AN ANALYSIS OF 24 SYLLOGISMS WITH ABSTRACT-SYMBOLIC CONTENT IN WHAT RESPECTS THEIR CAPACITY TO INDICATE THE ABILITY FOR THE CATEGORICAL SYLLOGISTIC REASONING

Lucia E. Faiciuc

Institutul de Istorie „George Barițiu", din Cluj-Napoca
Departamentul de Cercetări Socio-Umane

Abstract. *In the psychometric instruments that measure cognitive aptitudes like general intelligence, logical thinking, or formal reasoning, there are included, sometimes, some few categorical syllogisms, but why those particular syllogisms are selected is not very clear. On the other hand, in the empirical studies in which the theoretical models of the categorical syllogistic reasoning are tested, there are used syllogistic tasks in which tens of categorical syllogisms are included, as many as possible from the total number of 64 possible syllogisms (for the case in which only the two premises of a syllogism are taken into consideration). From pragmatic reasons (for example, administration time of a psychometric instrument), it would be desirable that, from those syllogistic tasks with a high number of categorical syllogisms to be selected for a psychometric purpose a small number of categorical syllogisms that have the highest potential to discriminate between the persons with a high ability for categorical syllogistic reasoning and the ones with a low ability in this respect, providing a more solid empirical base for the choice of the categorical syllogisms included in the cognitive tests. The present research is intended to be a contribution in that direction, an undertaking that, to my knowledge, lacks from the psychological literature. The research was based on the data obtained by me in a yet unpublished series of empirical studies, on a total number of 323 participants, having the same experimental design, through which there were investigated the cognitive processes involved in the syllogistic reasoning. In those studies, a syllogistic task with 24 categorical syllogisms with an abstract symbolic content, with two versions with a different linguistic format was used. In this research, from these 24 categorical syllogisms, there were selected a small number of syllogisms (five) as having the highest discriminative potential, based on the way they were solved by two groups of participants taken out from the entire sample as having the highest, and, respectively, the lowest stable performance at this syllogistic task in its two versions. The majority of the selected syllogisms (three) were invalid ones, with two negative premises. Future studies are needed in order to investigate the measure in which the selected categorical syllogisms have indeed a predictive value for the ability to reason syllogistically, in particular, or even for the formal thinking, or for intelligence, in general.*

Keywords: *categorical syllogism, ability to reason syllogistically, psychometric measurement of the formal thinking.*

An. Inst. de Ist. „G. Barițiu" din Cluj-Napoca, Series Humanistica, tom. XV, 2017, p. 125–146

Categorical syllogisms are a kind of deductive inferences through which a conclusion is drawn in a formally way as logically necessary about the relationship between categories and other categories, their members, or their properties based on the information given in two judgments, named premises, in which it is stated also information about the relationship between categories and other categories, their members, or their properties. The most famous illustration of a categorical syllogism is the following one:

All men are mortal.
Socrates is a man.
Therefore Socrates is mortal.

In a categorical syllogism, there are three syllogistic terms, two are named extreme terms, and one is named middle term. Those syllogistic terms represent categories, members of categories, or defining properties for categories. The conclusion of a categorical syllogism determines the relationship between the two extreme terms based on the information from the two premises, each of them stating the relationship of one of the two extreme terms with the middle term. Syllogistic judgments (the conclusion and the two premises) are of four types, depending on their quantity, i.e. the kind of quantification of the syllogistic terms that have the role of logical subject in those judgments (universal or particular) and on their quality, i.e. the affirmative or negative nature of those judgments. So, a universal affirmative judgment (for example, in an abstract version, in which the syllogistic terms are replaced with letters as symbols: "All S are P.") is traditionally symbolised with the letter A (or, if the syllogistic terms are included, as "SaP"). A particular affirmative judgment ("Some S are P.") is traditionally symbolised with the letter I (or, if the syllogistic terms are included, as "SiP"). A universal negative judgment ("All S are not P.") is traditionally symbolised with the letter E (or, if the syllogistic terms are included, as "SeP"), and a particular negative judgment ("Some S are not P.") with the letter O (or, if the syllogistic terms are included, as "SoP"). There are several types of categorical syllogisms, based on the particular combination of types of judgments of their premises, that combination defining the mood of a categorical syllogism. The position of the middle term in the two syllogistic premises is the one that defines the figure of a syllogism. Symbolizing the two extreme terms with S, and, respectively, P, and the middle term with M, there are four types of syllogistic figures, as follows:

