and paper-and-pencil forms of test. This allows to reveal the factors which complicate the use of computer-based test or significantly affect its results and interpretation.

The article presents the results of comparison of computer-based and paper-and-pencil psychological test forms verifying the effect of specific conditions of computer-based psychodiagnostic situation on the assessment results.

2. Problem Statement

Assessment results can be sensitive to the psychodiagnostic situation and procedure. Such factors as subjects' motivation, placement of time counter on the screen, presence or absence of psychologist/teacher, etc. have to be considered in computer-based psychodiagnostics. These aspects could affect not only the results of tests but also the conclusions about equivalence of computer-based and paper-and-pencil test

forms. The procedure of computer-based test development has to be organized so that it would be possible to observe these factors, control and, as a consequence, correct them. The problem is that not for all psychodiagnostic techniques the procedures of paper-and-pencil form performance could be the similar to computer-based one. For instance, in paper-and-pencil tests time limitation is controlled by a psychologist and a subject does not watch the time. The performance of tests on the computer does not imply psychologist's presence, and a feedback is provided by a time counter on the computer screen. If a subject has a high level of anxiety, countdown placed on a screen becomes a stressor which could cause erroneous results. In this case, is it correct to claim that the forms and psychodiagnostics procedure are equivalence? Thus, this kind of computer-based psychodiagnostic techniques requires serious validation procedure.

3. Research Questions

The research questions are the following:

- How to consider specific characteristics of computer-based psychodiagnostic situation while identifying the equivalence between computer-based and paper-and-pencil test forms?
- What should a researcher pay attention to when developing a computer version of a PPT?

4. Purpose of the Study

The purpose of the study is to reveal the influence of specific conditions of computer-based psychodiagnostic situation on the results of equivalence between computer-based and paper-and-pencil test forms on the example of Toulouse-Pieron Test.

5. Research Methods

5.1. Participants

The study involved 70 pupils of Russian schools (44 of them were female students and 26 male students) at the age of 10-12, the mean age being 11.2 years (SD=0.4). The study was conducted in the spring semester. In the autumn semester, pupils had classes on computer skills thereby they all have an experience in working on the computer.

5.2. Research Methods and Instruments

To reveal the specifics of computer-based psychodiagnostics of cognitive abilities, the Toulouse-Pieron Test of attention was chosen. The instrument allows to assess perception and concentration attention (Toulouse & Pieron, 1986). According to the instruction, a subject should look through the list of very similar graphic symbols, identify and cross out those, which are the same as shown in the instruction, and underline others. The time for task performance is limited (10 minutes, the start and the end of every minute have to be recorded in order to calculate the number of correct answers for each minute). The stimuli included 600 symbols.

5.3. Research Procedure

To provide the instruction clarity of computer-based Toulouse-Pieron Test version, its testing was conducted on the first stage of the study. It was important due to the one of specifics of computer-based psychodiagnostic situation: if the assessment is distance, there is no psychologist who can give the additional explanations for the instruction. 10 pupils read the instruction on a computer screen and estimated its clarity answering the following questions: «Evaluate how difficult the task will be» (7-point scale where 1 is very easy and 7 is very difficult); «Evaluate whether the instruction was clear» (7-point scale where 1 is totally unclear and 7 is totally clear); «What would you like to change in the instructions?» (open-ended question). Two questions were included to assess the usability of the program interface: «Did you feel comfortable while doing the task?» (7-point scale where 1 is totally uncomfortable and 7 is totally comfortable); «In what form would it be more comfortable to perform the task?» (open-ended question). The instruction and the program interface were corrected according to the results of the survey. The text instruction was difficult to understand, thus the animated demonstration of actions on the example of certain symbols was added. The feedback when the wrong action with a symbol was chosen by a subject (e.g., if a symbol has to be underlined, but a subject crosses it out) was also added. It was organized by appearance of the window with the text "Mistake!", after its disappearance a subject could check the sample and choose the correct action. On the second stage of the study, pupils performed both computer-based and paper-and-pencil test forms, the time length between assessments was two weeks. To identify the influence of the test performance sequence on the results, the sample was divided into two groups (Table 01).

