Title:

Report on on-line trial test for fourth-grade students – May 2009

Authors:

Srdjan Verbić, Boris Tomić, Vesna Kartal

Date:

11/22/2010

Abstract:

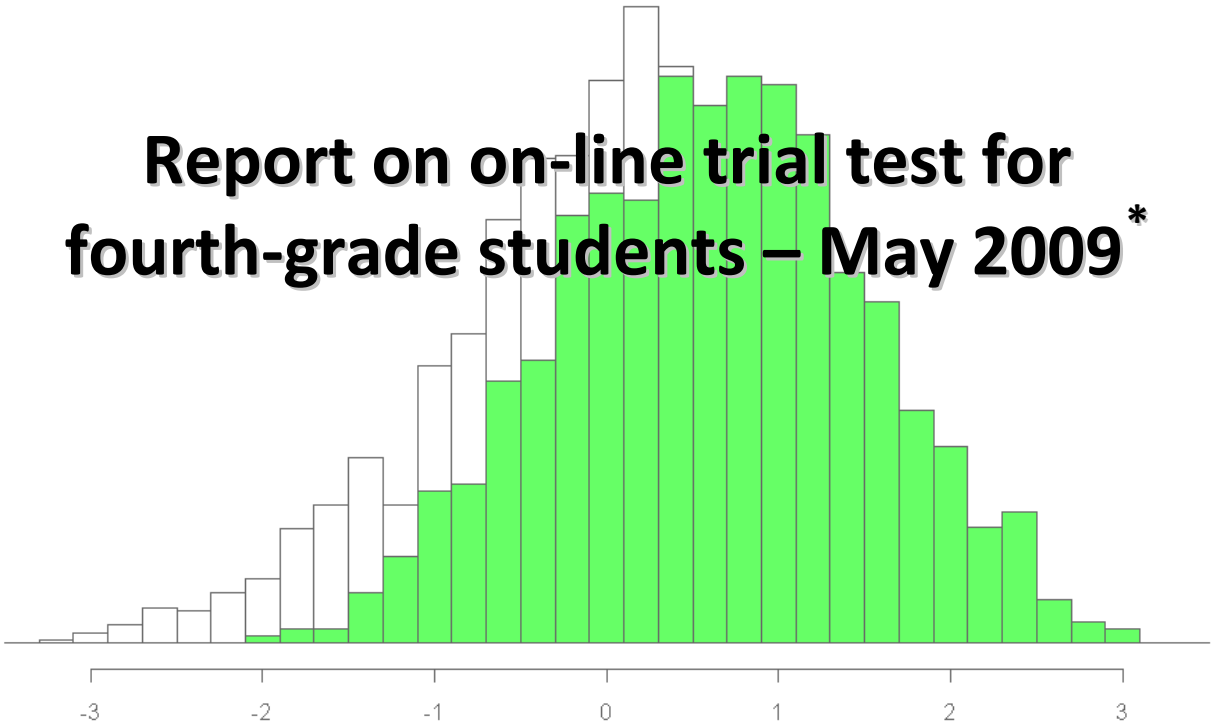
On-line trial testing for fourth-grade students was an exploratory study realized as a part of the project "Developing annual test of students' achievement in Nature & Society" realized by Institute for Education Quality and Evaluation. Main ideas of the study were to explore possibilities for on-line testing at national level in Serbia, and to explore the trial test characteristics on a convenience sample in partially controlled testing conditions.

This report presents evidences concerning possibilities for on-line testing as well as detailed analysis and comparison of results obtained from the annual and the trial test. In total, 926 students from 50 elementary schools participated in the trial test. Results of the trial and the annual tests were analyzed using Classical Test Theory and Item Response Theory. We have separately analyzed all alternatives and options in multiple choice and multiple response questions (including non-answers). We have also analyzed differential item functioning and item response time in order to collect additional information about questions and the test as a whole.

Keywords: trial test, on-line testing, IRT, analysis of alternatives, DIF, item response time

Report contains 40 figures, 72 tables, and 9 references.

Report on on-line trial test for fourth-grade students – May 2009*



Srdjan Verbić, Boris Tomić, Vesna Kartal
Institute for Education Quality and Evaluation

Belgrade, 2009

* Translated and adapted version of the original report in Serbian available at http://ceo.edu.rs.

Content

Introduction

On-line trial testing for fourth-grade students is an exploratory study realized as a part of the project "Developing annual test of students' achievement in Nature & Society" realized by Institute for Education Quality and Evaluation [1]. Basic idea of the study was to enable students to participate in a large-scale testing from their own schools through the Internet. All schools where teachers were interested in the testing were eligible for participation in the study. Motivation for this mode of test delivery was the fact that the schools in Serbia have much better Internet connections today than they had only a year ago.

Computer-based testing (or e-assessment) enables realization of trial test on convenience sample with simple administration and test analysis. It makes such studies much cheaper and easier to organize. Using Internet for computer-based testing, i.e. on-line testing, enables studies where communication with students and teachers who facilitate testing goes via Internet.

Basic aims of this study were:

1. to explore possibilities for on-line testing at national level in Serbia and
2. to explore the trial test characteristics on a convenience sample in partially controlled conditions.

This report presents evidences concerning possibilities for on-line testing as well as detailed analysis and comparison of results obtained from the annual and the trial test.

Methods

The on-line testing was realized in period May 5-19, 2009, immediately after the annual test in Nature & Society (April 2009). For the on-line testing modified Web application Moodle (with Ttest module) [2, 3] was used. Moodle had been used for all computer-based tests performed by the Institute in the past few years.

The sample for the trial test was not specifically designed. All interested schools were enabled to participate. The only request we had was to choose participants among fourth-grade students who attend schools in Serbia. The schools were required to provide us with information about the way selection was done.

Public media and direct e-mail communication were used to announce the on-line testing and to invite schools in Serbia to participate. Fifty six schools responded to our invitation (Appendix 2). Each school had a chance to select students for the test according to their own needs and opportunities. In total, the schools applied 1341 students. Accounts for the testing portal were made for all these students. Files with the Web address of testing portal, usernames, passwords, and the number of test intended for all the students were sent to all participating schools, as well as exact date and time when they should start testing.

Two weeks before the trial testing, a short example-test was available for students at the testing portal so they could practice and become familiar with the testing procedure and the way of responding to the test. This example-test has been taken by 247 students and several curious teachers.

Finally, 926 students from 50 elementary schools participated in the trial test. Elementary school "Sveti Sava" from Batočina had the most participants (81 students), while the least students had school "Djura Jakšić" from village Ravni (one student only). Because of numerous technical obstacles, like sudden interruption of Internet connection, obtained data is not valid for all the students. The trial test data analysis is performed for responses of 903 students.

Conditions

The on-line testing was deliberately performed in partially controlled conditions. Teachers in schools, where the testing was performed, had the opportunity to select students who would do the test. We gave instructions about the way teachers should realize testing, but we did not control the realization.

The trial test was composed of four sections, with eight questions each. Using these sections, we have made four different versions of the same test with the same number of questions – 32. That was the way to help teachers to hinder students to peek into each other's test.

All the students were provided with test instructions. Short instructions had appeared on screen immediately after logging in the e-assessment portal. After the reading of short instructions, students listened to detailed instructions red by teacher. Students were not allowed to start their tests until teacher had finished the reading and finally gave them password to start the test.

Computer classrooms in Serbian schools are equipped with computers with different keyboards and different language settings. Such a situation can cause problems with logging which may upset students at the very beginning of the testing [4]. Therefore, we decided to use numerical usernames and passwords. Once they start the test, there is no need for them to use keyboard any more. Students responded to all questions by using mouse, i.e. by selecting some alternative or option given on the screen. When students finished the test, they were able to see their total score.

At the end, all students were asked to fulfill a short questionnaire (Appendix 4).

Test

All test questions were taken from questions pool used for the annual test in Nature & Society. The trial test had 32 questions (eight from each of four booklets at the annual test). Because of potential problems with a keyboard usage, only multiple choice and multiple response questions were chosen. Distribution of questions according to their difficulty and topics was similar to the distribution for the annual test. For administrative reasons, time limit for both tests was set to 50 minutes although students needed less time to answer all the questions.

Feedback

Computer-based testing, generally, enables automatic test scoring and a lot of feedback. Our trial-test had limited feedback so it would not influence performance and reliability of the study. After the students had submitted the answers, they were able to see their total scores but not which questions they

answered correctly. Teachers involved in testing realization had also a chance to see results of their students. For that purpose, all teachers had accounts with necessary privileges for approach to the system.

Several days after the test was closed, all the questions and results became available to students and teachers. They used the same accounts for viewing the results as they used for taking the test earlier.

Result analysis tools and methods

Results of the trial and the annual tests were analyzed using Classical Test Theory (CTT) and Item Response Theory (IRT). Statistical analysis was done by using statistical software R with package irtoys [5] and standalone software IRT Command Language [6]. Standard errors were determined by bootstrapping method. Estimation of students' ability was done by using *a posteriori* method [7].

We have separately analyzed all alternatives in multiple choice questions (including non-answers). In the case of multiple response questions, we have analyzed responses to all available options. For each alternative/option difficulty index, discrimination index and median response time were determined.

Ttest module enables measuring responding time for each answer in the test. We have analyzed these data also in order to collect additional information about questions and the test as a whole.

Two different scales were used for representing questions' difficulty and students' ability: the annual test scale and the trial test scale. Estimations of students' ability are given on the annual test scale. For the analysis of alternatives we had used the trial test scale.

Additional sources of information

For this study we have used data collected for the annual test in Nature & Society and data collected in the preparation of the trial test for standard setting. Since the annual test did not have questionnaires for students and teachers like the trial test did, we used comparable data from students' database maintained by Institute for Education Quality and Evaluation.

Comparison of results for the annual and the trial test

Students who took the trial test, in average, had significantly higher achievement than students at the annual test. We can estimate ability of all students on both tests and present it on the same scale if we use IRT parameters obtained for the annual tests as benchmark. That way we can see how the convenience sample (used for the trial test) was different from the representative one (used for the annual test).

The distributions of estimated ability for the annual and the trial test are given on Figure 1. We can see that both tests had samples with fairly normal distributions of similar width. These two distributions differ significantly by the mean value.

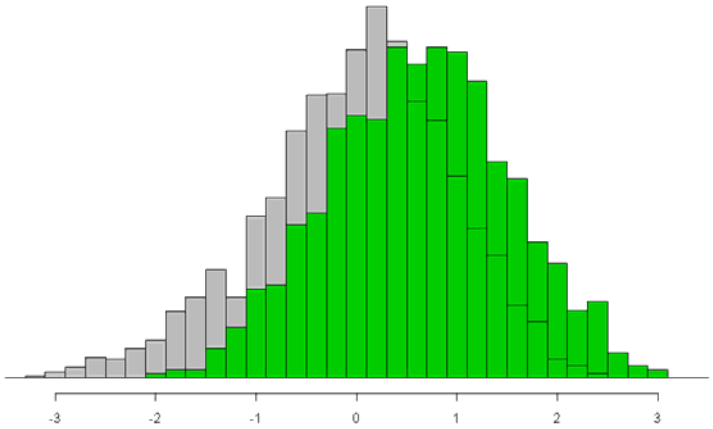


Figure 1: The distributions of estimated ability for the annual test (gray) and the trial test (green)

Ability distribution for the annual test can be described as normal curve with the mean value μ_a= 0.00 and the standard deviation σ_a=0.88. For the trial test, the mean value is μ_t= 0.62 and the standard deviation σ_t=0.80. In other words, the same question's difficulty parameter should be smaller at the trial test than at the annual test. That difference should be approximately the same as the difference between estimated abilities for these two tests.

Comparison of questions' metric characteristics for the annual and the trial test

If we do not have standard questions which we could use for the scale calibration, question parameters can be determined relatively for the sample that we have. That way, we can estimate parameters independently for the annual and the trial test. In two-parameter IRT model, in the ideal case, discrimination (parameter *a*) should have the same value regardless of the sample we use. Difficulty (parameter *b*) for two tests should differ by constant value determined by the properties of used samples [8].

The estimations of parameters *a* and *b*, as well as their standard deviations for the trial and the annual tests, are given in Table 1. We can notice that estimations of discrimination are approximately equal for both tests/samples, just as we expected. The estimation of questions' difficulty, however, vary more than we expected. Differences are particularly big for very hard questions at the annual test (b_a>2) and very easy questions at the trial test (b_t<-2). This deviation from the theoretical expectation can be an evidence of possibility that these two tests are different not only because of different samples but because of other characteristics that we did not measure.

