

Parallel Serbian Versions of BLOT Test: An Empirical Examination

Ivana Stepanović Ilić¹, Aleksandar Baucal², and Trevor G. Bond³

¹*Institute of Psychology, Faculty of Philosophy, University of Belgrade, Serbia*
²*Department of Psychology, Faculty of Philosophy, University of Belgrade, Serbia*
³*School of Education, James Cook University, Australia*

Logical Operations Test (BLOT) was developed for assessing the transition to formal operational thinking. BLOT is a 35 item multiple-choice test which examines all of the operations which comprise the logico-mathematical structure of formal operations in Piaget's theory. The test was translated into Serbian and used in previously reported research. This work deals with two additional parallel versions of the Serbian BLOT. For each original BLOT item two more parallel items have been constructed by changing the item content and leaving the logical structure of the item the same. Sample consisted of 517 primary and secondary school students. Rasch analysis confirmed that the vast majority of items maintained invariance across at least two test versions: for 19 original items *both* parallel items maintained their invariance, for 14 items *one* of the parallel items had similar parameters and only 2 items did not remain invariant in the parallel tests.

Keywords: formal operational thinking, BLOT, parallel versions in Serbian, Rasch analysis

INTRODUCTION

This work deals with the examination of three versions of BLOT test in Serbian. Two of them have been constructed recently as parallel versions of the original BLOT test translated from English. The research has practical implications on developing a proper parallel version of the test which can be used in further formal operational thinking studies. The results related to psychometric characteristics of items from two constructed parallel versions can also be considered in the framework of competence-performance problem, i.e. in the context of an important theoretical issue regarding the relationship between form and content of thinking on the formal operations stage.

The concept of formal operations is introduced in *The Growth of Logical Thinking from Childhood to Adolescence (GLT)* by Bärbel Inhelder and Jean Piaget (1958). The authors described qualitative changes in cognitive processes evolving in adolescence comparing the new form of thinking to the one on

the previous stage since a development is a process of restructuring, with each structure being incorporated in the next one. Gruber and Vonèche (1995) emphasize five transformations which mark the shift from concrete-operational level to formal operations stage and they are classically considered as distinctive characteristics of formal operational thinking: (1) combinatorial thinking, (2) the ability to differentiate between the real and the possible, (3) hypothetico-deductive thinking, (4) propositional thinking and (5) separating form from the content. Formal operational thinking adolescents are capable to generate system of every possible combination in which all the elements are intertwined in such a way that moving from one element to the other is always possible. Instead of simply coordinating the facts related to reality, hypothetico-deductive thinking makes implications on the basis of possible propositions and, in that way, reaches a unique synthesis of the possible and the necessary (Piaget, 1953). Adolescents can treat content of propositions hypothetically and think about them correctly focusing on their logical connections. Piaget and Inhelder (1969) claim that abstract thinking, in which structure dominates over content of thinking, enables combining the two forms of reversibility. Within that system, inversion and reciprocity are connected in such a way that each operation within the system is at the same time inverse to some other operation, and reciprocal to some third operation, so that there are *four transformations*: identity, negation, reciprocal, and correlative (group INRC). Inhelder and Piaget (1958) introduced different problem-solving situations, usually resembling experiments in physics, which required formal operational thinking for their complete solution. The behavior of children confronted with these problems was systematically analyzed and related to the concept of the formal structured operational schemata. The schemata reflect concepts that emerge in the interaction with certain contents. They are derived from co-ordinations of the operations rather than operations upon objects in the environment.

Studies of formal operations are very different in respect to instruments applied to their examination. In the original investigation of formal operations (Inhelder and Piaget, 1958) the specific Genevan investigative technique called *methode clinique*, or *methode critique*, was used (Bond, 2010). This methodology incorporates a constructive dialog between researcher and subject in which researcher's remarks and enquiries are aimed at exploring organization of subject's behavior in order to infer about underlying intellectual competences. The clinical method is described in a number of sources (Piaget, 1964; Inhelder, 1969; Inhelder, 1989; Bond & Jackson, 1991). Many investigations of formal operations have been based on original tasks from *GLT* (Lovell, 1961; Bart, 1971–1972; Neimark, 1975; Kuhn, Langer, Kohlberg, & Haan, 1977; Martorano, 1977; Bond & Bunting, 1995) although not all of them did strictly follow Genevan procedures of collecting and analyzing data. It was also common to investigate formal operations with adaptations of original tasks or tasks constructed to examine particular formal operation engaged in problem-solving situations described by Inhelder and Piaget in *GLT* (Rowell & Hoffman, 1975; Kuhn & Brannock, 1977; Kuhn, Ho, & Adams, 1979; Lawson, Karplus, & Adi, 1978; Noeltling, 1980; Mwamwenda, 1999). Some researchers developed

instruments on the basis of logico-mathematical model of formal operations structure, which cover a certain number of formal operations (*Piagetian Reasoning Tasks* – see Wylam & Shayer, 1978; *How is your logic?* – see Gray, 1976a, 1976b; *Butch and Slim* – see Ward, 1972). There are also a group of purportedly related studies in which subjects solved tasks based on the rules of classical formal logic (Brainerd, 1976; Ennis, 1978; Goswami, 2001).