Table 01. Research Design

Group	N	Test form for the first assessment	Test form for the second assessment
1st group	20	Paper-and-pencil form	Computer-based form
2nd group	50	Computer-based form	Paper-and-pencil form

The instruction was accompanied by the animated demonstration of how to perform the task on a computer. The choice of symbol to be crossed out or underlined provided by computer mouse. The subject should hover a cursor over the symbol and pressed left mouse button (LMB) if he/she decided to cross out it or hovered a cursor over the lower part of the symbol and pressed LMB if he decided to underline it. When hovering a cursor over a symbol, a line, crossing out or underlining it, appeared (Figure 01). The test forms included training task (with 60 symbols) which allowed subjects to understand the instruction better before doing the main task (Figure 02). Subjects could not start the main task until they finished the training one.

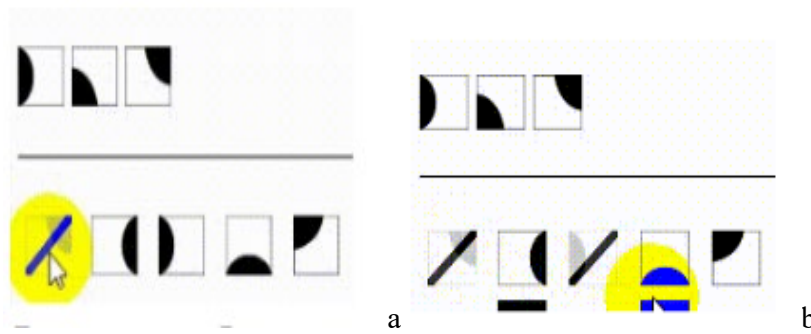


Figure 01. The example of stimuli in the animated instruction for computer-based Toulouse-Pieron Test

Note: three symbols over the line is the sample which are needed to be crossed out; with the symbols under the line it is shown how they could be crossed out or underlined with the use of computer mouse. The circle (yellow-colored) appeared only during the animated demonstration in the instruction to draw subject's attention to the action.

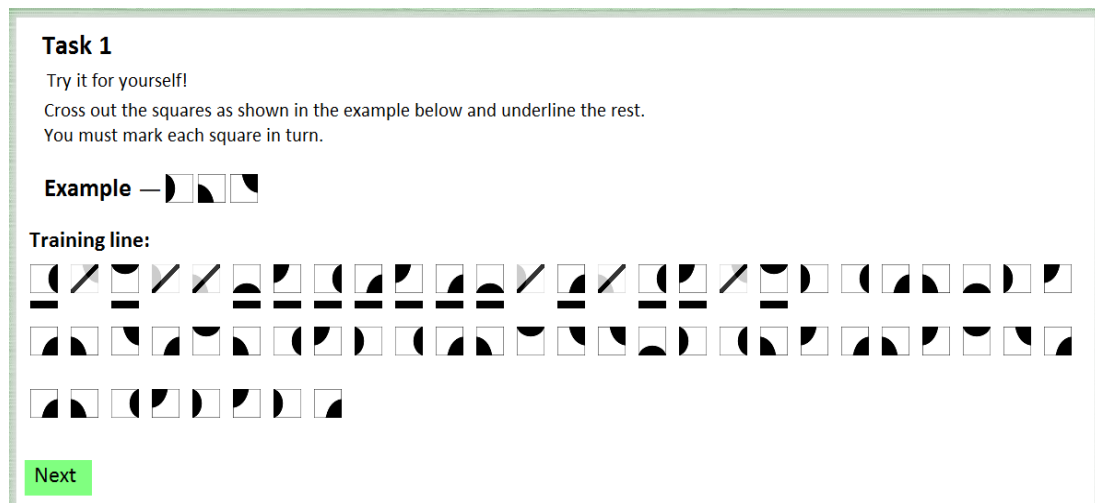


Figure 02. Training task performance for the computer-based Toulouse-Pieron Test

The observation over the pupils performing PPT and CBT was organized to compare pupils' behavior. Different subjects' reactions on the task in traditional and computer-based psychodiagnostic situations allowed to reveal specific characteristics of these situations and test hypotheses of their impact on the results. In the case of paper-and-pencil task, the feedback about time remaining was given by psychologist. In the situation of computer-based psychodiagnostics psychologists did not communicate with subjects. The computer program recorded automatically the end of every minute without giving a feedback to subjects.