Figure 1:
Premise 1: M–P
Premise 2: S–M
Conclusion: S–P

Figure 2:
Premise 1: P–M
Premise 2: S–M
Conclusion: S–P

Figure 3:
Premise 1: M–P
Premise 2: M–S
Conclusion: S–P

Figure 4:
Premise 1: P–M
Premise 2: M–S
Conclusion: S–P

Based on its mood and its figure, a particular categorical syllogism is traditionally symbolized by mentioning the corresponding symbols for the type of judgments of its premises, then the number of its figure, and, finally, the corresponding symbol for the type of judgment of its conclusion, or, if there is no valid conclusion, the symbol of the letter N for this type of answer. For example, the easiest and, maybe, the most frequently encountered categorical syllogism is AA1A, i.e. the one for which the two premises are universal affirmative judgments, and the conclusion is also a universal affirmative judgment. An example of this type of categorical syllogism would be:

All men are mortal. (MaP)
All Greeks are men. (SaM)
Therefore: All Greeks are mortal. (SaP)

Taking into consideration only the possible combinations of the four types of judgment for the two syllogistic premises and their combinations with the four syllogistic figures, there are possible 64 types of categorical syllogisms. Of these 64 possible categorical syllogisms, 18 are valid syllogisms (four of figure 1, four of figure 2, five of figure 3, and five of figure 4), i.e. categorical syllogisms for which a logically necessary conclusion can be drawn based on the information from the premises and the remaining ones are invalid categorical syllogisms, i.e. syllogisms for which no valid, logically necessary, conclusion can be drawn based on the information from the premises.

Although the categorical syllogistic reasoning, as verbal reasoning, is frequently encountered in the day to day thinking, it is rarely included in the psychometric tests that measure cognitive abilities or general intelligence, this psychometric instruments comprising mainly tasks of inductive reasoning (series completion, categorization) and of analogical reasoning, in the most cases as nonverbal tasks and sometimes as verbal ones. Nevertheless, as it is noted by Lohman (2005), the results obtained with psychometric instruments that use tasks of nonverbal reasoning, as it is the case with Raven's Progressive Matrices, are poorer predictors for the ability to learn in a real environment or for the academic performance than the results obtained with psychometric instruments that use tasks of verbal reasoning. Furthermore, Lohman (2005) shows that the sources of difficulty and the cognitive processes involved in the categorical syllogistic reasoning are, in part, different than the ones involved in the inductive reasoning or other deductive reasoning tasks. As a consequence, the psychometric instruments that intend to measure the reasoning ability or logical thinking should include also a task of categorical syllogistic reasoning. To the measure in which these psychometric instruments would want to measure formal-abstract thinking, the categorical syllogisms included in them should have an abstract content, with syllogistic terms represented through symbols or pseudowords (Stanovich, West & Toplak, 2016). The importance of the abstract categorical syllogistic reasoning as a predictor for the general intelligence and for the usual life has been revealed through a series of relatively recent studies. For example, Okamoto, Leighton &