Bond's Logical Operations Test (BLOT) was developed for assessing the transition to formal operational thinking. The test is unique because it was designed to examine all the operations which comprise the logico-mathematical structure of formal operations and all the logical schemata of the formal operations stage. The delineation of formal thought structures explicated in Chapter 17 *Concrete and formal structures* in *GLT*, rather than the behavioral descriptions and their ordering, was the starting point for BLOT items development (Bond, 1976, 1978, 1980, 1995). BLOT consists of 35 items in multiple choice format designed as instantiations of the calculus of the sixteen binary operations of truth functional logic and the INRC four-group of operations from Piaget's logical model (see Table 2). Since this work deals with examination of parallel versions of BLOT and regarding the fact that construction of parallel items will be discussed later, it is important to mention that certain BLOT items (several groups of 2–4 items) are linked i.e. their content is related. The report of the test development (Bond, 1976) indicated that BLOT has construct validity as well as a test-retest reliability correlation of .91 for an interval of greater than six weeks. It is confirmed that BLOT items have high levels of concurrent validity with the original Inhelder and Piaget tasks (Bond, 1980, 1989). Further studies underlined the validity and utility of the test (Morley, 1979; Christiansson, 1983; Smith & Knight, 1992; Bond & Jackson, 1991). Besides that, it was shown (Bond, 1995; 1997) that BLOT and PRTIII (which has very good face, construct and predictive validity – see Shayer, 1978, 1979; Shayer & Adey, 1981) measure the same underlying psychological trait, the development of formal operations. Almost a decade ago, BLOT was translated into Serbian and successfully applied to investigation of formal thinking in Serbian adolescents (Stepanović, 2004a).

THE PROBLEM

This research deals with empirical examination of parallel versions of Serbian BLOT. The main goal is to test two new versions of BLOT in Serbian and to discover whether they could be considered as parallel versions of the original test translated in Serbian. The parallel versions were developed for the purpose of testing the influence of peers' interaction on the development of formal operations (Stepanović, 2010). The interaction study had balanced experimental design. In order to prevent the effect of memorizing the tasks (used in the intervention phase) on the post-test results, it was necessary to construct parallel versions of BLOT items. For that reason, it was important to develop items with psychometric properties which will match with those of the

original BLOT items as much as possible. The two parallel versions of Serbian BLOT have been constructed in order to provide enough items with good parameters so that they can be an adequate replacement for the original items in the intervention phase. For that reason, the focus of this paper is comparison among parameters of items from parallel versions and original BLOT items. That examination should indicate whether constructed tests can be considered as parallel versions of Serbian BLOT. An even more important question is: Do we have enough items that fit originals or do we have at least one parallel item for each item from the original test? Apart from its practical value, the examination of parallel versions of Serbian BLOT has theoretical importance as well because it is related to the problem of relationship between form and content and to the existence of horizontal decalage phenomenon on the stage of formal operations. It was mentioned in the introduction that one of the crucial characteristics of formal operations is the ability to separate form of the problem from its content. Describing the stage of concrete operations, Piaget (1953) claims that one of the constrictions of concrete operations is their dependence of a content they operate upon. That is a reason why the decalages between thought in one domain and another appear on this stage. In contrast to that, formal operations represent the structured whole which exceeds limits of the previous stage. However, a certain number of studies indicated that subjects are not equally successful on different formal operation tasks which suggested that horizontal decalage appears on this stage as well. These findings provoked numerous discussions regarding form-content relation and phenomenon of horizontal decalage on the stage of formal operations (Stepanović, 2004b). Although this issue is not central problem of the paper in our opinion the investigation that will be presented here can contribute to this topic.

Method

Subjects. The convenience sample consisted of 517 Serbian students: 162 subjects in sixth grade primary school (mean age 12;7, mode – 12;5); 197 subjects from eighth grade primary school (mean age 14;6, mode – 14;4) and 158 secondary school students (mean age 16;5, mode – 16;8). Six classes from each grade were tested, three from urban and three from rural primary schools. Secondary school students came from three grammar school classes and three vocational school classes because secondary schools exist only in towns.

Instruments. Three Serbian versions of the BLOT test were used: Version 1 (V1) – the original BLOT translated in Serbian; Versions 2 (V2) & 3 (V3) are the parallel versions constructed in Serbian language. All versions followed the original 35 items in multiple choice format.

BLOT V1 was translated into Serbian and used in previous research (Stepanović, 2004a). The results of that study showed that Serbian version had good measurement characteristics and that items parameters were very similar to those from the studies which used BLOT in English (Stepanović, 2004a).

Two additional Serbian language versions of BLOT (V2 & V3) were made for the purpose of investigating the role of peer interaction on formal operation development. For each original item V1 two parallel V2 and V3 items were constructed. Parallel items were developed in such a way that changes were made to the items' content only, while the logical structure of the original items remained the same. Wherever possible, the format of items

(grammatical structure and meaning of sentences) was also preserved so that the only change was replacement of one term with another. For the majority of BLOT items it was possible to construct parallel items which so preserved the format of the original item. In order to illustrate such items (see the Example 1, items are presented in Serbian and their translation in English is provided).