	Trial test		Annual test	
Question	a_t	b_t	a_a	b_a
#1	1.1(0.1)	-1.1(0.1)	1.1(0.1)	-0.1(0.1)
#2	0.8(0.1)	0.8(0.1)	0.4(0.1)	3.4(0.7)
#3	1.7(0.2)	-1.6(0.1)	1.4(0.2)	-1.2(0.1)
#4	0.8(0.1)	1.8(0.2)	0.7(0.1)	1.9(0.3)
#5	1.0(0.1)	-2.0(0.2)	1.0(0.1)	-1.3(0.2)
#6	0.8(0.1)	-0.9(0.1)	0.8(0.1)	-1.3(0.2)
#7	0.8(0.1)	-0.6(0.1)	0.5(0.1)	1.8(0.4)
#8	0.6(0.1)	-0.6(0.2)	0.6(0.1)	0.4(0.2)
#9	0.6(0.1)	0.7(0.2)	0.5(0.1)	2.9(0.6)
#10	0.5(0.1)	-0.8(0.2)	0.5(0.1)	0.1(0.2)
#11	0.7(0.1)	0.4(0.1)	0.8(0.1)	1.8(0.3)
#12	1.0(0.1)	-0.9(0.1)	0.7(0.1)	-0.7(0.2)
#13	1.0(0.1)	-2.0(0.2)	1.1(0.2)	-1.6(0.2)
#14	0.7(0.1)	-1.4(0.2)	0.8(0.1)	0.6(0.2)
#15	0.9(0.1)	-1.5(0.2)	0.7(0.1)	0.1(0.1)
#16	0.7(0.1)	1.7(0.2)	0.4(0.1)	3.7(0.7)
#17	1.0(0.1)	-0.5(0.1)	1.0(0.1)	-0.3(0.1)
#18	1.2(0.1)	-1.1(0.1)	1.5(0.2)	-0.4(0.1)
#19	0.5(0.1)	-0.3(0.2)	0.5(0.1)	0.9(0.3)
#20	0.7(0.1)	1.0(0.2)	0.6(0.1)	1.8(0.3)
#21	0.9(0.1)	-1.0(0.1)	0.8(0.1)	0.6(0.2)
#22	0.8(0.1)	-1.7(0.2)	1.1(0.2)	-1.0(0.2)
#23	1.1(0.1)	-1.1(0.1)	1.5(0.2)	-0.4(0.1)
#24	0.9(0.1)	-0.6(0.1)	1.0(0.1)	0.1(0.1)
#25	0.6(0.1)	1.8(0.3)	0.7(0.1)	2.1(0.4)
#26	0.8(0.1)	-0.5(0.1)	0.7(0.1)	-0.3(0.2)
#27	1.0(0.2)	-3.1(0.5)	0.8(0.2)	-3.3(0.6)
#28	0.9(0.1)	-1.9(0.2)	0.4(0.1)	-0.2(0.3)
#29	0.9(0.1)	-2.5(0.3)	1.4(0.2)	-1.8(0.2)
#30	0.7(0.1)	-2.5(0.4)	0.8(0.1)	-1.2(0.2)
#31	1.3(0.2)	-1.3(0.1)	1.3(0.2)	-0.6(0.1)
#32	0.9(0.1)	-2.7(0.4)	1.0(0.1)	-1.6(0.2)

Table 1: The comparison of IRT parameters for all questions for the annual and the trial test

The comparison of discrimination parameters for the two tests is given on Figure 2. Dashed line displays theoretically expected values for the trial test, i.e. values equal to those at the annual test. The comparison of difficulty parameters for the two tests is given on Figure 3. Dashed line presents values from the annual test decreased by 0.62, i.e. expected value for the trial test.

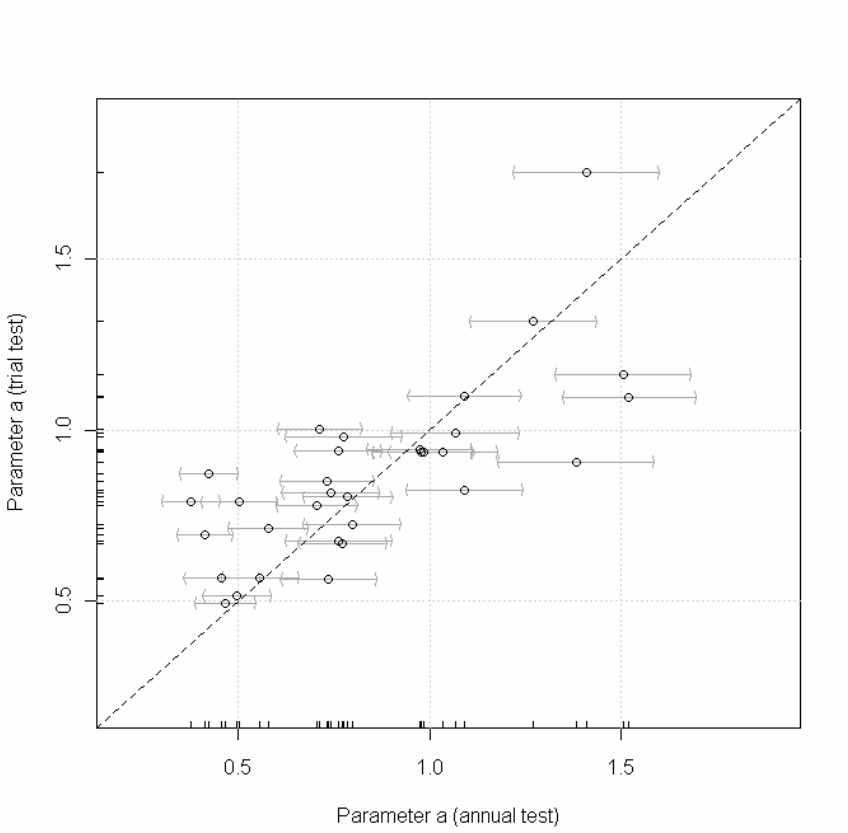


Figure 2: The comparison of questions' discrimination parameter (*a*) estimated for the annual and the trial test

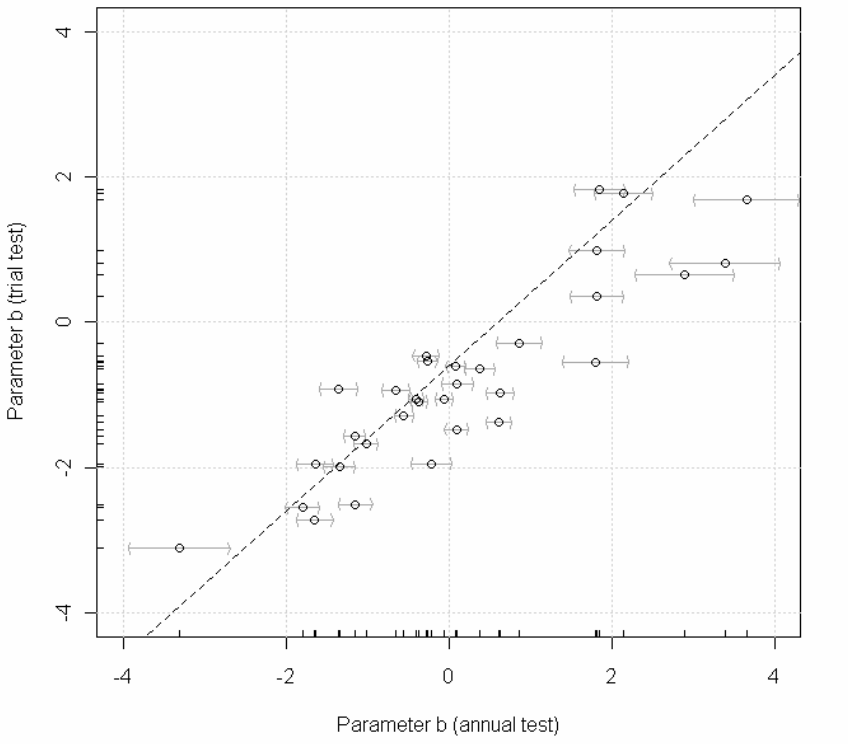


Figure 3: The comparison of questions' difficulty parameter (*b*) estimated for the annual and the trial test

Number of omitted questions for the annual and the trial test

There is a great difference between the number of omitted questions for annual and trial tests. In the first case, there are questions where up to 20% of students did not answer. In the second case, the percentage of omitted questions does not rise above 3%. This difference can be a consequence of the way of examination (mode of test delivery), but also of different way the selection of students for the testing was done.

From the total number of non-answers for the trial test, 44 of them have response time 0 seconds. It means that students have not seen that question at all, so we can treat it as missed instead of omitted one. Since the response time can be measured for computer tests only, we can not measure the response time and know information about the number of missed questions for the annual test.

The ratio of omitted questions for each item in both tests is presented in Table 2 and on Figure 4.

Question	Question codename	Ratio on non-answers (annual test)	Ratio of non-answers (trial test)
#1	PP031031	0.03	0.01
#2	PP013012	0.02	0.02
#3	PP041047	0.01	0.01
#4	PP061069	0.01	0.01
#5	PP071081	0.02	0.01
#6	PK093104	0.02	0.03
#7	PK123146	0.17	0.01
#8	PD171192	0.05	0.01
#9	PP033036	0.01	0.01
#10	PP013010	0.01	0.01
#11	PP043051	0.03	0.01
#12	PP063072	0.03	0.01
#13	PP073082	0.02	0.01
#14	PK111131	0.13	0.01
#15	PK121136	0.13	0.01
#16	PK123144	0.15	0.02
#17	PP043052	0.03	0.02
#18	PP061067	0.02	0.01
#19	PK093107	0.02	0.01
#20	PK101113	0.02	0.01
#21	PK123142	0.20	0.02
#22	PD141159	0.02	0.01
#23	PD133154	0.11	0.01
#24	PP013009	0.03	0.02
#25	PP083095	0.05	0.02
#26	PK101118	0.02	0.02
#27	PK111132	0.01	0.01
#28	PK121139	0.14	0.01
#29	PD143165	0.02	0.01
#30	PD153176	0.03	0.02
#31	PP021015	0.02	0.01
#32	PK093106	0.01	0.01

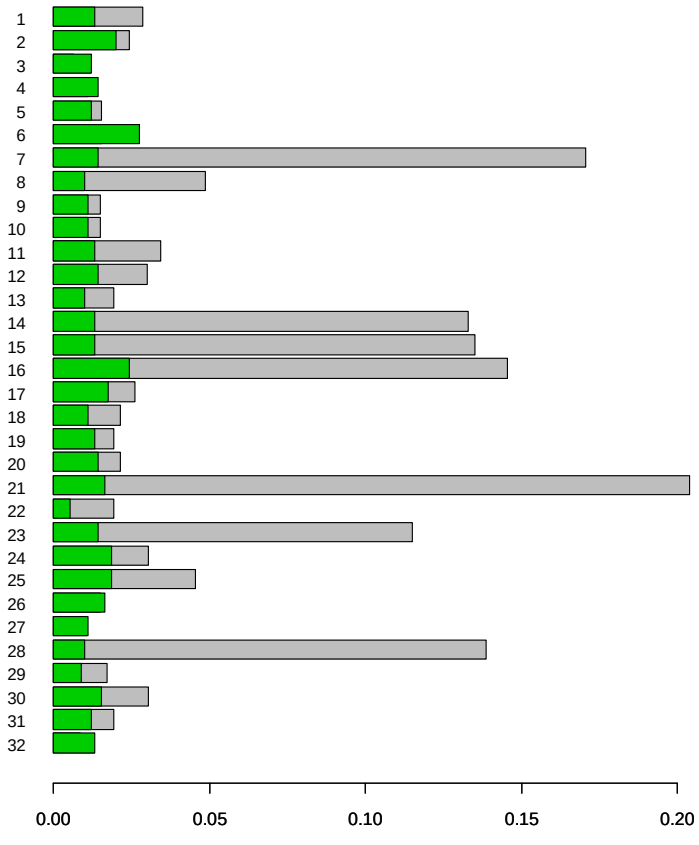


Figure 4: The ratio of non-answers for the annual (gray) and the trial test (green)

DIF by gender

Differential Item Functioning (DIF) for examinees grouped by gender is analyzed using Mantel-Haenszel method [9]. DIF *p*-value is determined separately for the annual and the trial test. All items where DIF *p*-value is less than 0.05 behaves statistically different for boys and girls. This discrepancy does not need to be an indicator of item bias, but it is a reason to check what and how that item measures.

p-Values for the annual and the trial test are presented in Table 3 and on Figure 5. Here we can see that items #18 and #23 are DIF flagged at both representative and convenience sample, while items #10, #17, and #25 are not. For item #23 significance of DIF difference is biggest for both tests (p<0,01).