Cor (2008) note that categorical syllogisms are used in the academic activity in order to determine the truth of some statements, in order to find an interpretation for the relationships among various statements, in order to determine if a proposed conclusion is valid or not, or, in the day to day thinking, in order to determine the validity of some current presuppositions, to understand the social rules, or to critically analyze the information received from various sources, like the one encountered in mass-media. Rayside & Kontogiannis (2001) show the importance of the categorical syllogisms for the IT domain, noting that the categorical syllogistic reasoning has a similar structure to the one of the object-oriented programming languages, and, therefore, a higher ability for the categorical syllogistic reasoning would be able to facilitate the learning of such programming languages. Shikishima, Hiraishi, Yamagata, Sugimoto, Takemura, Ozaki, ... & Ando (2009), in a study on a very large sample of monozygotic twins from Japan, which was aimed to investigate the heritability of the general intelligence (g factor) and of the ability to reason syllogistically, obtained data that indicate that the ability for categorical syllogistic reasoning has a high loading on the g factor, as high as other cognitive abilities through which they measured general intelligence (spatial and verbal abilities). The conclusion of the above-mentioned authors based on their data was that solving categorical syllogisms involves a comprehensive general ability that integrates several aspects of the human intelligence and that this type of syllogisms has symbolized human intelligence in philosophy, beginning with Aristotle, with good reason, because it represents one of the best markers of g from a psychological and behavioural genetics point of view.

The importance of the categorical syllogism is frequently underestimated, being ignored, because the way it is encountered in the usual life, i.e. in a shortened, verbally compressed, form, in which one of its two premises or sometimes even its conclusion are implicit (Kirby & Goodpaster, 2007, apud Okamoto *et al*., 2008), a different form than the one encountered in the handbooks for logic. An example of such an implicit categorical syllogism (named enthymeme) is given by Okamoto *et al*. (2008): "Michael Jean is the General Governor of Canada and, therefore, he deserves respect". The complete, explicit, corresponding syllogism would be: "All General Governors of Canada deserve respect. Michael Jean is the General Governor of Canada. Therefore, he is a person who deserves respect." Such usual categorical syllogisms are realized very fast and, because of that, most of the time, they are not even consciously acknowledged by the speakers.

On the one hand, in the past, there have been a few psychometric instruments for measuring cognitive abilities that include a relatively small number of categorical syllogisms, taking into account the importance of the categorical syllogisms. Examples of such psychometric instruments are the False Premises test, elaborated by Thurstone in 1938 (apud French, Ekstrom & Price, 1963), Logical reasoning test, elaborated by Guilford in 1954 (apud French et al., 1963), Morgan Test of Logical Reasoning elaborated in 1955 (apud McCoy, 1966), and, more recently, in Romania, the Deductive Reasoning subscale of the Analytic Reasoning

scale of the CAS++ instrument (Miclea *et al*., 2009), this instrument comprising three valid categorical syllogisms with a meaningful concrete content (EA1E, AO2O, IA3I), for which the option "no valid conclusion" was not provided. The criteria that were used in the selection of the categorical syllogisms for such instruments are not clear, lacking a solid theoretical and empirical support. Currently, because of the desire to construct cognitive ability tests that are as less as possible influenced by cultural and educational factors, there is a tendency that the categorical syllogisms tasks to be completely excluded from the psychometric instruments because of their verbal content or restrained to a very small number of categorical syllogisms, for which there lacks a solid theoretical or empirical ground.