Example 1:

Q3, original BLOT in Serbian (V1): *Istraživač ruda je otkrio da se neki plemeniti metali ponekad javljaju zajedno. Ponekad je pronalazio zajedno zlato i srebro; ponekad je nalazio samo srebro; a u svim ostalim slučajevima nije pronalazio ni zlato ni srebro. Koje od sledećih pravila ovaj istraživač smatra istinitim?*

Q3, Parallel test (V2): *Istraživač livadskog bilja je otkrio da se neke lekovite biljke ponekad javljaju zajedno. Ponekad je pronalazio zajedno nanu i kamilicu; ponekad je nalazio samo kamilicu; a u svim ostalim slučajevima nije pronalazio ni nanu ni kamilicu. Koje od sledećih pravila ovaj istraživač smatra istinitim?*

Q3, original BLOT: *A prospector has found that some rich metals are sometimes found together. In his life he has sometimes found gold and silver together; sometimes he has found silver by itself; every other time he has found neither silver nor gold. Which of the following rules has been true for this prospector?*

Q3, translation of the item from V2: *A botanist has found that some medicinal herbs are sometimes found together. In his life he has sometimes found mint and chamomile together; sometimes he has found chamomile by itself; every other time he has found neither chamomile nor mint. Which of the following rules has been true for this botanist?*

However, some parallel items were changed to a greater extent than this because we were not able to find a phenomenon very similar to the one mentioned in the content of the original item. In particular cases it was difficult to replace phenomena from original items with new ones which would be close to everyday life and students' experience (mentioned problems regarding the items' content are labeled with *c* in Table 2). Some problems in construction of parallel items were related to the fact that the found phenomena could not fit the format of original item precisely (labeled with *f* in Table 2). Certain linked items were also problematic because it was hard to fit the new content in a specific manner in which contents of original linked items refer to each other (labeled with *l*). Sometimes two of those problems occurred together, and sometimes all of them (see Table 2). Because of the reasons discussed above several parallel items preserved the logical form of original item but their format was changed (see the Example 2)

Example 2:

Q33, original BLOT in Serbian (V1): *Novčić je bačen u vazduh 10 puta i pri tom je pao na pod. Koja od sledećih tvrdnji bi predstavljala najverovatniji rezultat?*

Q33, Parallel test (V3): *Dečak ima dva identična ključa na privesku. Samo jedan od njih otključava njegov stan. Tokom 10 dana on otključava stan probajući jedan od dva ključa. Koja od sledećih tvrdnji bi predstavljala najverovatniji rezultat?*

Q33, original BLOT: *A coin is flicked into the air and allowed to fall on a flat surface ten (10) times. Which of the following would be the most likely result?*

Q33, translation of the item from V3: *A boy has two identical keys on his key ring. Only one key unlocks door of his apartment. He unlocks the door for 10 days trying one of these two keys. Which of the following would be the most likely result?*

Procedure. All three versions of BLOT were administrated as group tests in 18 classes. Three versions were distributed in each class randomly; thus, one third of students in each class was tested by V1, one third by V2 and one third by V3.

The analytical technique. Rasch analysis (Rasch, 1960; Wright & Masters, 1982; Bond & Fox, 2001) is held to be the most appropriate for this purpose because it is sensitive to the explicitly developmental nature of Piagetian accounts (Bond 1995; Bond, 1997; Bond & Fox, 2001) and it provides an estimation of the unidimensionality of the data set under analysis, which is relevant for comparison among three versions of Serbian BLOT.

Results

The results of Rasch analysis reveal very good psychometric characteristics of all three Serbian BLOT tests. The reliability coefficients for items are more than satisfactory and similar across three versions of BLOT (Table 1). Reliability of subjects for three tests is a bit lower but still very well (0.83 – 0.84). Fit statistics suggests that each of the three versions fit a unidimensional model, i.e. each measures one underlying trait.

Table 1. *Items' measures for three versions of BLOT*

	Version 1	Version 2	Version 3
Summary of item estimates			
Mean	-0.70	-0.54	-0.16
SD	0.99	0.83	0.71
SD (adj.)	0.97	0.81	0.68
Reliability of estimate	0.96	0.95	0.94
Items with zero scores	0	0	0
Items with perfect scores	0	0	0
Fit statistics			
Infit mean square			
Mean	0.99	1.0	1.0
SD	0.12	0.14	0.15
Infit ZSTD			
Mean	0.0	-0.1	-0.1
SD	1.6	1.9	1.8
Outfit mean square			
Mean	0.99	1.0	1.02
SD	0.20	0.2	0.31
Outfit ZSTD			
Mean	0.0	0.0	-0.1
SD	1.5	1.7	1.8

The mean values for items differ across three versions of the test. The original BLOT is the easiest for students ($M=-0.70$). V2 is slightly harder ($M=-0.54$) than the original BLOT, but existing difference can be neglected and difficulties of two tests considered as equal since item error estimates are about 0.20 logits. However, V3 ($M=-0.16$) seems to be more difficult than V2, and it is definitely more difficult than the original BLOT because the means difference is statistically significant. Found differences were the reason to take a closer look to the parameters of items in different version of BLOT. Figure 1 presents items' estimates across three versions. Axis x represents the items, from easiest

to most difficult. On axis *y* is located item's difficulty, in logit units. For each item three estimates, one from each version are displayed.