Question	Question codename	DIF *p*-value (annual test)	DIF *p*-value (trial test)
#1	PP031031	0.98	0.95
#2	PP013012	0.92	0.95
#3	PP041047	0.60	0.77
#4	PP061069	0.39	0.92
#5	PP071081	0.07	0.38
#6	PK093104	0.02	0.23
#7	PK123146	0.53	0.62
#8	PD171192	0.53	0.20
#9	PP033036	0.15	0.93
#10	PP013010	0.98	0.04
#11	PP043051	0.37	0.77
#12	PP063072	0.24	0.80
#13	PP073082	0.38	0.87
#14	PK111131	0.26	0.43
#15	PK121136	0.11	0.31
#16	PK123144	0.97	0.51
#17	PP043052	0.31	0.04
#18	PP061067	0.04	0.05
#19	PK093107	0.95	0.36
#20	PK101113	0.27	0.28
#21	PK123142	0.70	0.84
#22	PD141159	0.27	0.88
#23	PD133154	0.00	0.00
#24	PP013009	0.92	0.13
#25	PP083095	0.46	0.05
#26	PK101118	0.51	0.55
#27	PK111132	0.70	0.56
#28	PK121139	0.39	0.97
#29	PD143165	-	0.92
#30	PD153176	0.77	0.89
#31	PP021015	0.44	0.97
#32	PK093106	0.81	0.61

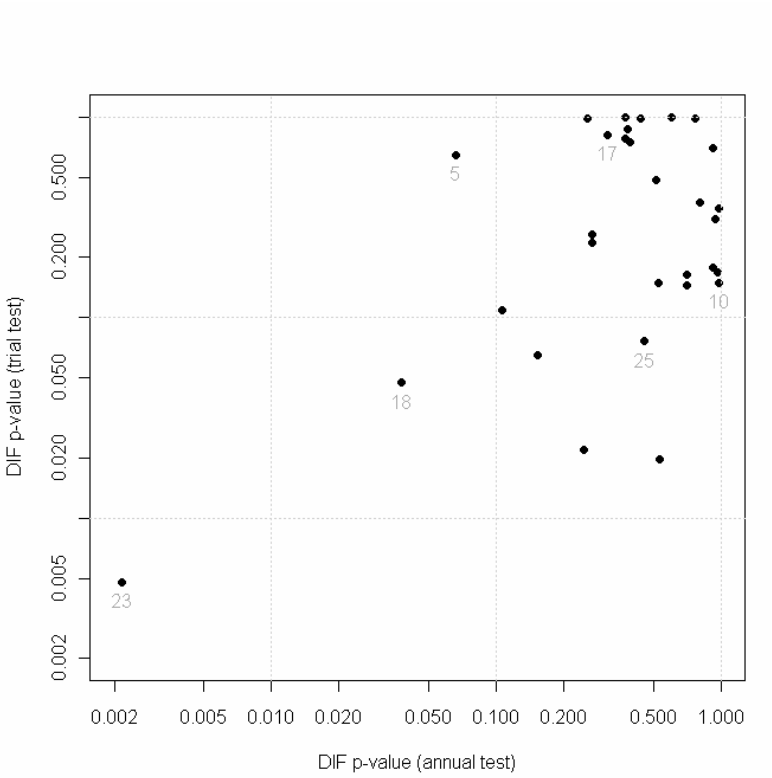


Figure 5: DIF *p*-values for the annual and the trial test

Conclusions

On-line test in Nature & Society is realized using computer-based testing system where students participated from their own schools through the Internet. All elementary schools in Serbia had opportunity to participate in this testing if the school had an interest and technical capabilities. The organization of on-line testing, as an innovation in educational practice, was accompanied by many technical problems. Because of the quality and reliability of Internet connections, such a testing was impossible only two years ago. Today, we can say that on-line testing is feasible in the majority of schools in Serbia. The quality of Internet connections in schools is still the biggest obstacle in the realization of e-testing. Due to technical difficulties, in 6 out of 56 engaged schools, testing was not finished successfully. The second biggest obstacle for e-testing was the quality of Institute's Internet connection where the e-testing server is located.

We have received the feedback concerning testing realization from 25 teachers. 14 Teachers said there were no technical problems of any kind, 3 of them said that some pages were loading slowly, 5 of them said that Internet connection was interrupted but it did not violate the testing, while 3 teachers said they encountered too many technical problems so they can not say the testing was successful. 22 Teachers said that students were very motivated to answer all the questions as good as they could. 19 Teachers said that praxis of e-testing in Serbia would be very useful for all students while the others thought that just certain students would benefit from such a praxis. Generally, teachers were very interested in this kind of testing and they hoped such testing would be available more often in the future.

Psychometric characteristics of the trial test were examined through the results of both the annual and the trial test. The characteristics of individual items in the trial test were compared with the same item in the annual test. Results of both tests were analyzed using Classical Testing Theory (CTT) and 2-parameter Item Response Theory (2PL IRT). The annual test was performed on representative sample (1842 students in total) in paper & pencil mode, while the sample for the trial test was convenience sample (926 students from 50 elementary schools). Because of unavoidable technical difficulties in the e-testing realization, some test taking attempts were not valid. Finally, 903 students had valid responses suitable for the analysis.

Trial test was composed of 32 questions (eight from each of four annual test booklets). All selected items were of multiple-choice type (with one or more correct answers). The distribution of items by topics and difficulties were approximately the same as for the annual test. Responding time for both test was limited to 50 minutes.

According to Item Response Theory, we can reconstruct item parameters for entire population if we know parameters for a sample that is large enough and normally distributed, and standard items for scale calibration. That means that estimation of item discrimination, in ideal case, should have the same value for all normally distributed samples, while item difficulty estimation should be linearly shifted in dependence on the mean ability of used sample. In this study, we have examined how good can we estimate item parameters for the entire

population on the basis of on-line trial test given to convenience sample of students. The assumed similarity of item response curves for these two tests opens another research question – Can we detect differential item functioning for entire population if we look at the results of the trial test only?

Results show that differences in the estimation of item discrimination for the annual and the trial test are small and comparable to their standard errors. For item difficulty we can see that there is linear dependence between estimations obtained for these two tests with correlation coefficient r=0.89. Such a correlation with theoretical model, however, does not enable us to make good prediction of parameters for individual items. Difference in testing conditions certainly have great influence on the estimation of individual item parameters. (We have some indications that conditions in certain schools were not completely regular. That is also a reason why results do not fit in theoretical model as good as we expected.) Results of numerical simulations show that these two tests differ more than can be explained by the difference in samples. Obviously, there is a difference in students' attitude toward the tests. This discrepancy is reflected in the number of non-answered questions for the annual (5.0% per student in average) and the trial test (1.3%). The analysis of differential item functioning shows that two out of three items are flagged correctly (DIF p-value less than 0.05) for the annual test using the results of the trial test only.

In this study we have also analyzed alternatives, i.e. options for multiple choice questions. The analysis of alternatives frequency against estimated ability can help us detect bad alternatives or strong distractors. Information about item response time fulfills this picture because it enables us to detect strong distractors even when difficulty and discrimination parameters of alternatives do not indicate such behavior (e.g. the fourth alternative of question #9).

The general conclusion of this study is that on-line testing is feasible, that teachers need them and that such a practice motivate students. In psychometric domain, on-line trial tests in partially controlled conditions are useful for item and test quality analysis and usability of alternatives in multiple choice questions. The estimation of item difficulty is also possible but under constrains that have to be thoroughly examined in the future.

Appendix 1: Metric characteristics of items

All items from the trial test and their metric characteristics are given in this Appendix. Number and codename of items are displayed in the header of each page. Response codes ①, ②, ③, ④ are associated with alternatives in multiple-choice questions. Code ⑨ is reserved for non-answers. Correct (✓) and incorrect answers (✘) are labeled for all alternatives. If there is a question with a few correct answers, we consider it successfully answered if examinee responded correctly to all options like it was set for the annual test.

CTT parameters are determined for all items (item difficulty – *p*-value and item discrimination – *r*-value). Item difficulty is determined as the ratio of number of correct answers and the total number of answers for particular item. Item discrimination is obtained as item-total correlation with exclusion. 2PL IRT parameters are also determined for all items (difficulty – *b* and discrimination – *a*). For the sake of comparison, all item parameters are determined for both: the annual and the trial test.

For the trial test, all alternatives (or options for multiple response questions) were analyzed separately. The number of responses, *r*-value, *p*-value and characteristic response time are given in tables for all alternatives or options. Characteristic response time is given in seconds as the median value of response time distribution for each alternative or option.

Frequency curves for particular alternatives, or options, are given graphically for the trial test. Here we used non-calibrated scale of ability determined for the trial test. The curves of alternatives or options are colored according to the response code.

Question #1 PP031031

У природна станишта убрајамо:

Одабери један одговор.

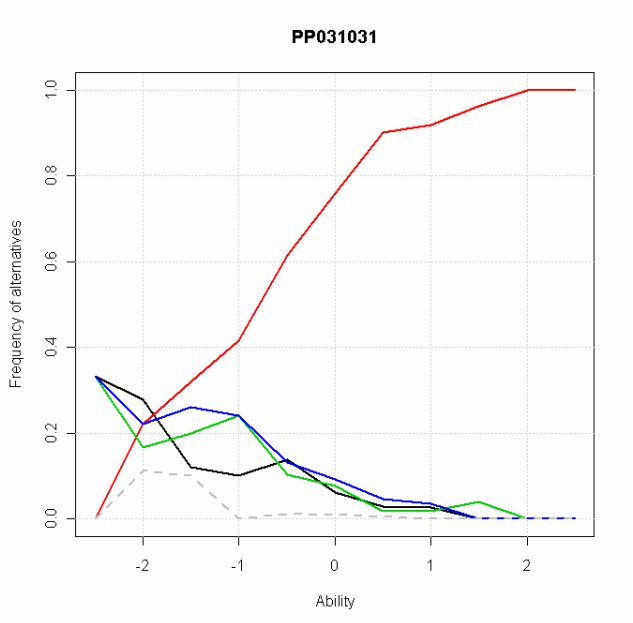
○	реку, њиву и језеро	✘	①
○	шуму, реку и бару	✓	②
○	виноград, фарму и воћњак	✘	③
○	ливаду, њиву и повртњак	✘	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.40	0.51		annual test	1.1	-0.1
trial test	0.35	0.72		trial test	1.1	-1.0

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	65	0.07	-0.14	43
②	639	0.72	0.35	39
③	80	0.09	-0.18	41
④	94	0.11	-0.18	40
⑨	10	0.01	-0.10	76



Question #2 PP013012

Корење биљака спречава:

Одабери један одговор.

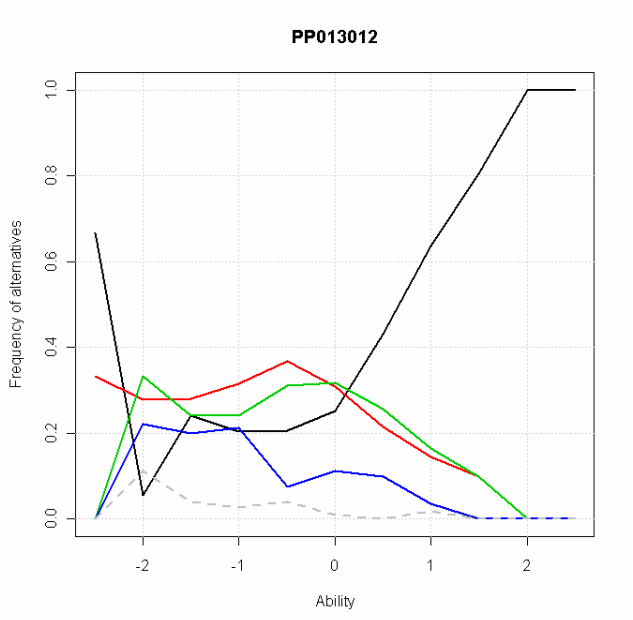
○	спирање земљишта	✓	①
○	загађивање земљишта	✘	②
○	исушивање земљишта	✘	③
○	ђубрење земљишта	✘	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.09	0.23		annual test	0.4	3.4
trial test	0.28	0.37		trial test	0.8	0.8

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	327	0.37	0.28	46
②	230	0.26	-0.13	43
③	221	0.25	-0.05	48
④	93	0.10	-0.15	39
⑨	17	0.02	-0.08	58



Question #3 PP041047

Шта од наведеног **не убрајамо** у природна богатства?

Одабери један одговор.

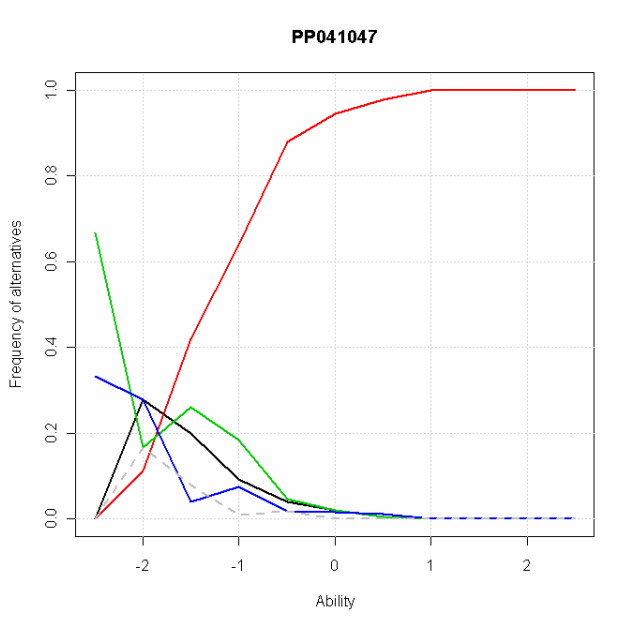
○	биљни и животињски свет	✘	①
○	куће	✓	②
○	руде	✘	③
○	воду	✘	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.44	0.77		annual test	1.4	-1.2
trial test	0.39	0.87		trial test	1.7	-1.6

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	36	0.041	-0.19	40
②	770	0.87	0.39	27
③	51	0.06	-0.26	34
④	23	0.03	-0.14	43
⑨	8	0.01	-0.14	37



Question #4 PP061069

Која својства има ваздух?

Одабери бар један одговор.