On the other hand, in those studies in which the theoretical models of the categorical syllogistic reasoning processes were investigated, there have been used reasoning tasks that included as many categorical syllogisms as possible from the total number of 64 categorical syllogisms (usually, tens of them). It would be preferable, from pragmatic reasons (for example, the completion time), that, for those instruments that have a psychometric goal, from this number of 64 possible categorical syllogisms, only a few to be selected, in an empirically grounded way, i.e. only those categorical syllogisms that would allow the best discrimination between those persons who have a high ability for the categorical syllogistic reasoning and those persons with a low ability in this respect. The present research is intended to be a contribution for the completion of such a desideratum, an undertaking that, to my knowledge, is absent from the psychological literature. In order to accomplish that, I used data obtained by me in a series of yet unpublished empirical studies, with the same experimental design, in which there were investigated the processes involved in the syllogistic reasoning on a total number of 323 participants. In those empirical studies, there was used a syllogistic task with 24 categorical syllogisms with an abstract-symbolic content with two versions, corresponding to two different linguistic formats in which that content was formulated: a traditional format, i.e. the abstract symbolic expression encountered generally in the logic handbooks, in which the syllogistic terms are specified by using only the symbols of some letters, and a new format, through which there was specified also the logical status and meaning of the syllogistic terms symbolized with letters and of the syllogistic relationships of the syllogistic judgments. In these studies, of the total of 64 possible categorical syllogisms, only 24 were selected based on pragmatic and theoretical reasons. In the first place, all categorical syllogisms with the syllogistic figure 4 (16 syllogisms) were excluded, because they are similar with those with the syllogistic figure 1, as they differ mainly only by the order of their premises. Then, for each of the three remaining syllogistic figures, there were selected an equal number of categorical syllogisms (four valid, and four invalid), with the same combinations of types of premises (categorized based on their mood). Because, in the empirical studies used in this research it was important that the number of the valid categorical syllogisms to be equal with the number of the invalid categorical syllogisms, there were selected

almost all valid syllogisms remained after the exclusion of the valid categorical syllogisms with the syllogistic figure 4 (with the exception of one valid syllogism: IA3I, as it is similar with AI3I, because of the symmetry of the third syllogistic figure), i.e. 12 valid categorical syllogisms with the syllogistic figures 1, 2, and 3, and, from the remaining 48 invalid syllogisms, 12 invalid syllogisms were selected, by excluding those invalid syllogisms having at least one particular premise and for which there is no valid syllogism with the same combination of premises in the other syllogistic figures. All empirical studies had the same experimental design, a factorial mixed design, with two experimental groups, testing the same hypotheses in what respects the processes involved in the syllogistic reasoning. The 24 syllogisms from the syllogistic task had the same order for all participants, the two versions of the syllogistic task (with a different linguistic format) being administered successively (without pause) to all participants, the two experimental groups differing only by the administration order of the two versions of the syllogistic task. The valid and invalid syllogisms were interspersed, alternating a valid with an invalid syllogism. Syllogisms of figure 1 one were presented first, then those of figure 2, and, finally, those of figure 3.

The present research was initiated based on the assumption that the ability for categorical syllogistic reasoning should be higher in the case of those participants who have a high total score for the syllogistic task in both of its administered versions, because a superior performance for the two successive administrations of the syllogistic task in its two versions would indicate stableness in time (although for a very short term, in this case) of the capacity to solve correctly categorical syllogisms, and, also, more importantly, stableness of this capacity against the linguistic format of the syllogisms, a linguistic format that would be able to facilitate or not the correct solving of the categorical syllogisms, as it was assumed in the general hypothesis of the empirical studies that generated the data used in this research (the new linguistic format was assumed to be able to facilitate the solving of the valid categorical syllogisms, whereas the traditional format was assumed to be neuter or, at most, to be able to facilitate the solving of the invalid categorical syllogisms, the unpublished results obtained in the empirical studies mentioned here tending to support those assumptions). Therefore, three performance groups were formed with a relatively equal number of participants, based on the value of the total score at the syllogistic task in its version administered in the first position (the cutting points were determined separately for each of the two experimental groups). For the following procedure, only the two groups with extreme (high, and, respectively, low) scores were taken into consideration. For the participants from these two groups, a difference score was computed by subtracting the score obtained for the syllogistic task at its second administration from the scored obtained for the syllogistic task at its first administration. Based on the value of this difference score, the participants from the two groups with extreme syllogistic performance were, again, divided into three groups with a relatively equal number of participants, having high, medium, or, respectively, small