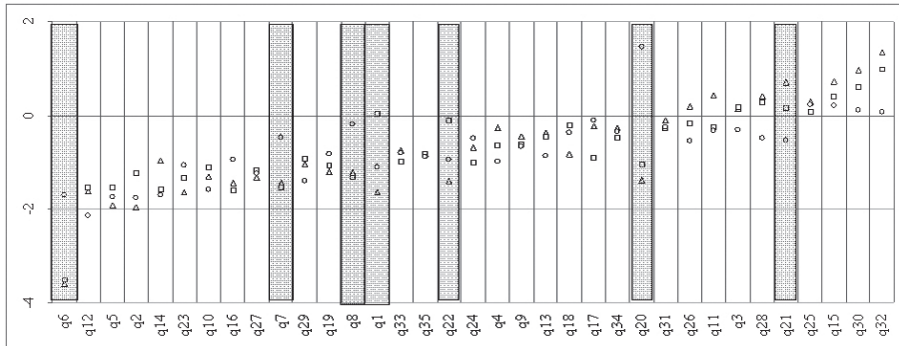


Figure 1. Items' estimates across three Serbian language versions of BLOT

Legend: triangle – V1 (Original Serbian BLOT); square – V2 (Parallel test); circle – V3 (Parallel test)

Some items (q9, q13, q27, q31, q33, q34, q35) have the same difficulty in three versions of the test. For certain items (q3, q10, q12, q15, q19, q33) the differences in estimates are notable but not large. However, difficulty of items (q1, q6, q7, q8, q20, q21, q22) marked with gray rectangles varies in greater degree across three tests.

For more precise information regarding items difficulty across three versions of BLOT it is necessary to test the difference between pairs of estimates, i.e. to compare original items with the corresponding items from V2 and V3. In order to make such comparison, the original items are plotted against the items from parallel versions and invariance of item-difficulty estimates across items measuring the same operation was tested. Pairs of items estimates (original item difficulty and difficulty of parallel item) were plotted onto a scatter plot. If we draw a diagonal line (45° or slope 1) through the point representing the group means of items estimates for two versions of BLOT (the original and one of parallel versions), we construct a line that represented an ideal situation in which all corresponding items would have exactly the same difficulty and lie along that line. Usefully, Rasch modeling provides us with error estimates for items' estimate, and we can use these to construct control lines to see whether the distribution of the plotted item points is close enough to the modeled relationship diagonal line for the measures to be regarded as sufficiently identical (i.e. identical within the limits of measurement error). The formula for constructing the control lines for 95% confidence band around diagonal through the mean item estimates was derived originally from Wright and Stone (1979) and it is presented in Bond and Fox book about Rasch analysis (2001). The precision of the plotting Rasch model estimates depend on the size of the error estimates (Bond & Fox, 2001). Since we had more than 150 subjects per each test, it is reasonable to believe that error of items' estimates will be relatively small. This was justified by the fact that the item error estimates in all versions are approximately the same for all items, mostly about 0.20 logits.

The results of plotting the original items (V1) against items from the Version 2 (V2) are shown on Figure 2. Majority of items (28 out of 35) from the V2 fall within confidence band. Seven items (q3, q6, q7, q8, q12, q13, q35) lie on, or very near, the ideal diagonal line, which means that differences in their estimates are smaller than 0.1 logits. If we have in mind the mentioned item error estimates, we can say that the 14 items from V2 have the same difficulty (within measurement error) as the corresponding originals, since the difference between pairs of estimates is smaller than 0.2 logits. There are 5 items (q2, q11, q14, q17, q18) that are positioned on the control lines, or very close to them, and only 2 items (q1, q22) misfit the invariance model.

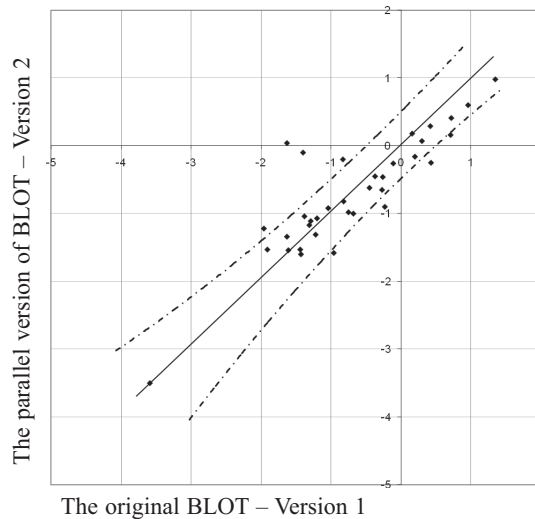


Figure 2. Plotting the original BLOT against Version 2

Comparison among items from Version 3 (V3) and the original items is presented in Figure 3. The majority of items have invariant estimates in these two versions of BLOT (24 of 35). Five items (q25, q27, q33, q34, q35) lie on the ideal diagonal line. If we have less strict criteria and tolerate the difference between corresponding items lower than 0.2 logits, as we did in case of V2, it can be stated that 7 items from V3 represent the perfect match for corresponding original BLOT items. Four items from V3 (q4, q11, q14, q26) are on (or near) control lines. Estimates of 8 items (q6, q7, q8, q20, q21, q28, q30, q32) differ to a greater extent from the original items which measure the same operations.