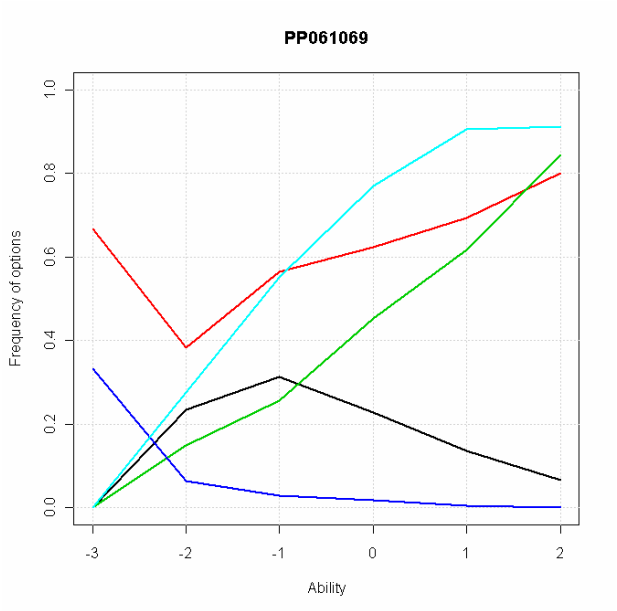
☐	има сталан облик	✘	❶
☐	креће се	✓	❷
☐	заузима простор	✓	❸
☐	има боју	✘	❹
☐	у гасовитом је стању	✓	❺

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.25	0.23		annual test	0.7	1.9
trial test	0.27	0.22		trial test	0.8	1.8

Analysis of options

Option	Number of responses	*p*-value of option	*r*-value of option	Median response time [s]
❶	191	0.22	-0.14	41
❷	555	0.63	0.10	44
❸	403	0.46	0.25	39
❹	17	0.02	-0.11	35
❺	657	0.74	0.32	41



Question #5 PP071081

Шта ће се десити ако у посуду са водом ставимо две лоптице: једну од **стиропора**, а другу од **стакла**?

Одабери један одговор.

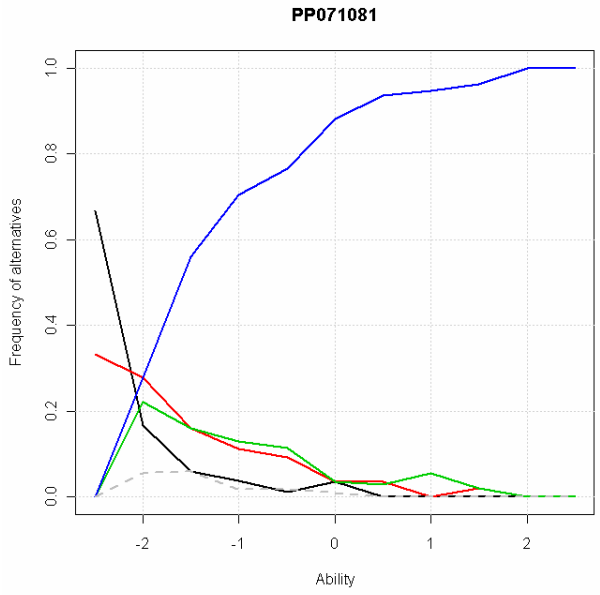
○	Обе ће потонути.	✘ ①
○	Обе ће плутати.	✘ ②
○	Лоптица од стиропора ће потонути, а стаклена ће плутати.	✘ ③
○	Лоптица од стиропора ће плутати, а стаклена ће потонути.	✓ ④
	(no response)	✘ ⑨

Item parameters

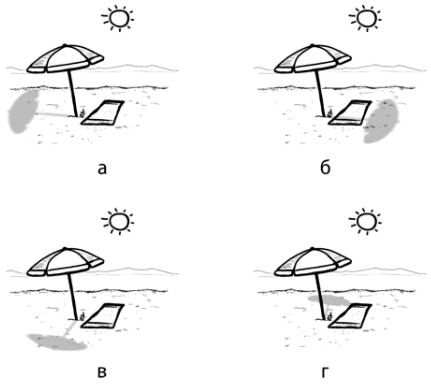
CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.36	0.75		annual test	1.0	-1.3
trial test	0.28	0.83		trial test	1.0	-2.0

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	21	0.02	-0.15	42
②	54	0.06	-0.17	51
③	65	0.07	-0.12	50
④	736	0.83	0.27	42
⑨	12	0.01	-0.07	60



Question #6 PK093104



На којој слици је тачно приказан положај сенке?

Одабери један одговор.

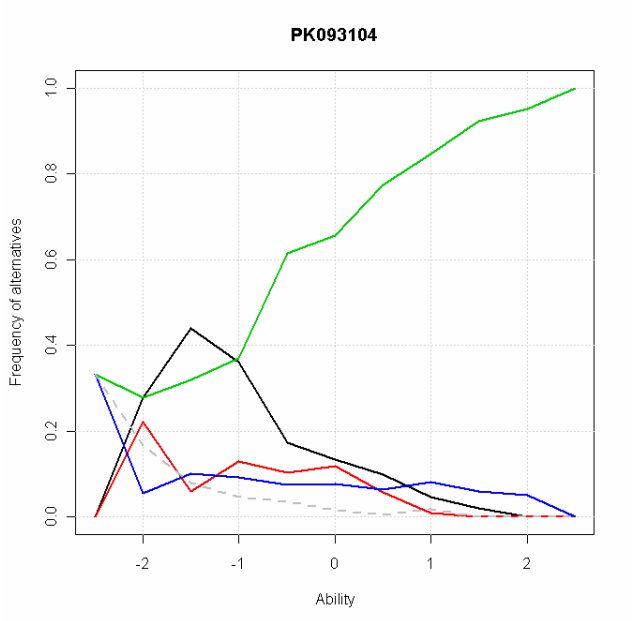
○	а	✖	①
○	б	✖	②
○	в	✓	③
○	г	✖	④
(no response)		✖	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.31	0.71		annual test	0.8	-1.3
trial test	0.30	0.66		trial test	0.8	-0.9

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	141	0.16	-0.24	53
②	71	0.08	-0.11	51
③	584	0.66	0.31	43
④	67	0.08	-0.01	54
⑨	25	0.03	-0.15	50



Question #7 PK123146

Шта је био **повод** за избијање Првог српског устанка?

Одабери један одговор.

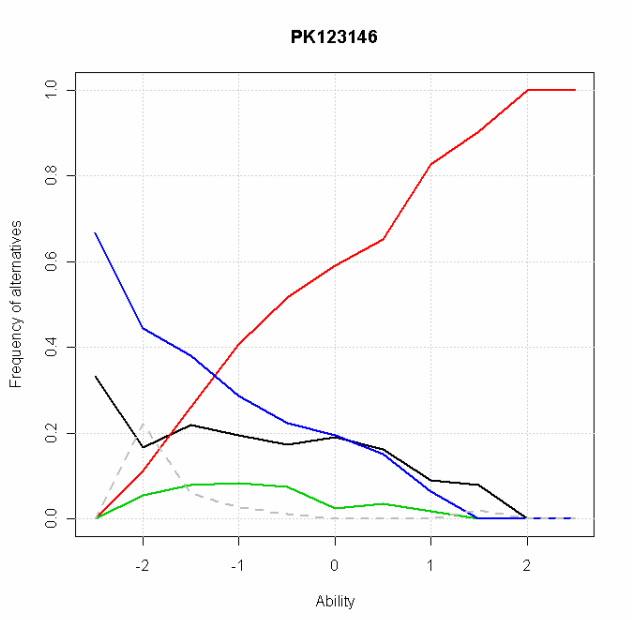
○	данак у крви	✘	①
○	сеча кнезова	✓	②
○	зидање Ћеле куле	✘	③
○	сеоба Срба	✘	④
	(no response)	✘	⑨

Item parameters

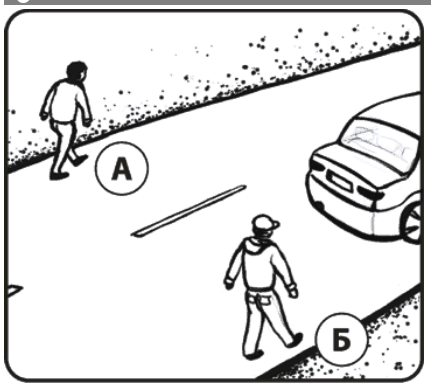
CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.18	0.31		annual test	0.5	1.8
trial test	0.31	0.60		trial test	0.8	-0.6

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	140	0.16	-0.07	38
②	529	0.60	0.31	30
③	40	0.05	-0.10	49
④	167	0.19	-0.23	34
⑨	12	0.01	-0.16	42



Question #8 PD171192



Ко се креће **правилно** улицом без тротоара?

Одабери један одговор

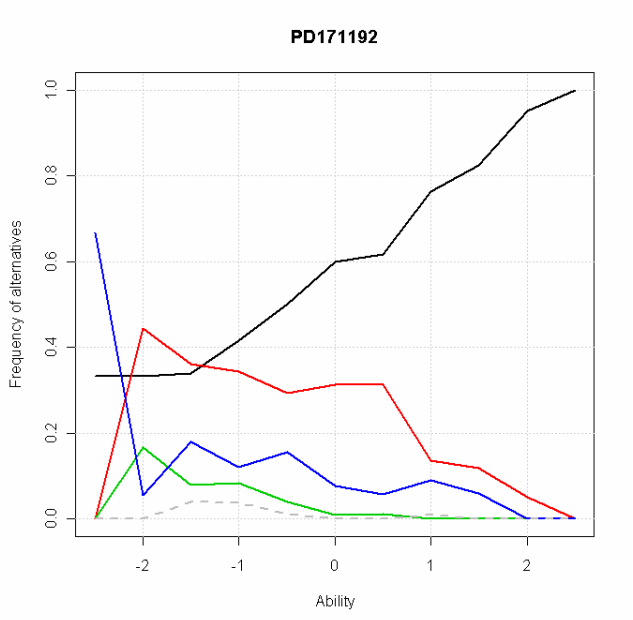
○	пешак А	✓ ①
○	пешак В	✗ ②
○	и пешак А и пешак Б	✗ ③
○	ни пешак А ни пешак Б	✗ ④
	(no response)	✗ ⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.23	0.45		annual test	0.6	0.4
trial test	0.23	0.58		trial test	0.6	-0.6

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	519	0.58	0.23	44
②	243	0.27	-0.09	42
③	27	0.03	-0.18	48
④	90	0.10	-0.11	46
⑨	9	0.01	-0.09	75



Question #9 PP033036

Који се услови за живот у стаништима највише мењају током године?

Одабери један одговор

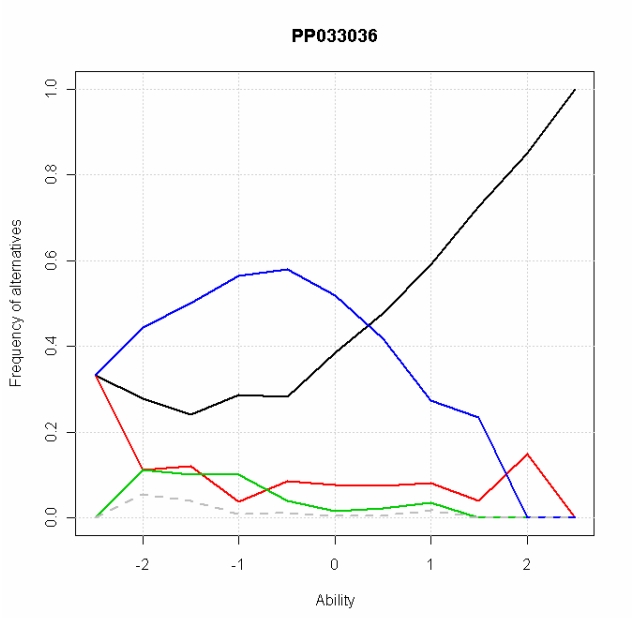
○	светлост и топлота	✓	①
○	количина воде	✘	②
○	количина ваздуха	✘	③
○	број живих бића	✘	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.14	0.23		annual test	0.5	2.9
trial test	0.21	0.42		trial test	0.6	0.7

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	370	0.42	0.22	50
②	68	0.08	0.00	55
③	35	0.04	-0.12	52
④	404	0.45	-0.16	40
⑨	11	0.01	-0.05	26



Question #10 PP013010

На трави је стајала канта са ђубретом. После неколико дана трава испод канте је пожутела. Због чега?

Одабери један одговор.

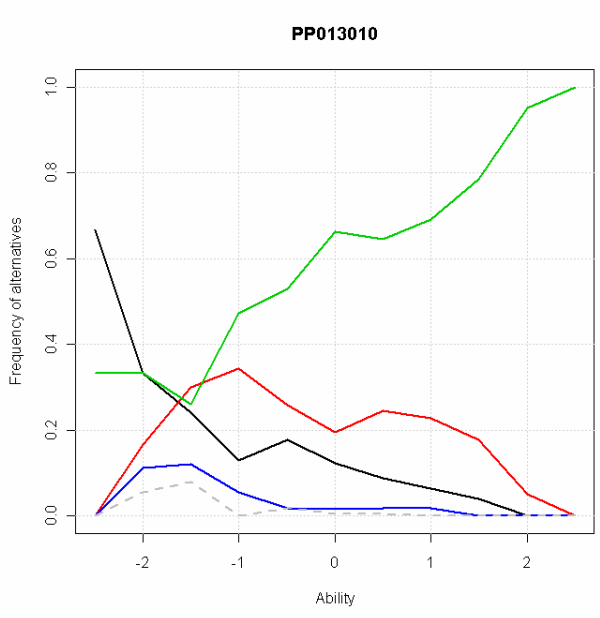
○	Није имала довољно воде.	✘	①
○	Није имала довољно ваздуха.	✘	②
○	Није имала довољно сунчеве светлости.	✓	③
○	Није имала довољно топлоте.	✘	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.20	0.49		annual test	0.5	0.1
trial test	0.21	0.60		trial test	0.5	-0.8

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	111	0.12	-0.16	46
②	212	0.24	-0.04	46
③	528	0.59	0.20	42
④	25	0.03	-0.13	43
⑨	12	0.01	-0.07	44



Question #11 PP043051

Које природно богатство људи користе и као извор енергије и као сировину?