difference scores (computed as mentioned above), separately for each of the two orders of administration of the two versions of the syllogistic task. In the group of participants assumed to potentially have a higher ability for categorical syllogistic reasoning, there were included only those participants who had a high score for the syllogistic task at its first administration, and medium or a low score for the difference between the total performance at the first and the second administration of the syllogistic task (the latter category indicating an even better performance at the syllogistic task at its second administration). Correspondingly, in the group of participants assumed to potentially have a lower ability for categorical syllogistic reasoning, there were included only those participants who had a low score for the syllogistic task at its first administration, and medium or high score for the difference between the total performance at the first and the second administration of the syllogistic task (the latter category indicating an even worsened performance at the syllogistic task at its second administration). Once the two groups of participants with a potentially high and, respectively, low ability for the categorical syllogistic reasoning were formed, based on the performance at the syllogistic task administered in the first position and on its stableness, assessed by taking into consideration the performance at the syllogistic task administered in the second position, there were analyzed all the 24 syllogisms in their traditional linguistic format, in order to indentify those syllogisms for which statistically significant differences occur between their solving in the two groups with an assumed different ability level to solve categorical syllogisms. Only the traditional linguistic format was chosen for this analysis, because it is the linguistic format that is usually used in the psychological literature, and because it is presumably a rather neuter format, which does not facilitate significantly the syllogistic performance as it was assumed in the case of the new linguistic format. This analysis was made in order to select a smaller number of syllogisms that, if included in future tests for the measurement of the formal thinking, would require a shorter administration time, and, in the same time, would allow a better discrimination between the persons with a higher ability for the categorical syllogistic reasoning and the ones with a lower ability in that respect.

METHOD

Participants: 323 participants with an average age of 19 years (247 girls, 74 boys, 2 without a declared gender), who were high school pupils (51 pupils) or students in the first and second year at the University of Fine Arts and Design from Cluj-Napoca (272 students). In the group with the first administration order of the two versions of the syllogistic task (traditional version first, new version second) there were 165 participants, and in the group with the second administration order of the two versions of the syllogistic task (new version first, traditional version second) there were 158 participants.

Instruments: A syllogistic task with 24 syllogisms, with two versions with a different linguistic format (see Appendix 1).

Procedure: The two successive versions of the syllogistic task were administered collectively, without time limit, and without pause between them. No practice syllogisms were given before administering the two versions of the syllogistic task or as part of them. The participants were distributed randomly into the two groups with a different administration order of the two versions of the syllogistic task. In the case of the high school pupils, the participants were distributed in the two groups by choosing at random that the pupils placed on the right side of the three rows of school desks from their classroom to be in first group, and the ones placed on the left side to be in the second group. In the case of the students, they were distributed at random in the first, and, respectively, in the second group, in accordance with their place in the rows of the lecture hall, in an alternative manner: the first student in the row in the first group, then the second one in the second group, the third one in the row in the first group, the fourth one in the second group and so on. The participants were required to write on the completion form of the both versions of the syllogistic task a pseudonym at their choice instead of their names, in order to secure the confidentiality of their result and to discourage the tendency to copy the answers from their neighbors. In order to be motivated supplementarily, the participants were told that, based on their pseudonym, they can find out, at their request, the obtained correctness score. A total correctness score was computed by adding all the correct answers for the 24 syllogisms.

RESULTS

A statistical significance threshold of $p = .05$ was chosen for this research. The reported results of the statistical tests used for the comparisons made in the following analyses were the ones for the two-tailed case, even though, for some of them, a one-tailed option would have been warranted, especially for the situation in which it is to be expected a better performance for each of the 24 syllogisms in the group with the highest total scores at the syllogistic task than in the group with the lowest total scores at that task.