Presented figures show that V2 is closer to the original BLOT than V3. More items from V2 have the same or very similar estimates as the original items than from V3. This is in accordance with the data regarding the mean values for three versions of BLOT (Table 1).

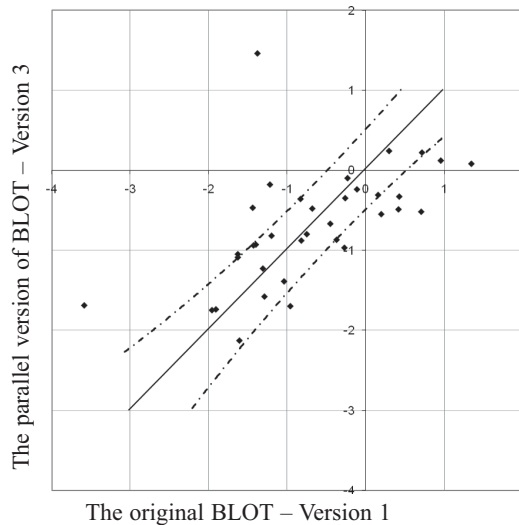


Figure 3. Plotting the original BLOT against Version 3

Rasch measurement showed that for 19 original items **both** parallel items maintained their invariance. For 14 items **one** of the parallel items had similar parameters. Only 2 items did not remain invariant in the parallel tests (q11 and q14).

Since the logical structure and format of q11 were preserved in parallel items, like it was illustrated by the Example 1, it is hard to explain existing difference in their estimates ($V1=0.43$, $V2=-0.25$, $V3=-0.33$). The original q14 (-0.96) is also more difficult than its parallel versions ($V2=-1.58$, $V3=-1.70$), but in this case the parallel items do not have exactly the same meaning as the original item (Example 3). From two parallel items one can easily conclude that one person will become a winner eventually (every week, or every month) and, according to that fact, to estimate the chances of winning. However, in the original item it is not that obvious that every type of raffle will have a winner and maybe that makes this item more difficult than the parallel items.

Example 3:

Item 14, original BLOT (only 2 alternatives are presented to understand the meaning of the item): *A man buys a raffle ticket in 4 different raffles each week. Which raffle does he have the best chance of winning?*

- a raffle with 50 tickets sold.
- a raffle with 10 tickets sold.

Item 14, Version 2: *A man takes part in the competition in which every week one car is the main price. Which week does he have the best chance of winning the car?*

- (a) the first week with 200 tickets sold.
- (b) the second week with 60 tickets sold.

Item 14, Version 3: *A man takes part in the competition in which every month the main prize is a trip. Which month does he have the best chance of winning the trip?*

(a) in January with 100 tickets sold.

(b) in February with 40 tickets sold.

DISCUSSION

The main goal of this research was to test two new versions of BLOT in Serbian and to discover whether they can be considered as parallel versions of the original test. The result show that majority of items, in both new constructed versions, do not differ from the original items regarding their estimates. However, the number of items that fit invariance model was bigger in Version 2 (28 out of 35, only 2 items strongly misfit the model) than in Version 3 (24 out of 35, 8 with misfit estimates).

Table 2 represents information summary of parallel items: their logical form, content linkage with other items, particular problems occurring during the construction, and the information about the comparison between parallel and original items estimates.

Table 2. *Parallel items summary*

Q	Operation	linked with the other(s) item(s)	Version 2		Version 3	
			problems in construction	estimate significantly different from the original item	problems in construction	estimate significantly different from the original item
1	SCHEMA: Mechanical Equilibrium	q2		√	c + 1	
2	SCHEMA: Mechanical Equilibrium	q1		Yes, but close to the control lines	c + 1	
3	Implication	No				
4	Incompatibility	No				Yes, but close to the control lines
5	Multiplicative Compensation	No				
6	SCHEMA: Correlations	q7, q8	c + 1		c + 1	√
7	SCHEMA: Correlations	q6, q8	c + 1		c + 1	√
8	SCHEMA: Correlations	q6, q7	c + 1		c + 1	√
9	Conjunction	No				
10	Disjunction	No				