Одабери један одговор.

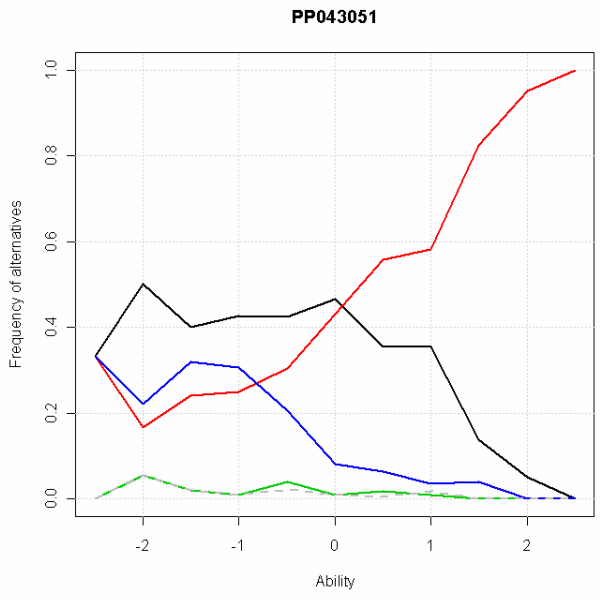
○	гас	✖	①
○	дрво	✓	②
○	камен	✖	③
○	земљиште	✖	④
	(no response)	✖	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.24	0.23		annual test	0.8	1.8
trial test	0.26	0.45		trial test	0.7	0.4

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	343	0.39	-0.07	39
②	400	0.45	0.26	40
③	16	0.02	-0.05	45
④	117	0.13	-0.25	38
⑨	12	0.01	-0.04	38



Question #12 PP063072

Једног летњег јутра Марко је приметио да су сви плочници суви, а да на трави у баштама има капљица воде. Чиме се то може објаснити?

Одабери један одговор.

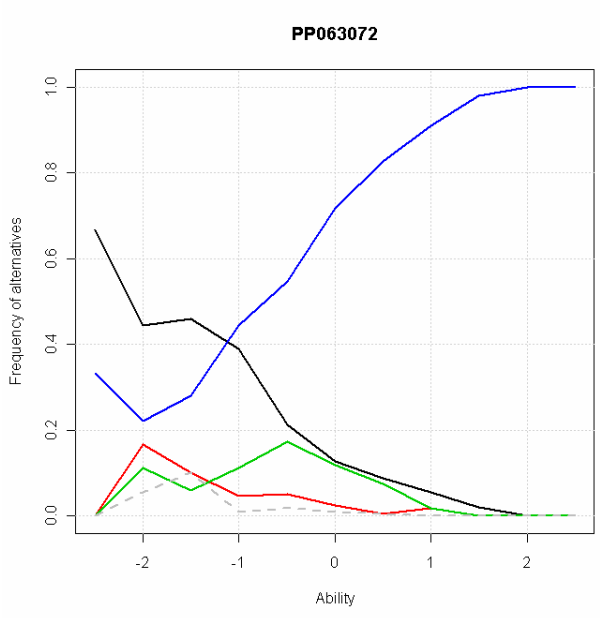
○	Падала је киша.	✘	①
○	Падао је град.	✘	②
○	Трава је пустила сокове.	✘	③
○	На трави се згуснула водена пара.	✓	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.28	0.60		annual test	0.7	-0.7
trial test	0.35	0.69		trial test	1.0	-0.9

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	155	0.17	-0.29	54
②	29	0.03	-0.12	56
③	84	0.09	-0.07	58
④	607	0.68	0.36	52
⑨	13	0.01	-0.11	66



Question #13 PP073082

Зашто се посуде за кување хране праве од метала?

Одабери један одговор.

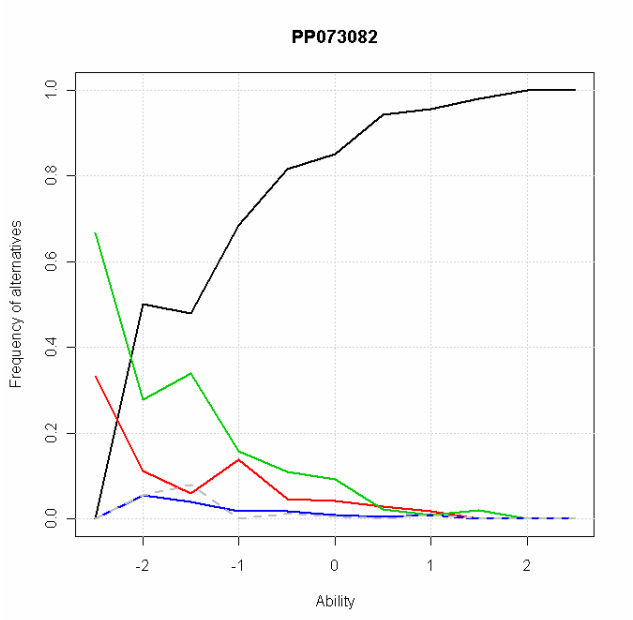
○	зато што метал добро проводи топлоту	✓	①
○	зато што метал проводи електричну струју	✘	②
○	зато што се метал не ломи	✘	③
○	зато што метал није провидан	✘	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.39	0.81		annual test	1.1	-1.6
trial test	0.30	0.83		trial test	1.1	-2.0

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	739	0.83	0.30	40
②	43	0.05	-0.12	39
③	83	0.09	-0.24	44
④	12	0.01	-0.06	45
⑨	11	0.01	-0.10	47



Question #14 PK111131

Косовска битка се одиграла 1389. године. Који је то век?

Одабери један одговор.

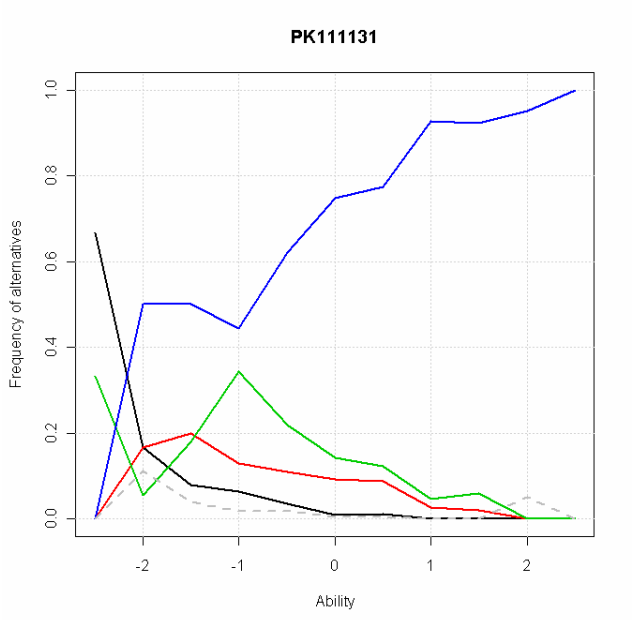
○	једанаести век	✘	①
○	дванаести век	✘	②
○	тринаести век	✘	③
○	четрнаести век	✓	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.29	0.40		annual test	0.8	0.6
trial test	0.27	0.71		trial test	0.7	-1.4

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	26	0.03	-0.19	45
②	83	0.09	-0.11	34
③	137	0.15	-0.13	40
④	631	0.71	0.27	31
⑨	11	0.01	-0.10	38



Question #15 PK121136

Ко је био вођа првог српског устанка?

Одабери један одговор.

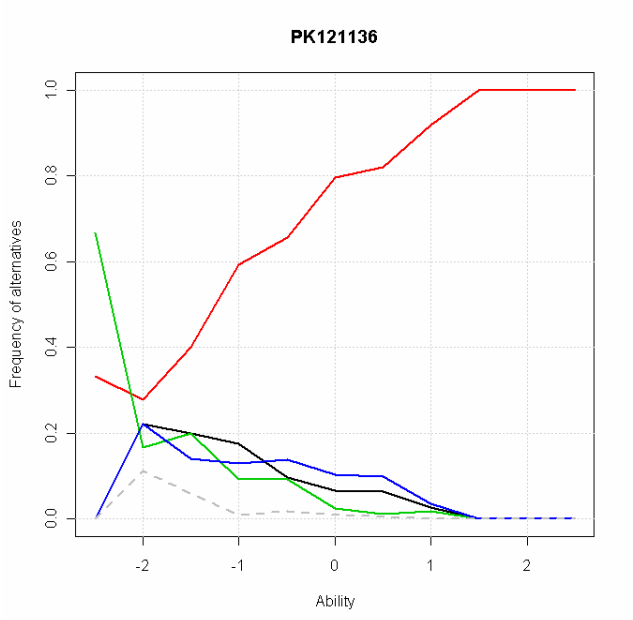
○	Милош Обреновић	✘	①
○	Карађорђе	✓	②
○	Цар Душан	✘	③
○	Петар Први Карађорђевић	✘	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.28	0.49		annual test	0.7	0.1
trial test	0.29	0.75		trial test	0.9	-1.5

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	74	0.08	-0.15	27
②	667	0.75	0.29	24
③	50	0.06	-0.21	30
④	86	0.10	-0.08	30
⑨	11	0.01	-0.11	37



Question #16 PK123144

Које су тврдње о животу у време Немањића тачне?

Одабери бар један одговор.

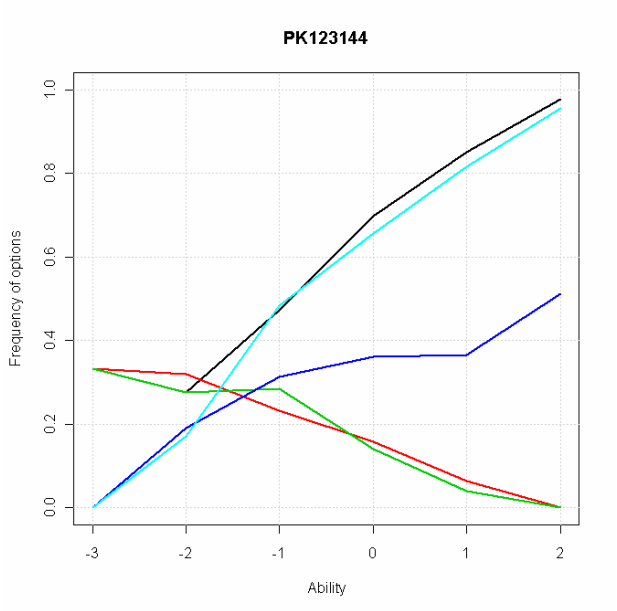
☐	највећи број становника чинили су сељаци	✓	❶
☐	куће су обично биле грађене од цигала	✗	❷
☐	сва деца су ишла у школу	✗	❸
☐	највећи број људи бавио се трговином	✗	❹
☐	посуђе се правило од дрвета и печене глине	✓	❺

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.10	0.20		annual test	0.4	3.7
trial test	0.25	0.27		trial test	0.7	1.7

Analysis of options

Option	Number of responses	*p*-value of option	*r*-value of option	Median response time [s]
❶	600	0.68	0.31	65
❷	134	0.15	-0.19	72
❸	133	0.15	-0.22	67
❹	309	0.35	0.10	68
❺	576	0.65	0.33	65



Question #17 PP043052

Зашто је вода обновљив извор енергије?

Одабери један одговор.

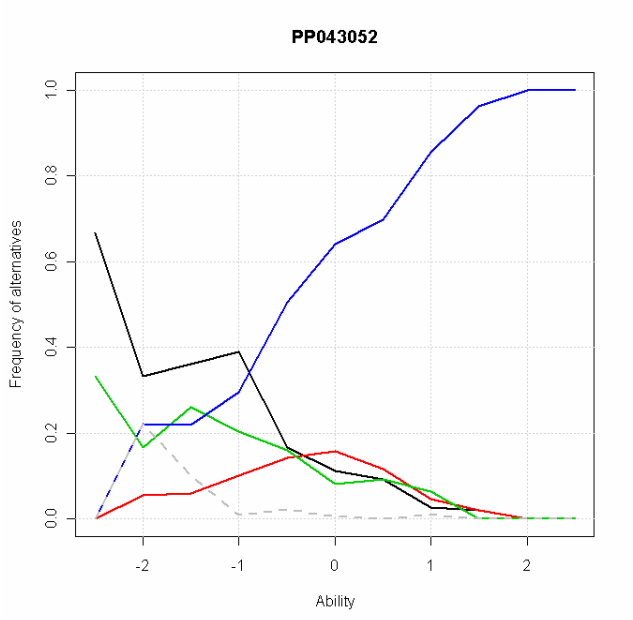
○	Зато што се налази свуда око нас.	✘	①
○	Зато што се може наћи у три агрегатна стања.	✘	②
○	Зато што је има највише на Земљи.	✘	③
○	Зато што непрекидно кружи у природи.	✓	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.37	0.56		annual test	1.0	-0.3
trial test	0.33	0.61		trial test	1.0	-0.5

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	134	0.15	-0.26	42
②	95	0.11	-0.01	47
③	104	0.12	-0.15	41
④	541	0.61	0.33	38
⑨	14	0.02	-0.14	39



Question #18 PP061067

Које је **заједничко својство** воде и свих других течности?