By following the procedure described in the introductory section through which two groups were formed, based on their performance at the categorical syllogistic task used in this research and its stability, there was obtained a group with 48 participants (35 girls, 13 boys, one without a declared gender) with the highest stable level of categorical syllogistic performance and a group of 30 participants (20 girls, 9 boys and one without a declared gender) with the lowest stable level of categorical syllogistic performance. There was no statistically significant gender difference between the two groups. By applying the Mann-

Whitney test, a statistically significant age difference between the two groups ($U = 448.5$, $p = .010$, $N = 78$) was obtained, the participants from the group with a better syllogistic performance being younger (Mdn = 19 years) than those from the group with a worse syllogistic performance (Mdn = 20 years). No statistically significant difference occurred between the two groups in what respects the proportion of high school or undergraduate participants. Because the distribution of the total correctness score was significantly departed from the normal distribution in the two performance groups, a nonparametric test was used to compare their performance, in order to verify if the procedure through which the two groups were formed was effective. Appling the Mann-Whitney test, the total correctness score was indeed statistically significantly different between the two groups ($U = 0$, $p < .001$, $N = 78$, Mdn = 7, for the group with the lowest level of categorical syllogistic performance, Mdn = 14, for the group with the highest level of categorical syllogistic performance). In the group with the first administration order of the two versions of the syllogistic task, the group with the better syllogistic performance had 21 participants, and the group with the worse syllogistic performance had 23 participants. In the group with the second administration order of the two versions of the syllogistic task, the group with the better syllogistic performance had 27 participants, and the group with the worse syllogistic performance had 7 participants. Applying $\chi 2$ test with Yates' continuity correction, there was a statistically significant difference between the two groups with a different administration order of the two versions of the syllogistic task in what respects the percentage of their participants included in the group with the highest syllogistic performance, or, respectively, in the group with the lowest syllogistic performance: $\chi 2\ (1,\ 78) = 6.851$, $p = .009$. This difference was presumably caused mainly by the experimental manipulation in the empirical studies that generated the data used in this research, i.e. by the direct effect of the linguistic format and by the order effects influenced by that linguistic format effect. Taking into consideration this difference, the following data analysis will be made separately for the two groups with a different administration order of the two versions of the syllogistic task.

Based on the analysis realised by using the $\chi 2$ test with Yates' continuity correction (and, where the number of expected cases in a cell was less than five, using Fisher's exact test) in what respects the comparison of the frequency of the correct answers for each of the 24 syllogisms formulated with the traditional linguistic format in the two groups with different levels of categorical syllogistic performance, statistically significant results were obtained for the syllogisms from the Table 1, noted by using the international traditional symbolization system for categorical syllogisms (see the introductory section and the corresponding syllogisms in the Appendix 1). As mentioned above, the analysis was made separately for the two groups with a different administration order of the two versions of the syllogistic task. Order 1 is the one in which the syllogistic task was first administered in the traditional format, and then in the new format, in which the logical meaning of the syllogistic

Table 1

Syllogisms for which there were obtained statistically significant differences in what respects their correct solving between the two groups with a different level of categorical syllogistic performance