5.4. Mathematical and statistical analysis

The concentration attention index and the speed of test performance were analyzed and compared between paper-and-pencil and computer-based forms of Toulouse-Pieron Test. Mathematical and statistical methods were the following: comparability of means (Mann-Whitney U, Wilcoxon signed-rank test),

standard deviations and dispersion (F-test), correlation analysis (r-Spearman), comparability of reliabilities (Cronbach's alpha). Data management and analysis were performed using SPSS 20.0.

6. Findings

To identify whether the results of Toulouse-Pieron Test are different depending on the test performance sequence, the data of two groups was compared. Significant differences for concentration attention index and speed of test performance were revealed (Table 02).

Table 02. The comparison of the Toulouse-Pieron Test results between paper-first and computer-first groups (Mann-Whitney U)

Characteristics		N	Mean rank	Rank sum	Z	p-value
Speed of paper-and-pencil test performance	1st group	20	26.10	522.00	-2.444	0.015
	2nd group	50	39.26	1963.00		
Concentration attention index of paper-and-pencil test	1st group	20	24.10	482.00	-2.964	0.003
	2nd group	50	40.06	2003.00		
Speed of computer-based test performance	1st group	20	46.68	933.50	-2.906	0.004
	2nd group	50	31.03	1551.50		
Concentration attention index of computer-based test	1st group	20	26.40	528.00	-2.389	0.017
	2nd group	50	39.14	1957.00		

The observation over the pupils' behavior during performance of different test forms showed that in the situation of computer-based psychodiagnostics subjects spent the first few seconds scrolling down the information (list of symbols) on the computer screen to identify the task scope. The countdown was conducted from the beginning of the stimuli presentation on the screen, however, in fact, the time they spent was not directly related to the task performance. Due to the importance of data of test performance speed in the analysis of results, time not related to the task could influence on concentration attention index. Thus, the obtained time index for the computer test was not coincide with the actual task performance time (since subjects proceeded to look through the first line of symbols later than the countdown started). In this regard, the time index of the first line of symbols (included 26 symbols) was not considered in the analysis of the «computer» test indexes and the data were counted again (Table 03). All amount of symbols was left in the analysis in the «paper» test.

Table 03. The comparison of groups of pupils performed Toulouse-Pieron Test after the time index of the first line of symbols was excluded from the analysis (Mann-Whitney U)

Characteristics		N	Mean rank	Rank sum	Z	p-value
Speed of paper-and-pencil test performance	1st group	20	26.10	522.00	-2.444	0.015
	2nd group	50	39.26	1963.00		
Concentration attention index of paper-and-pencil test	1st group	20	24.10	482.00	-2.964	0.003
	2nd group	50	40.06	2003.00		
Speed of computer-based test performance	1st group	20	45.78	915.50	-2.671	0.008
	2nd group	50	31.39	1569.50		
Concentration attention index of computer-based test	1st group	20	28.65	573.00	-1.781	0.075
	2nd group	50	38.24	1912.00		

After the exclusion of time index of the first line from the analysis, there were not significant differences for concentration attention index of «computer» test between two groups of pupils ($p \leq 0.075$). Before the exclusion it was significantly higher in the computer-first group in comparison with the paper-first one.

Thus, the time which subjects spent on scrolling down the list of symbols on the computer screen could be one of the factors influencing on the differences in results of computer-based and paper-and-pencil test versions.

Descriptive statistics and the results of analysis for dispersions equivalence are presented in Tables 4 and 5. Of the two formats of the test, mean scores for the index «Speed of test performance» were higher in paper-and-pencil test than in computer-based one (Table 04). However, after exclusion of time index of the first symbols line from the analysis, mean scores of «computer» test became a little bit nearer to mean scores of «paper» version. Whereas means are still significantly differ (see results of Wilcoxon signed-rank test), correcting data for analysis on the base of observation over subject's behaviour could be the way to clarify the results of equivalence between computer-based and paper-and-pencil test forms. Standard deviation and dispersion in «paper» form were higher than in computer-based one (except the results of the first group). Dispersions of «paper» and «computer» test forms were not different for the whole sample. In contrast, for the concentration attention index mean scores were higher in computer-based form in comparison with «paper» one (Table 05). Standard deviation in PPT was slightly different from CBT; dispersion in «computer» form was higher than in «paper» one (except the results of the first group). Significant differences between dispersions of «paper» and «computer» test forms for the whole sample were revealed.