Одабери један одговор.

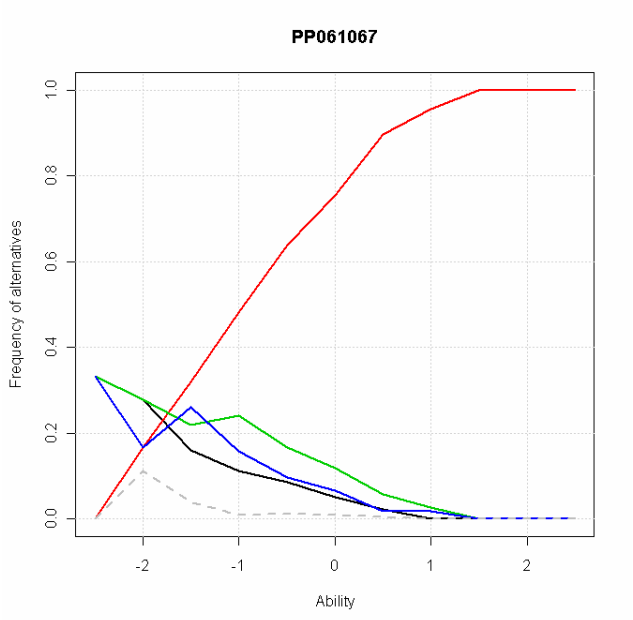
○	нема боју	✘	①
○	нема сталан облик	✓	②
○	нема мирис	✘	③
○	нема укус	✘	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.50	0.60		annual test	1.5	-0.4
trial test	0.36	0.73		trial test	1.2	-1.1

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	54	0.06	-0.18	44
②	649	0.73	0.35	37
③	106	0.12	-0.18	43
④	67	0.08	-0.19	38
⑨	12	0.01	-0.04	70



Question #19 PK093107

Због чега ће згужвани лист папира брже пасти на тло него исти лист папира који није згужван?

Одабери један одговор.

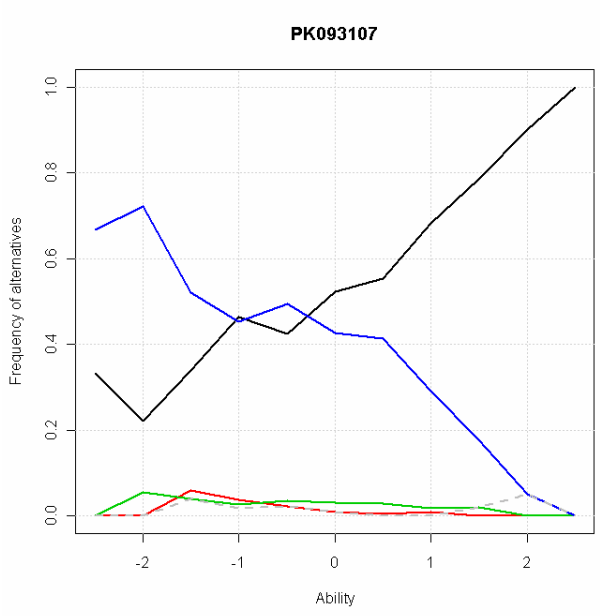
○	због облика	✓	①
○	због храпавости	✘	②
○	због дебљине	✘	③
○	због тежине	✘	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.17	0.41		annual test	0.5	0.9
trial test	0.20	0.54		trial test	0.5	-0.3

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	474	0.53	0.19	42
②	15	0.02	-0.09	37
③	26	0.03	-0.02	46
④	361	0.41	-0.15	40
⑨	12	0.01	-0.05	22



Question #20 PK101113

Ако је кућа на истоку, на којој страни света је шума у односу на кућу?

Одабери један одговор.

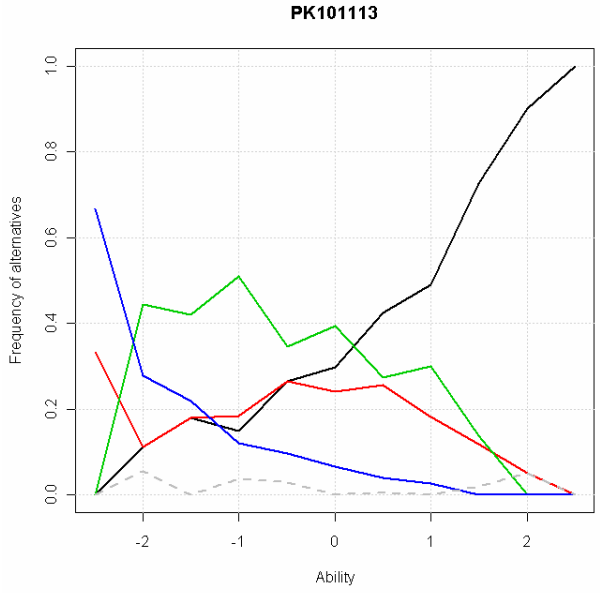
○	на југу	✓	①
○	на северу	✘	②
○	на западу	✘	③
○	на истоку	✘	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.20	0.28		annual test	0.6	1.8
trial test	0.26	0.35		trial test	0.7	1.0

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	312	0.35	0.26	47
②	192	0.22	0.00	50
③	304	0.34	-0.14	41
④	67	0.08	-0.20	46
⑨	13	0.01	-0.05	54



Question #21 PK123142

Цар Душан је био велики освајач и моћни владар Србије. Због тога је у народу запамћен као **Душан Силни**. Због чега је његовог сина народ назвао **Урош Нејаки**?

Одабери један одговор.

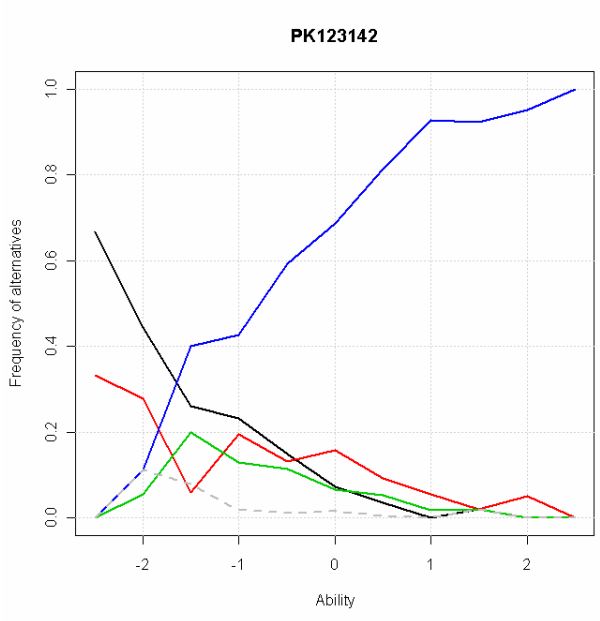
○	Био је најмлађи Душанов син.	✘	①
○	Био је нежне и крхке грађе.	✘	②
○	За време његове владавине није донет нови законик.	✘	③
○	За време његове владавине великаши су распарчали царство.	✓	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.30	0.40		annual test	0.8	0.6
trial test	0.32	0.69		trial test	0.9	-1.0

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	94	0.11	-0.28	56
②	102	0.11	-0.07	56
③	68	0.08	-0.13	56
④	611	0.69	0.32	48
⑨	13	0.01	-0.07	61



Question #22 PD141159

Шта су од наведеног **симболи** сваке државе?

Одабери бар један одговор.

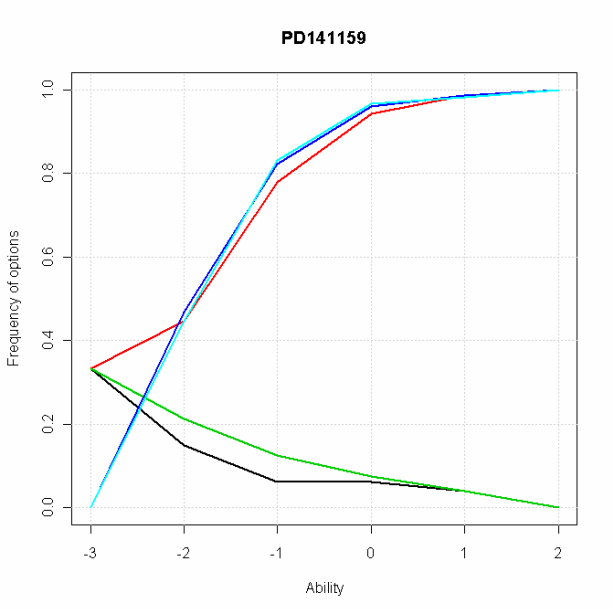
☐	територија	✘	❶
☐	химна	✓	❷
☐	главни град	✘	❸
☐	застава	✓	❹
☐	грб	✓	❺

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.39	0.71		annual test	1.1	-1.0
trial test	0.26	0.77		trial test	0.8	-1.7

Analysis of options

Option	Number of responses	*p*-value of option	*r*-value of option	Median response time [s]
❶	53	0.06	-0.08	47
❷	791	0.89	0.31	37
❸	71	0.08	-0.12	46
❹	807	0.91	0.32	38
❺	807	0.91	0.31	38



Question #23 PD133154

Ана и Милица су се посвађале за време школског одмора. Ана је тужила Милицу код учитељице. Учитељица је изгрдила Милицу, не саслушавши шта она има да каже о томе. Које се дечје право **не поштује** у овом примеру?

Одабери један одговор.

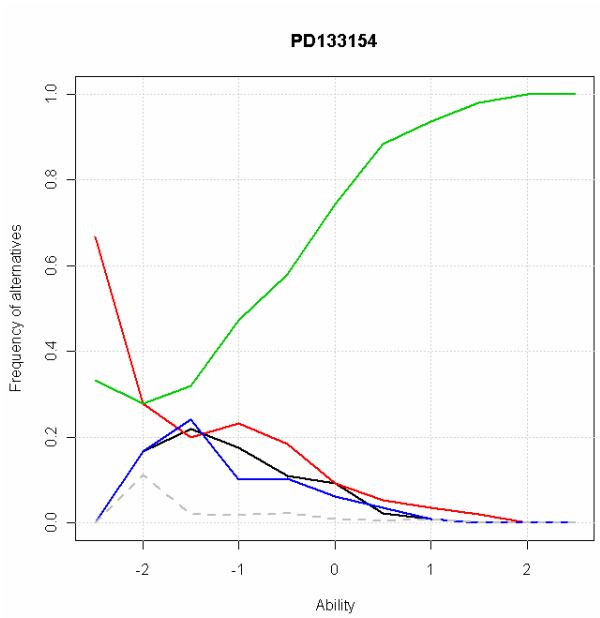
○	право на слободно време и игру	✘	①
○	право на заштиту када их одрасли злостављају	✘	②
○	право на изражавање властитог мишљења	✓	③
○	право на личне тајне	✘	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.51	0.61		annual test	1.5	-0.4
trial test	0.36	0.72		trial test	1.1	-1.1

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	73	0.08	-0.19	64
②	103	0.12	-0.18	62
③	639	0.72	0.36	57
④	61	0.07	-0.17	51
⑨	12	0.01	-0.06	61



Question #24 PP013009

Животиње настањене у земљишту помажу да оно буде:

Одабери један одговор.

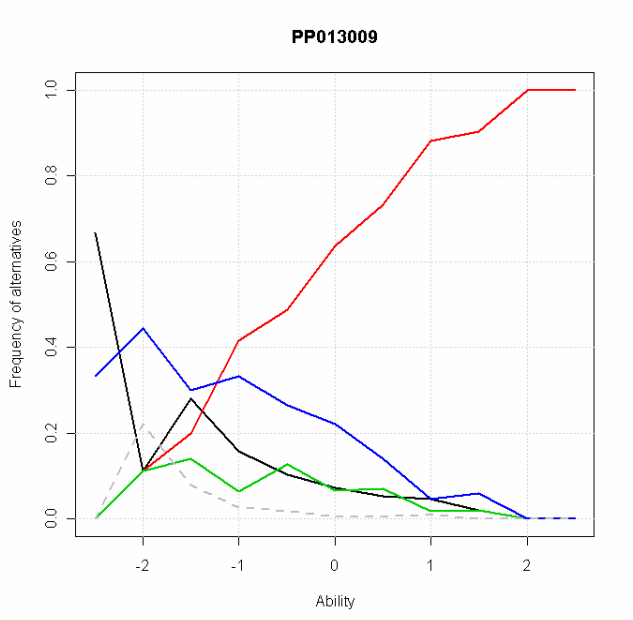
○	топлије	✘	①
○	растреситије	✓	②
○	влажније	✘	③
○	мање загађено	✘	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.38	0.49		annual test	1.0	0.1
trial test	0.34	0.62		trial test	0.9	-0.6

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	80	0.09	-0.17	44
②	549	0.62	0.34	40
③	66	0.07	-0.09	44
④	178	0.20	-0.19	44
⑨	15	0.02	-0.13	46



Question #25 PP083095

Три кашике соли сипане су у посуду са водом. Вода у којој је растворена со стављена је у шерпу да прокува. Вода са сољу је испаравала све док на дну шерпе није остала само со. Шта је истраживач хтео да покаже овим огледом?