11	Conjunctive Negation	No		Yes, but close to the control lines		Yes, but close to the control lines
12	Affirmation of p	No				
13	Reciprocal Exclusion	No				
14	Probability	No	f	Yes, but close to the control lines	f	Yes, but close to the control lines
15	Reciprocal Implication	No				
16	SCHEMA: INRC, Reciprocal (to negate Identity)	q17-q19	c + l + f		c + l + f	
17	SCHEMA: INRC, Identity (to negate Reciprocal)	q16, q18, q19	c + l + f	Yes, but close to the control lines	c + l + f	
18	SCHEMA: INRC, Negation (to negate Correlative)	q16, q17, q19	c + l + f	Yes, but close to the control lines	c + l + f	
19	SCHEMA: INRC, Reciprocal (to cause disequilibrium)	q16-q18	c + l + f		c + l + f	
20	SCHEMA: INRC, Negation (to cause disequilibrium)	q21-q23				√
21	SCHEMA: INRC, Correlative & Negation → equilibrium	q20, q22, q23				√
22	SCHEMA: INRC, Reciprocal & Negation → disequilibrium	q20, q21, q23		√		
23	SCHEMA: INRC, Correlative & Identity → disequilibrium	q20-q22				
24	SCHEMA: Co-ordination of 2 system of reference – Correlative & Identity	No			c	

25	Complete Negation	No			
26	Complete Affirmation	No			Yes, but close to the control lines
27	Negation of p	No			
28	Non-implication	No			√
29	Affirmation of q	No			
30	Equivalence	No	c	c	√
31	Negation of q	No			
32	Negation of Reciprocal Implication	No			√
33	SCHEMA: Probability	No	c	c	
34	SCHEMA: Co-ordination of 2 systems of reference – Reciprocal (to negate Identity)	q35			
35	SCHEMA: Co-ordination of 2 systems of reference (to negate Negation)	q34			

Legend: c – Problems in finding a phenomenon similar to the one from the *content* of the original item.

l – Problems in fitting the new content in the *linked* items

f – Found phenomena could not precisely fit the *format* of the original item

√ – Big difference between the estimates of the parallel item and the original.

The data about items estimates show that vast majority of original items have at least one parallel item, more than half of them have two parallel items and only two items are left without “twin” item, since q11 and q14 from both parallel versions have different estimates from the original items. Fortunately, the estimates of two misfitted items in both versions are close to the confidence band, which means that the difference between their difficulty and the difficulty of the original item is not large. In Table 2 one can notice that problems in items construction sometimes were followed by the estimate difference between a particular item and corresponding original item (q14, q17, q18 from V2; q6-q8, q14, q30 from V3). Such cases reveal potential reasons of found discrepancies and they can direct work on further development of parallel items. But, sometimes the difference in estimates appeared without previous construction problems which suggested that it is important to get back to the content of each misfitted item and try to find the difference in comparison to the content of the original item, as we did in the case of q11 and q14. The biggest difference (q20, V3) was the effect of an accidental replacement of the word *less* with word *more* which led to a completely different logical form of this parallel item. For some

other discrepancies in items' estimates it was not easy to find an explanation because constructed items preserved the format of the original items (q22 from V2; q26 and q32 from V3), as it was the case with q11. However, for majority of misfitted items it was possible to determine particular points that could be regarded as potential sources of differences. Our analysis has discovered certain patterns. Some items differed from original in the text length (q1, V2) which could make them more difficult in the terms of the information processing. The other discrepancies could be related to the different meaning of the original and parallel items, and they can be ascribed to different kind of phenomena or to a different relationship between events mentioned in the items' content. Even though in the construction of parallel items we intended to make minimal interventions trying to replace just one term with another wherever possible, those substitutes were not always addressing the phenomena of the same nature. For example, an absolute phenomenon was replaced with a continuous one, (q4, V3) or a phenomenon which can be changed was substituted by a phenomenon which cannot be changed (q30, V3).

Although horizontal decalage on the formal operation stage was not the central problem of this research, the question regarding the influence of changed tasks content on students' performances can be raised. In that respect, the previously mentioned item 11, whose estimates vary across all three versions of BLOT, is particularly interesting. This item was minimally changed in comparison to the original item. Items like that are the simplest case for analysis because different students' performances on them can be the result only of changes in single terms/concepts, since logical structure and format of the original item remained constant. In order to answer the question about the relationship between form and content it is very important to define precisely our understanding of content because it can represent different aspects which were not recognized and sufficiently differentiated in the mentioned discussions addressing the horizontal decalage phenomenon. The case of q11 content can be related to the type of decalage which Chapmen (1988) called "procedural decalage" when discussing concepts of stage, structure and developmental synchronicity in the context of Piaget's theory. He referred to the fact that the different versions of the same task are nevertheless solved by subjects at different ages and related that to the competence-performance distinction.

CONCLUSION

Presented data lead to the conclusion that Version 2 is closer to the original BLOT than to Version 3, and parameters of Version 2 allow us to regard it as a parallel version which needs some additional work in improving a few items. The relevance of this research is not just the usage of items with good psychometric characteristics in the research of peers' interaction on formal operation development but also in the fact that we consider this examination as the first step in the development of parallel versions of BLOT

which can be used in different researches. Furthermore, we are planning to translate Serbian parallel items into English and to conduct a research with an English speaking group of adolescents in order to develop parallel versions of the test in English.