Administration order 1 for the two versions of the syllogistic task (the version in the traditional linguistic format followed by the version in the new linguistic format)	**Administration order 2 for the two versions of the syllogistic task (the version in the new linguistic format followed by the version in the traditional linguistic format)**
EE1N (41%) $\chi2(1,44) = 26.397$, $p = .000$	EE1N (35.4%) $\chi2(1,34) = 19.959$, $p = .000$ ($p = .000$, Fisher's exact test)
EE2N (41.4%) $\chi2(1,44) = 17.989$, $p = .000$	EE2N (34.8%) $\chi2(1,34) = 14.744$, $p = .000$ ($p = .000$, Fisher's exact test)
EE3N (48.5%) $\chi2\ (1,44) = 36.409$, $p = .000$	EE3N (41.1%) $\chi2\ (1,34) = 12.790$, $p = .000$ ($p = .000$, Fisher's exact test)
EO3N (37.6%) $\chi2\ (1,44) = 7.393$, $p = .007$	EO3N (30.4%) $\chi2\ (1,34) = 7.421$, $p = .006$ ($p = .002$, Fisher's exact test)
EI1O (47.9%) $\chi2(1,44) = 10.955$, $p = .001$	EI1O (62%) $\chi2(1,34) = 14.842$, $p = .000$ ($p = .000$, Fisher's exact test)
EA1E (73%) $\chi2(1,44) = 13.391$, $p = .000$	EA1E (72.2%) $\chi2(1,34) = .545$, $p = .460$ ($p = .206$, Fisher's exact test)
EO1N (25.5%) $\chi2(1,44) = 9.898$, $p = .002$	EO1N (17.7%) $\chi2(1,34) = 2.56$, $p = .110$ ($p = .069$, Fisher's exact test)
AE2E (69.7%) $\chi2(1,44) = 7.343$, $p = .007$	AE2E (60.1%) $\chi2(1,34) = 1.692$, $p = .193$ ($p = .151$, Fisher's exact test)
EO2N (27.3%) $\chi2\ (1,44) = 17.344$, $p = .000$	EO2N (21.5%) $\chi2\ (1,34) = 4.216$, $p = .069$ ($p = .069$, Fisher's exact test)
EA2E (64.3%) $\chi2(1,44) = .448$, $p = .503$	EA2E (57%) $\chi2(1,34) = 4.664$, $p = .031$ ($p = .020$, Fisher's exact test)

terms was specified. Order 2 was the one in which the syllogistic task was first administered in the new format, and then in the traditional format. In the Table 1, between the round brackets there were mentioned the percentage of the correct answers for those syllogisms for the entire group with the corresponding administration order. In the Table 1, but also in the Appendix 1, those syllogisms for which there were obtained values for the $\chi2$ test statistically significant for the both administration orders of the two versions of the syllogistic tasks have emboldened fonts (five syllogisms), and those having a value for the $\chi2$ test statistically significant only for one of the two above-mentioned administration

orders were marked with italics (five syllogisms). In Table 2, for each syllogism from the Table 1, it is presented the frequency of the correct answers (marked with number 1) and of the incorrect ones (marked with number 0) for the two syllogistic performance groups, separately for each administration order. Finally, in the Table 3, there were included the discrimination indices computed for each syllogism from the Table 1, by using the formula provided by Albu (1998) for the computation of a discrimination index of a test item, as a difference between the ratio of the correct answers in the group with a high syllogistic performance to the number of participants from that group and the ratio of the correct answers in the group with a low syllogistic performance to the number of participants of this group.

Table 2

The frequency of the correct and incorrect answers for each of the ten selected syllogisms, in the case of each administration order, for the two syllogistic performance groups

Syllogism	**Value of the answer (1: correct answer, 0: incorrect answer)**	**Administration order 1 for the two versions of the syllogistic task (the version in the traditional linguistic format followed by the version in the new linguistic format)**		**Administration order 2 for the two versions of the syllogistic task (the version in the new linguistic format followed by the version in the traditional linguistic format)**	
		Low performance group	**High performance group**	**Low performance group**	**High performance group**
EE1N	**0**	22	3	7	2
	1	1	18	0	25
EE2N	**0**	21	5	7	4
	1	2	16	0	11
EE3N	**0**	23	1	7	5
	1	0	20	0	22
EO3N	**0**	19	8	7	9
	1	4	13	0	18
EI1O	**0**	19	6	6	2
	1	4	15	1	25
EA1E	**0**	11	0	1	0
	1	12	21	6	27
EO1N	**0**	23	12	7	16
	1	0	9	0	11
AE2E	**0**	12	2	0	9
	1	11	19	7	18
EO2N	**0**	23	8	7	16
	1	0	13	0	11
EA2E	**0**	11	7	4	3
	1	12	14	3	24

Table 3

Discrimination indices for each of the ten selected syllogisms, in the case of each administration order