Одабери један одговор

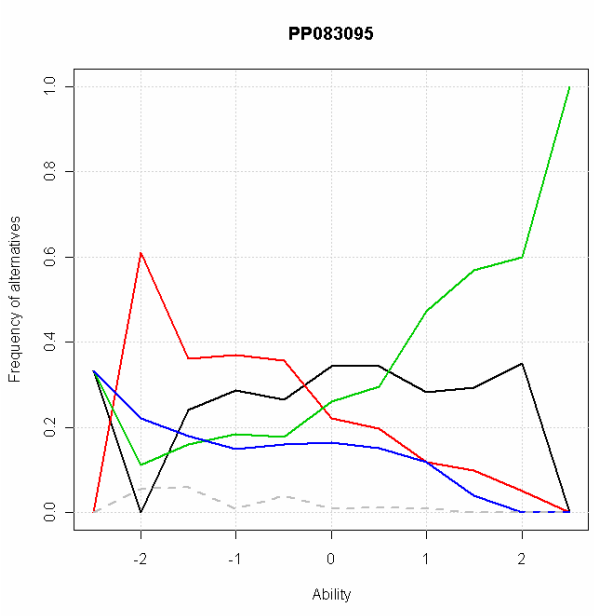
○	да вода испарава када се јако загреје	✘	①
○	да се со раствара у води	✘	②
○	да је растварање соли у води повратна промена	✓	③
○	да је растварање соли у води неповратна промена	✘	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.23	0.20		annual test	0.7	2.1
trial test	0.21	0.29		trial test	0.6	1.8

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	265	0.30	0.08	82
②	217	0.24	-0.20	73
③	256	0.29	0.21	81
④	131	0.15	-0.09	90
⑨	19	0.02	-0.08	71



Question #26 PK101118

Ако си окренут лицем према изласку сунца, која је страна света **иза** тебе?

Одабери један одговор.

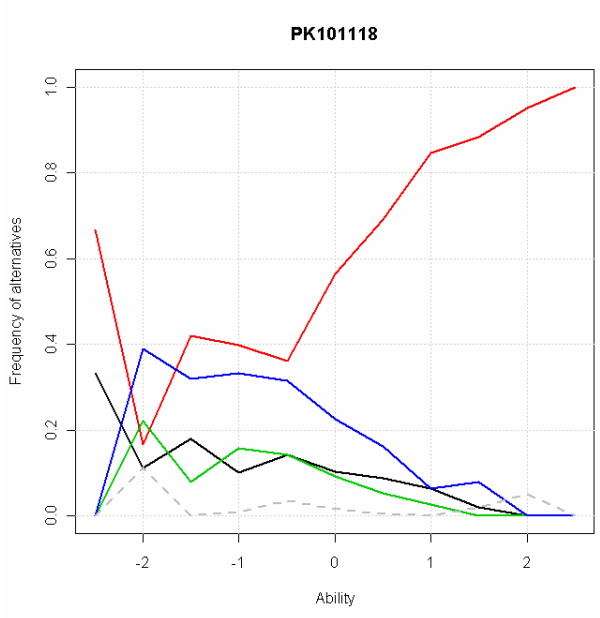
○	исток	✘	①
○	запад	✓	②
○	север	✘	③
○	југ	✘	④
(no response)		✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.29	0.55		annual test	0.7	-0.3
trial test	0.30	0.58		trial test	0.8	-0.5

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	83	0.09	-0.09	39
②	517	0.58	0.29	37
③	78	0.09	-0.12	37
④	195	0.22	-0.19	41
⑨	15	0.02	-0.04	27



Question #27 PK111132

Које време показује часовник?

Одабери један одговор.

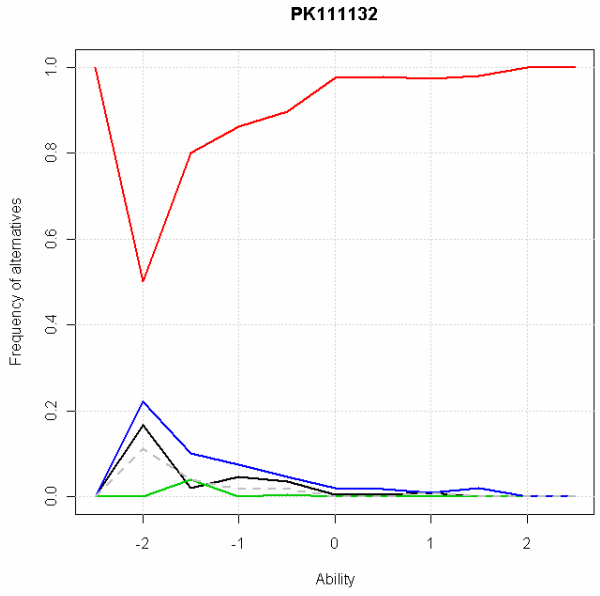
○	пола дванаест	✘	①
○	пола један	✓	②
○	пет минута до шест часова	✘	③
○	шест часова и пет минута	✘	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.22	0.90		annual test	0.8	-3.3
trial test	0.21	0.93		trial test	1.0	-3.1

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	18	0.02	-0.10	42
②	823	0.93	0.21	31
③	3	0.00	-0.07	59
④	34	0.04	-0.14	44
⑨	10	0.01	-0.10	80



Question #28 PK121139

Прва српска држава била је:

Одабери један одговор.

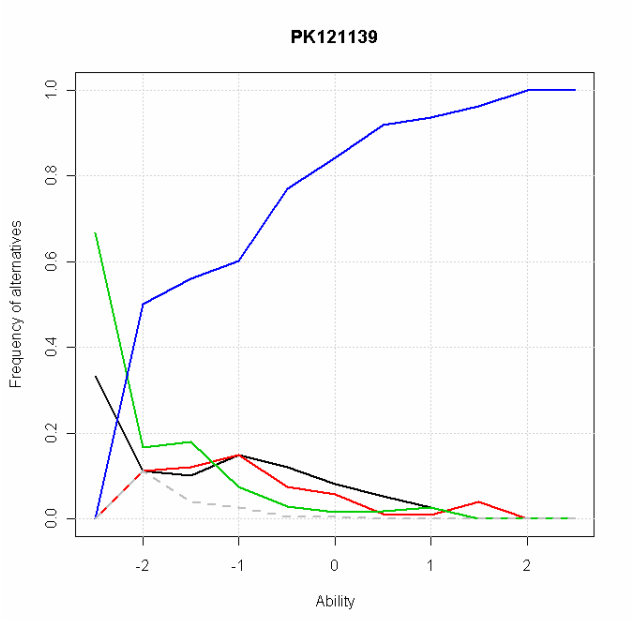
○	Кнежевина Србија	✘	①
○	Душаново царство	✘	②
○	Велика Србија	✘	③
○	Рашка	✓	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.16	0.52		annual test	0.4	-0.2
trial test	0.27	0.81		trial test	0.9	-1.9

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	73	0.08	-0.11	31
②	53	0.06	-0.13	34
③	36	0.04	-0.18	38
④	718	0.81	0.27	20
⑨	8	0.01	-0.12	37



Question #29 PD143165

Са којом од наведених држава се граничи Република Србија?

Одабери један одговор

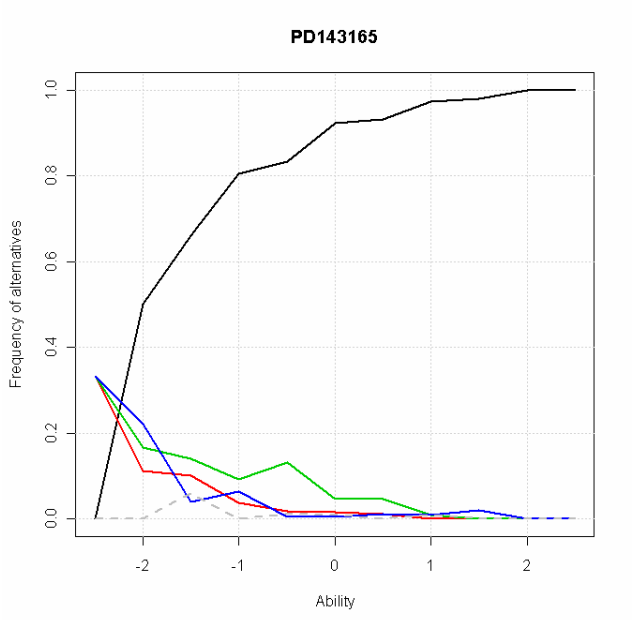
○	Мађарска	✓	①
○	Грчка	✘	②
○	Словенија	✘	③
○	Немачка	✘	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.40	0.86		annual test	1.4	-1.8
trial test	0.24	0.88		trial test	0.9	-2.5

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	781	0.88	0.24	24
②	20	0.02	-0.13	35
③	60	0.07	-0.14	34
④	19	0.02	-0.14	35
⑨	8	0.01	-0.04	27



Question #30 PD153176

Шта се најчешће производи у равничарским крајевима?

Одабери један одговор.

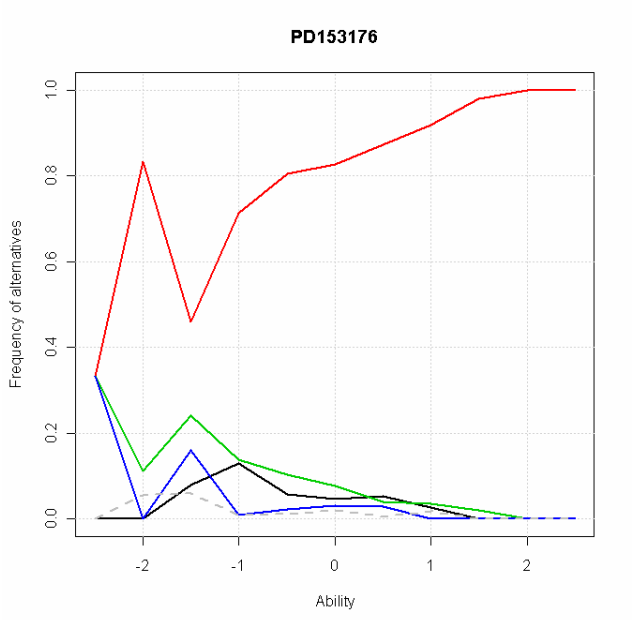
○	аутомобили	✘	①
○	храна	✓	②
○	бакар	✘	③
○	намештај	✘	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.30	0.68		annual test	0.8	-1.2
trial test	0.19	0.83		trial test	0.7	-2.5

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	47	0.05	-0.05	39
②	732	0.82	0.20	28
③	72	0.08	-0.15	35
④	23	0.026	-0.09	39
⑨	14	0.02	-0.05	21



Question #31 PP021015

По чему се биљке разликују од осталих живих бића?

Одабери један одговор.

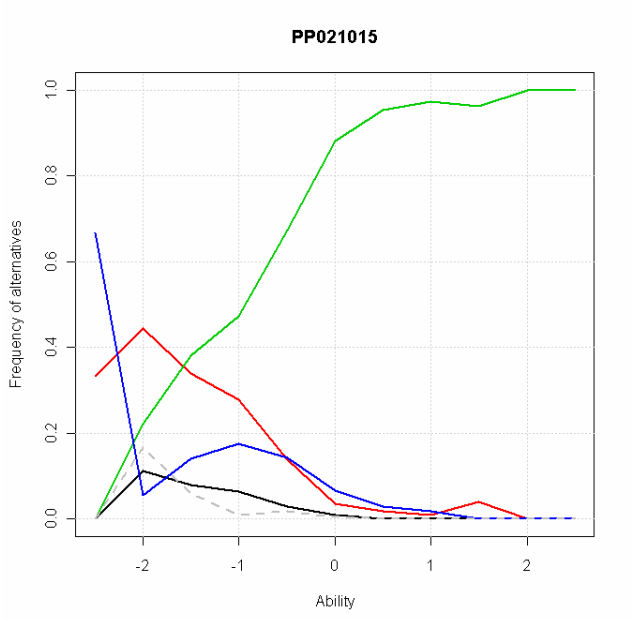
○	дишу	✘	①
○	расту и развијају се	✘	②
○	стварају храну	✓	③
○	остављају потомство	✘	④
	(no response)	✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.45	0.63		annual test	1.3	-0.6
trial test	0.39	0.78		trial test	1.3	-1.3

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	20	0.02	-0.14	46
②	92	0.10	-0.29	40
③	693	0.78	0.39	36
④	74	0.08	-0.14	47
⑨	9	0.01	-0.13	43



Question #32 PK093106

а

б

в

г

Ком раднику ће бити најлакше да вуче терет?

Одабери један одговор.