As we mentioned before, this research could be related to the important theoretical problem of the existence of horizontal decalage phenomenon on the stage of formal operations. The generalization of formal operations across different domains was reconsidered by Piaget (1972), but that issue was also addressed by many other authors interested in this period of development (Neimark, 1975, Kuhn & Brannock, 1977, Wason, 1977, Chapman, 1988, Overton, Ward, Noveck, Black, & O'Brien, 1987). The main question is: Can this type of decalage be considered as in accordance with Piaget's theory, especially with his concept of formal operations? We can say that this kind of data does not contradict Piaget's theory, having in mind that he was interested in the universal course of development, i.e. structural invariants and their formal characteristics, and not in the individual subject, individual differences or in the specific context in which the structures are manifested. It is often emphasized that theoretical synchronicity does not imply the empirical one, and that Piaget always talked about the former not the latter (Chapman, 1988; Lourenco & Machado, 1996; Baucal & Stepanović, 1999). According to Bond (1995) it is a naïve point of view that competence in formal operational thinking requires an immediate transfer to all performance situations. On the other hand, Gray (1990) emphasizes the concept of pseudo-necessity, mentioned by Furth in his famous book *Piaget and Knowledge* (1969), as a more adequate explanation of the discussed problem. Gray argues that sometimes experiences which contradict an adaptational structure and usually lead to a developmental change are not acknowledged as such because they are considered impossibility by the adaptations. The resistance of existing adaptational structures to a new kind of adaptation is explained by pseudo-necessity, i.e. an adaptationally created impossibility. No matter how we explain the horizontal decalage phenomenon on formal operational stage, it is true that form-content interaction is not adequately explicated in *GLT* and we can consider different aspects of the horizontal decalage phenomenon. Because of that, it is very important to distinguish those aspects and to organize researches which will deal with these issues and collect empirical data in order to investigate different performance aspects of the Piagetian competence model.

Acknowledgement

The paper is a part of the project "Identification, measurement and development of cognitive and emotional competences important for a society oriented toward European integrations" (No. 179018) being done by Institute of Psychology, University of Belgrade, and funded by Ministry of Science, Republic of Serbia.

REFERENCES

- Bart, W. M. (1971–1972). The Factor Structure of Formal Operations. *British Journal of Educational Psychology*, 41/42, 70–77.
- Baucal, A., & Stepanović, I., (1999). The Horizontal Decalage Hypothesis: An Empirical Evaluation. *The Genetic Epistemologist*, 27(2), 2–10.
- Bond, T. G. (1976). *The development, validation and use of a test to assess Piaget's formal stage of logical operations*. Honours thesis, James Cook University of North Queensland.
- Bond, T. G. (1978). Propositional logic as a model for adolescent intelligence – additional considerations. *Interchange*, 9, 93–100.
- Bond, T. G. (1980). The psychological link across formal operations. *Science Education*, 64, 113–117.
- Bond, T. G. (1989). An investigation of the scaling of Piagetian formal operations. In P. Adey (Ed.), *Adolescent development and school science* (pp. 334–341). New York: Falmer Press.
- Bond, T. G. (1995). Piaget and measurement II: Empirical validation of the Piagetian model. *Archives de Psychologie*, 63, 155–185.
- Bond, T. G. (1997). Measuring development: Examples from Piaget's theory. In L. Smith et al. (Eds.), *Piaget, Vygotsky and beyond* (pp. 127–138). London: Routledge.
- Bond, T. G. (2010). Comparing décalage and development with cognitive developmental tests. *Journal of Applied Measurement*, 11(2), 158–171.
- Bond, T. G., & Bunting, E. (1995). Piaget and measurement III: Reassessing the methode clinique. *Archives de Psychologie*, 63, 231–255.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Bond T. G., & Jackson, I. (1991). The GOU protocol revisited: A Piagetian contextualization of critique. *Archives de Psychologie*, 59, 31–53.
- Brainerd, C. J. (1976). On the validity of propositional logic as a model for adolescent intelligence. *Interchange*, 7, 40–45.
- Chapman, M. (1988). *Constructive evolution: Origins and development of Piaget's thought*. Cambridge: Cambridge University Press.
- Christiansson, D. J. (1983). *An investigation of the relationship between cognitive developmental stage and quantitative skills in college students*. Doctoral dissertation, University of Pacific.
- Ennis, R. H. (1978). Conceptualization of Children's Logical Competence: Piaget's Propositional Logic and an Alternative Proposal. In: L. S. Siegel and C. J. Brainerd (Ed.) *Alternatives to Piaget: Critical Essays on the Theory*. Ontario, Canada: Academic Press, INC.
- Furth, H. (1969). *Piaget and Knowledge: Theoretical Foundations*. NJ: Prentice Hall.
- Goswami, U. (2001). Cognitive Development: No Stages Please-We're British. *British Journal of Psychology*, 92, 257–277.
- Gray, W. M. (1976a). *How is your logic?* (Experimental ed., Form A). Boulder, CO: Biological Sciences Curriculum Study.
- Gray, W. M. (1976b). *How is your logic?* (Experimental ed., Form B). Boulder, CO: Biological Sciences Curriculum Study.
- Gray, W. M. (1990). Formal Operational Thought. In W. F. Overton (Ed.), *Reasoning, Necessity and Logic: Developmental Perspectives* (pp. 227–253). New York: Hillsdale.
- Gruber, H. E., & Vonèche, J. J. (1995). The preadolescent and the propositional operations: introductory notes. In H. E. Gruber & J. J. Vonèche (Eds.), *The essential Piaget: An interpretative reference and guide* (pp. 394–398). Northvale, New Jersey: Jason Aronson.
- Inhelder, B. (1969). Some aspects of Piaget's approach to cognition. In H. Furth (Ed.), *Piaget and knowledge: Theoretical foundations* (pp. 22–42). Englewood Cliffs, NJ: Prentice-Hall.