Table 04. Descriptive statistics and the results of analysis for dispersion equivalence (F-test) for the index «Speed of test performance» before and after exclusion of time index of the first symbols line from the analysis

Groups	Paper-and-pencil form			Computer-based form			Dispersion equivalence
	Means	Standard deviation	Dispersion	Means	Standard deviation	Dispersion	p-value for F-test
1st group	39.42	7.9	63.9	36.93(a) 37.4(b)	9.9(a) 10.1(b)	99.4(a) 103.6(b)	0.344(a) 0.301(b)
2nd group	45.98	11.1	124.4	30.28(a) 31.04(b)	7(a) 7.2(b)	49.7(a) 52.6(b)	0.001(a) 0.003(b)
The whole sample	44.10	10.7	114.8	32.18(a) 32.86(b)	8.4(a) 8.6(b)	71.8(a) 74.3(b)	0.053(a) 0.072(b)

Note: for F-test the difference of dispersions is significant if $p \leq 0.05$; «a» means before the exclusion; «b» means after the exclusion; for PPT the time index of the first symbols line was not excluded from the analysis.

Table 05. Descriptive statistics and the results of analysis for dispersion equivalence (F-test) for the concentration attention index before and after exclusion of time index of the first symbols line from the analysis

Groups	Paper-and-pencil form			Computer-based form			Dispersion equivalence
	Means	Standard deviation	Dispersion	Means	Standard deviation	Dispersion	p-value for F-test
1 st group	0.92	0.05	0.002	0.94(a) 0.94(b)	0.04(a) 0.04(b)	0.001(a) 0.001(b)	0.353(a) 0.245(b)
2 nd group	0.95	0.03	0.001	0.96(a) 0.96(b)	0.02(a) 0.01(b)	0.0004(a) 0.0003(b)	0.0002(a) 0.00007(b)
The whole sample	0.94	0.04	0.001	0.95(a) 0.96(b)	0.02(a) 0.02(b)	0.0008(a) 0.0007(b)	0.002(a) 0.0006(b)

Note: for F-test the difference of dispersions is significant if $p \leq 0.05$; «a» means before the exclusion; «b» means after the exclusion; for PPT the time index of the first symbols line was not excluded from the analysis.

The means of the measures were compared across the modes of administration using Wilcoxon signed-rank test. Speed of test performance and the concentration attention in paper-and-pencil form significantly differed from the results of computer-based one (Table 06). However, after exclusion of time index of the first symbols line from the analysis, mean scores of «Speed of test performance» index in the paper-first group were not significantly different – the speed of «computer» test performance was the same as the speed of «paper» test performance.

Table 06. Comparison of mean scores of the Toulouse-Pieron Test results before and after exclusion of time index of the first symbols line from the analysis (Wilcoxon signed-rank test)

Z and p-values	1st group		2nd group		The whole sample	
	before	after	before	after	before	after
Speed of test performance						
Z (Wilcoxon)	-3.901	-1.344	-8.625	-6.115	-10.172	-6.633
p-value	0.000	0.179	0.000	0.000	0.000	0.000
Concentration attention index						
Z (Wilcoxon)	-6.879	-2.277	-3.266	-2.534	-7.551	-3.520
p-value	0.000	0.023	0.001	0.011	0.000	0.000

The correlation analysis showed that results of PPT and CBT positive correlated for the whole sample before/after exclusion of time index of the first symbols line from the analysis:

- «Speed of test performance» index: $r=0.501/0.520$, $p \leq 0.000$;
- Concentration attention index: $r=0.463/0.435$, $p \leq 0.000$.

The comparability of reliabilities of two test forms revealed the high internal consistency of each form before and after exclusion of time index of the first symbols line from the analysis (for PPT, $\alpha=0.925$ and for CBT, $\alpha=0.924$ (before)/0.920 (after)).