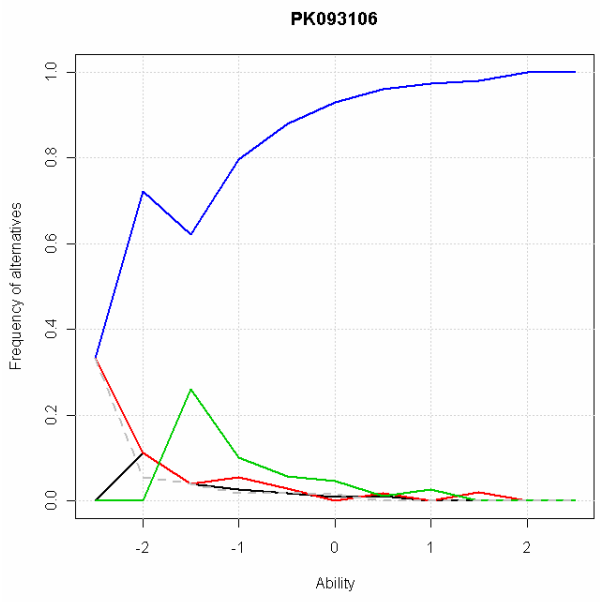
○	а	✘	①
○	б	✘	②
○	в	✘	③
○	г	✓	④
(no response)		✘	⑨

Item parameters

CTT				IRT		
	r-value	*p*-value			*a*	*b*
annual test	0.36	0.80		annual test	1.0	-1.6
trial test	0.23	0.90		trial test	0.9	-2.7

Analysis of alternatives

Response code	Number of responses	*p*-value of alternative	*r*-value of alternative	Median response time [s]
①	14	0.02	-0.10	52
②	20	0.02	-0.11	37
③	47	0.05	-0.13	48
④	796	0.90	0.23	39
⑨	11	0.01	-0.12	32



Appendix 2: List of elementary schools that participated in on-line testing

1 Bratstvo, Aradac
2 Jelena Ćetković, Beograd
3 Jovan Kursula, Varvarin
4 4. oktobar, Vojvoda Stepa
5 Dositej Obradović, Vrba
6 Bratstvo Jedinstvo, Vrbas
7 Petar Petrović Njegoš, Vrbas
8 Svetozar Miletić, Vrbas
9 Desanka Maksimović, Gornji Milanovac
10 Hristo Botev, Dimitrovgrad
11 Dr Aleksandar Sabovljev, Ečka
12 Moša Pijade, Žagubica
13 Desanka Maksimović, Zaječar
14 2. oktobar, Zrenjanin
15 Petar Petrović Njegoš, Zrenjanin
16 Milinko Kušić, Ivanjica
17 Feješ Klara, Kikinda
18 Đura Jakšić, Kovin
19 Đura Jakšić, Kragujevac
20 Jovan Popović, Kragujevac
21 Moma Stanojlović, Kragujevac
22 Nada Popović, Kruševac
23 22. jul, Krčedin
24 Dobrosav Radosavljević Narod, Mačvanska Mitrovica
25 Vuk Karadžić, Negotin
26 Braća Stefanović, Neuzina
27 Branko Miljković, Niš
28 Bubanjski heroji, Niš
29 Vožd Karađorđe, Niš
30 Miroslav Antić, Niš
31 Mirko Tomić, Obrež
32 Đura Jakšić, Oreškovica
33 Miroslav Antić, Palić
34 Jovan Jovanović Zmaj, Pančevo
35 Đura Jakšić, Ravni
36 Ivo Lola Ribar,Ruma
37 Bačka Palanka, Sveti Sava
38 Branko Radičević, Smederevo
39 Dimitrije Davidović, Smederevo
40 Ivo Lola Ribar, Sombor
41 Vuk Karadžić, Srbobran
42 Đura Jakšić, Srpska Crnja
43 Vuk St. Karadžić, Starčevo
44 Stefan Nemanja, Studenica
45 10. oktobar, Subotica
46 Ivan Goran Kovačić, Subotica
47 Majšanski put, Subotica
48 Sečenji Ištvan, Subotica
49 Matija Gubec, Tavankut
50 Aleksa Dejović, Temerin
51 Mladost, Tomaševac
52 Petar Drapšin, Turija
53 Slobodan Sekulić, Užice
54 Milica Pavlović, Čačak
55 Sveti Sava, Čitluk
56 Laza K. Lazarević, Šabac

Appendix 3: Trial test sample characteristics

The trial test was conducted on convenience sample that consisted of 926 students from 50 elementary schools. All schools interested in on-line testing took part in the study. They were allowed to select students for the testing according to their own needs. In some schools teachers intended to test all fourth-grade students; on the other hand, but also there were schools where teachers were just interested to see what is on-line testing like and how does it work. Schools that applied for the testing surely belong to better organized schools where computers in computer room have Internet connection and where teachers show interest to participate in the project of this kind.

Considering the data obtained from schools about the way selection of students was done, we can see that 50% of the sample consists of more or less randomly selected students (random selection, all students from a class, first 15 students from the register, etc.) while 40% of students were selected according to school marks, computer skills and motivation, or because students had asked to participate in the testing. For 10% of students schools did not provide data about the way of selection.

If we compare ability estimations[2] for each of these groups, we can see that students selected for the trial test performed significantly better than students at the annual test. Mean abilities for each of these groups are given in Table 4 with the estimated standard error is given in parenthesis.

	Number of students	Mean ability
Representative sample (the annual test)	1846	0.00(2)
Convenience sample (the trial test)	903	0.61(3)
• random sample within a class or all students in a class	• 454	0.52(4)
• students selected according to school marks, motivation and similar	• 354	0.74(5)
• without information about the way of selection	• 95	0.62(8)

Table 4: Mean ability for students selected in different way

In order to see how convenience sample is different from the representative one, we have compared students' grades in Nature & Society at the end of previous semester. We can notice that students from the convenience sample have much greater mean course grade (Table 5).

	Number of students	Mean course grade
Representative sample (the annual test)	4115	4.10(2)[3]
Convenience sample (the trial test)	903	4.54(3)
• random sample within a class or all students in a class	• 454	4.31(5)
• students selected according to school marks, motivation and similar	• 354	4.76(3)
• without information about the way of selection	• 95	4.82(5)

Table 5: Mean course grade in Nature & Society for students selected in different way

2 For the reasons of comparison on the same scale with annual test, mean estimated ability is displayed on the scale of the annual test.

3 This data is obtained from questionnaire for the preparation of standard setting test for the same population.

Appendix 4: Questionnaire for students

After the testing was finished, teachers asked students to fulfill short on-line questionnaire before they log out. This questionnaire had three questions concerning course grade in Nature & Society, the frequency of computer usage, and applied effort. The questionnaire was answered by 60% of students who took part in the testing.

- Question number 1: What grade did you have in Nature & Society at the end of previous semester?
  (proposed answers from 1 (lowest) to 5 (highest))

We can see that most of students had grade 5 (highest) in Nature & Society. The graph of estimated ability depending on the grade at the end of semester is given on the Figure 6. For the sake of comparison, we have denoted girls' results with red dots, and boys' results with blue. Obviously, students with higher course grade, in average, had better test results.

Grade	Number of students
1	2
2	21
3	40
4	94
5	375
unknown	371

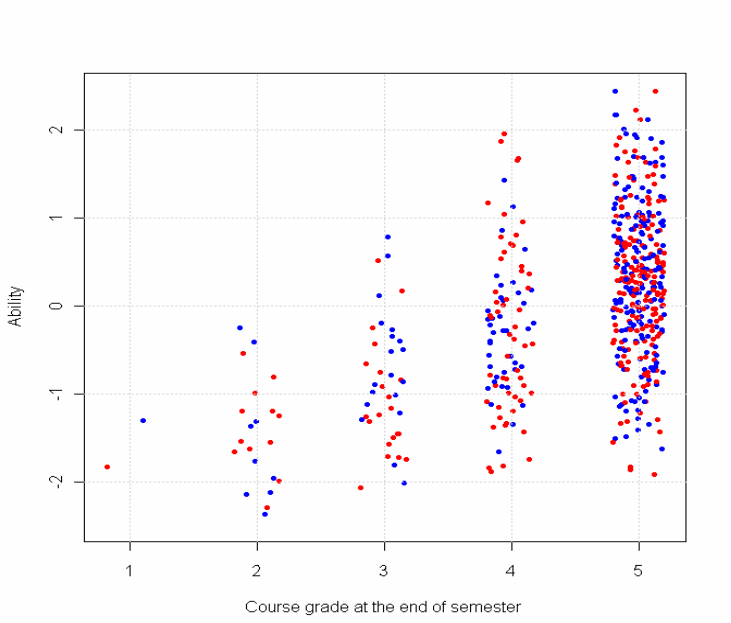


Figure 6

- Question number 2: How often do you use computer?
  1 this is my first time to use computer
  2 almost never
  3 approximately once a month
  4 several times a month
  5 almost every day

From answers to this question, we can see that great majority of students use computer regularly, and that more than a half of them use computer almost every day. Finding from our previous study in 2007 [4] that students who use computer more frequently have better results is not so obvious any more. The graph of estimated ability depending on the frequency of computer usage is given on Figure 7. We have denoted that girls' results with red and boys' results with blue dots. Here we can see that there is no significant difference between how often boys and girls use computer.

Frequency of computer usage	Number of students
1	8
2	33
3	22
4	125
5	342
unknown	373

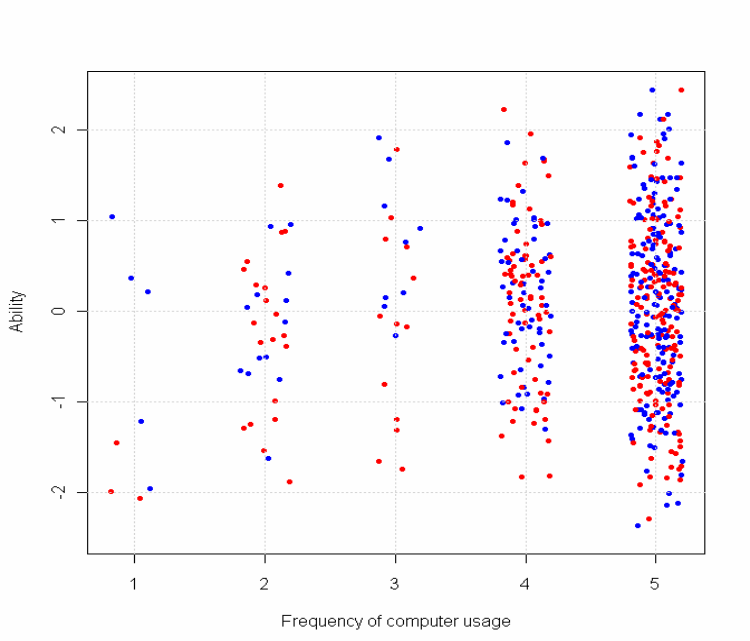


Figure 7

- Question number 3: How much effort did you apply to answer all question in the test?
  (proposed answers from 1 (not at all) to 5 (very much))

From the answers to the third question, we can see that more than a half of students said they applied great effort to do this test. The graph of estimated ability depending on applied effort is given on Figure 8. We have denoted girls' results with red and boys' results with blue dots. We can see that students who applied greater effort achieved better result at the test. Also, we can see there is no significant difference between effort applied by boys and girls.

Applied effort	Number of students
1	0
2	2
3	25
4	107
5	394
unknown	375

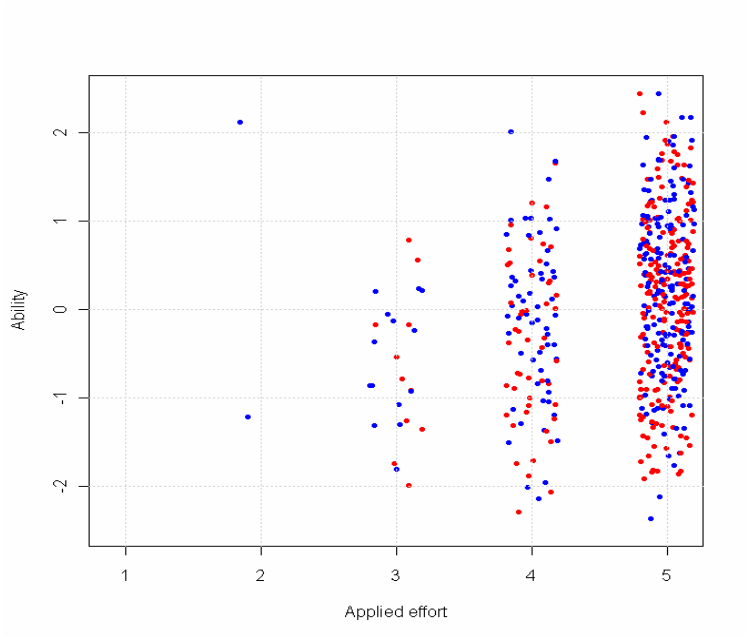


Figure 8

References

1. Картал, В., *Извештај о главном тестирању у школској 2008/2009. за предмет Природа и друштво*. 2009, Завод за вредновање квалитета образовања и васпитања: Београд.
2. Вербић, С. and Б. Томић, *Рачунарски тестови знања у софтверском пакету Moodle: Приручник за наставнике*. 2009, Београд: Завод за вредновање квалитета образовања и васпитања.
3. Tomić, B., I. Lacković, and S. Verbić, *Ttest module*. 2008. Available from http://tracker.moodle.org/browse/CONTRIB-972.
4. Verbić, S. and B. Tomić, *Primenljivost računarskih testova znanja u osnovnoj školi*, in *XIV naučni skup "Empirijska istraživanja u psihologiji"*. 2008, Filozofski fakultet: Beograd, Srbija. p. 39-40.
5. Partchev, I., *Simple interface to the estimation and plotting of IRT models*. 2008.
6. Hanson, B.A., *IRT Command Language*. 2002.
7. Bock, R.D. and R.J. Mislevy, *Adaptive EAP Estimation of Ability in a Microcomputer Environment.* Applied Psychological Measurement, 1982. **6**(4): p. 431.
8. Embretson, S.E. and S.P. Reise, *Item response theory for psychologists*. 2000, Mahwah, N.J.: L. Erlbaum Associates. xi, 371 p.
9. Zumbo, B.D., *A handbook on theory and methods of differential item functioning (DIF)*. Ottawa: National Defense Headquarters. 1999.