- Inhelder, B. (1989). Autobiography. In G. Lindzey (Ed.), *A history of psychology in autobiography*. Vol. 8 (pp. 208–243). Stanford, CA: Stanford University Press.
- Inhelder, B., & Piaget, J. (1958). *Growth of logical thinking from childhood to adolescence*. New York: Basic Books.
- Kuhn, D., & Brannock, J. (1977). Development of the isolation of variables scheme in experimental and “natural experiment” contexts. *Developmental Psychology*, *13*, 9–14.
- Kuhn, D., Ho, V., & Adams, C. (1979). Formal Reasoning among Pre- and Late Adolescents. *Child Development*, *50*, 1128–1135.
- Kuhn, D., Langer, J., Kohlberg, L., & Haan, N. (1977). The development of formal operations in logical and moral judgment. *Genetic Psychology Monographs*, *95*, 97–188.
- Lawson, A. E., Karplus, R., & Adi, H. (1978). The acquisition of propositional logic and formal operational schemata during the secondary school years. *Journal of Research in Science Teaching*, *15*, 465–478.
- Lourenco, O., & Machado, A. (1996). In Defense of Piaget’s Theory: A Reply to 10 Common Criticisms. *Psychological Review*, *103*(1), 143–164.
- Lovell, K. (1961). A follow up study of Inhelder and Piaget’s *The growth of logical thinking*. *British Journal of Educational Psychology*, *52*, 143–153.
- Martorano, S. C. (1977). A developmental analysis of performance on Piaget’s formal operations tasks. *Developmental Psychology*, *13*, 666–672.
- Morley, K. A. (1979). *The development of spatial ordering concepts in geography at the secondary school level*. Honours thesis, James Cook University of North Queensland.
- Mwamwenda, T. S. (1999). Undergraduate and Graduate Students’ Combinatorial Reasoning and Formal Operations. *The Journal of Genetic Psychology*, *160*(4), 503–506.
- Neimark, E. D. (1975). Longitudinal Development of Formal Operations Thought. *Genetic Psychology Monographs*, *91*, 171–225
- Noelting, G. (1980). The development of proportional reasoning and the ratio concept. Part I – Differentiation of satges. *Educational Studies in Mathematics*, *11*, 217–253.
- Overton, W. F., Ward, S. L., Noveck, I. A., Black, J., & O’Brien, D. P. (1987). Form and Content in the Development of Deductive Reasoning. *Developmental Psychology*, *21*, 692–701
- Piaget, J. (1953). *Logic and psychology*. Manchester: Manchester University Press.
- Piaget, J. (1964). *Judgment and reasoning in the child*. Paterson, NJ: Littlefield, Adams.
- Piaget, J. (1972). Intellectual Evolution from Adolescence to Adulthood. *Human Development*, *2*, 1–12.
- Piaget, J., & Inhelder, B. (1969). *Psychology of the child*. London: Routledge and Kegan Paul.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.
- Rowell J. A., & Hoffman, P. J. (1975). Group tests for distinguishing formal from concrete thinkers. *Journal of Research in Science Teaching*, *12*(2), 157–164.
- Shayer, M. (1978). *A test of the validity of Piaget’s construct of formal operational thinking*. Doctoral dissertation, University of London.
- Shayer, M. (1979). Has Piaget’s construct of of formal operational thinking any utility? *British Journal of Educational Psychology*, *49*, 265–276.
- Shayer, M., & Adey, P. (1981). *Towards a science teaching*. London: Heinemann.
- Smith, L., & Knight, P. (1992). Adolescent reasoning tests with history content. *Archives de Psychologie*, *60*, 225–242.
- Stepanović, I. (2004a). Istraživanje formalno-operacionalnog mišljenja na uzrastu 14–19 godina [The investigation of formal operational thinking on the age 14–19]. *Psihologija*, *37*(2), 163–181.

- Stepanović, I. (2004b). Formalne operacije: Pijaževov koncept, istraživanja i najvažnije kritike [Formal operations: Piaget's concept, reserch and main critiques]. *Psihologija*, 37(3), 311–334.
- Stepanović, I. (2010). *The role of asymmetrical peers' interaction in the development of formal operational thinking*. Doctoral thesis, University of Belgrade.
- Ward, J. (1972). The saga of Buch and Slim. *British Journal of Educational Psychology*, 42, 267–289.
- Wason, P. C. (1977). The Theory of Formal Operations — A Critique. In B. A. Geber (Ed.), *Piaget and Knowing, Studies in Genetic Epistemology* (pp. 119–135). London: Routledge & Kegan Paul.
- Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: MESA Press.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.
- Wylam, H., & Shayer, M. (1978). *CSMS Science Reasoning Tasks*. Windsor: NFER.