7. Conclusion

Due to the significant differences in means, dispersions and standard deviations, paper-and-pencil and computer-based forms of the Toulouse-Pieron Test can be considered not quite equivalent. The «computer» test was performed slower and more accurate than «paper» one. Consequently, the results of PPT

and CBT comparison highlight the need of further creation of the normative database scores for «computer» test form. However, consideration of specifics of computer-based psychodiagnostic situation in the analysis of results clarifies data on the equivalence, and sometimes can even improve it (as was shown after the exclusion of time index of the first symbols line from the analysis). In the study the characteristics of subject's behavior in the process of interaction with the computer-based form of time limited test were indicators to reveal such kind of specifics.

Thus, a strict control of conditions of computer-based testing on the stage of its development is needed. The observation over the subjects' performance of task on the computer, their interaction with the program interface and individual activity is required. It allows to identify certain behaviour patterns, compare them with those which are observed when subjects perform PPT and understand whether it is correct to compare paper-and-pencil and computer-based test forms as equivalent. Not only the technical aspects of the task performance can be the observation object, but also subjects' affective reactions to computer testing (e.g., computer anxiety) (Smith & Caputi, 2007), testing motivation (Chua, 2012) or testing mode preference (Ebrahimi, Toroujeni, & Shahbazi, 2019). That means that equivalence study of «paper» and «computer» test forms should be conducted in the presence of a psychologist or an observer, and not at a distance.

Another important point is that researcher should not make premature conclusions about the lack of equivalence of two forms of psychological techniques. The situation of computer-based test performance should be analyzed carefully to reveal the factors that could affect the difference in results and consider them in further analysis.

References

- American Psychological Association Committee on Professional Standards and Committee on Psychological Tests and Assessment. (1986). *Guidelines for computer-based tests and interpretations*. Washington, DC: Author.
- Bailey, S. K. T., Neigel, A. R., Dhanani, L. Y., & Sims, V. K. (2018). Establishing measurement equivalence across computer- and paper-based tests of spatial cognition. *Hum Factors*, 60(3), 340-350.
- Chua, Y. P. (2012). Effects of computer-based testing on test performance and testing motivation. *Computers in Human Behavior*, 5(28), 1580-1586.
- Ebrahimi, M. R., Toroujeni, S. M. H., & Shahbazi, V. (2019). Score equivalence, gender difference, and testing mode preference in a comparative study between computer-based testing and paper-based testing. *International Journal of Emerging Technologies in Learning*, 14(07), 128-143.
- Feenstra, H. E. M., Vermeulen, I. E., Murre, J. M. J., & Schagen, S. B. (2017). Online cognition: factors facilitating reliable online neuropsychological test results. *The Clinical Neuropsychologist*, 31(1), 59-84.
- Hansgen, K. D., & Perrez, M. (2001). Computer-based psychodiagnostics in family and education - Concepts and perspectives. *Psychologie in erziehung und unterricht*, 48(3), 161-178.
- 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.
- Kveton, P., & Klimusova, H. (2002). Methodological aspects of computer administration of psychodiagnostics methods. *Ceskoslovenska psychologie*, 46(3), 251-264.
- Moore, A. L., & Miller, T. M. (2018). Reliability and validity of the revised Gibson Test of Cognitive Skills, a computer-based test battery for assessing cognition across the lifespan. *Psychology research and behavior management*, 11, 25-35.
- Matarazzo, J. D. (1983). Computerized psychological testing. *Science*, 221, 323.

- Smith, B., & Caputi, P. (2007). Cognitive interference model of computer anxiety: Implications for computer-based assessment. *Computers in Human Behavior*, 23(3), 1481-1498.
- The International Test Commission (2006). International guidelines on computer-based and internet-delivered testing. *International Journal of Testing*, 6(2), 143-171. https://doi.org/10.1207/s15327574ijt0602_4
- Toulouse, E., & Pieron, H. (1986). Toulouse-Pieron. Pruebaperceptiva y deatencion [Test of perception and attention]. Madrid: TEA.
- 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.
- Webster, O., & Compeau, D. (1996). Computer-assisted versus paper-and-pencil administration of questionnaires. *Behav Research Methods, Instruments, and Computers*, 28, 567-576.