Syllogism	**Discrimination index in the case of the administration order 1 for the two versions of the syllogistic task (the version in the traditional linguistic format followed by the version in the new linguistic format)**	**Discrimination index in the case of the administration order 2 for the two versions of the syllogistic task (the version in the new linguistic format followed by the version in the traditional linguistic format)**
EE1N	0.81	0.92
EE2N	0.67	0.85
EE3N	0.95	0.81
EO3N	0.45	0.67
EI1O	0.54	0.78
EA1E	0.48	0.14
EO1N	0.43	0.40
AE2E	0.43	-0.33
EO2N	0.61	0.40
EA2E	0.14	0.46

DISCUSSION AND CONCLUSIONS

Based on the procedure used in this research, there were identified ten categorical syllogisms that were solved significantly different by the participants from the two groups with a different level of categorical syllogistic performance for the two administration orders of the two versions of the syllogistic task from the present research. Of these ten categorical syllogisms, five are encountered for the both administration orders. It can be assumed that these five would be the ones with the highest discriminative potential between those with a high ability for the categorical syllogistic reasoning and those with a low ability in this respect.

It can be noted that almost all categorical syllogisms identified as having a high discriminative potential have two negative premises, and no valid conclusion, and, the remaining categorical syllogisms have one negative premise and a negative conclusion. It could mean, on the one hand, that a possible characteristic of the people having a higher ability for the categorical syllogistic reasoning would be the explicit or implicit acknowledgment of the logical principle that from two negative premises it is logically impossible to infer a valid conclusion. On the other hand, the results for the categorical syllogisms with a high discriminative potential could indicate also the seriousness with which the syllogistic task was completed, i.e. the measure in which the syllogistic task was completed by guessing the answers based on chance or using simple heuristics (like choosing a universal affirmative conclusion for two universal negative premises, because the participants interpret the situation to be similar with a double negation one, or ignore the

negation altogether, because its processing would require a supplementary cognitive effort, or choosing a negative conclusion for two negative premises, based on a so called atmosphere heuristic, the A, E, and O conclusions being the incorrect answers that were selected by the participants for the syllogisms with the highest discriminative potential that have two negative premises).

There were categorical syllogisms with a relative medium level of the percentage of the correct answers (for example, AO2O: 65% and EI3O: 42%), similar with the ones identified to have a high discriminative potential, but which did not have a high discriminative potential between the two groups with a different level of syllogistic performance. This situation can be an indication that the difficulty of a syllogism, measured by the percentage of correct answers for it, might be a misleading index for the potential power of that categorical syllogism to discriminate between those with different levels of ability to reason syllogistically. It can be speculated that those categorical syllogisms that have a high frequency of correct answers, mostly those with valid conclusions, do not have a discriminative power between the two groups with a different level of performance for the syllogistic task used in this research because most of them are categorical syllogisms that are solved based on some well known and widespread inference schemata (encountered frequently during numerous argumentative experiences), which do not require the explicit or implicit knowledge of syllogistic principles or a high level of ability for formal thinking. On the other hand, there are valid, or invalid syllogisms with a medium or low frequency of correct answers that are similarly solved in the both groups with a different level of syllogistic performance because they can be easily confounded with those from the first category, i.e. those syllogisms with a high rate of correct answers. The categorical syllogisms with two negative premises, instead, are less likely to be confounded with the syllogisms from the first mentioned category by those persons for which the inference schemas of the syllogisms from this first category are well known and easily recognizable (presumably the persons that have a high ability for categorical syllogistic reasoning), as no categorical syllogism with a high rate of correct answers has two negative premises.

Of the valid categorical syllogisms from the syllogistic task used in this research, the only one that had a high potential discrimination power between the two performance groups was the syllogism EI1O, a syllogism which is from a syllogistic figure with an important cognitive function (see Faiciuc, 2009), presumably frequently encountered in the current argumentative experience, but, in the same time, it has a conclusion with a low degree of informativeness, a reason for which, this categorical syllogism might be less frequently encountered in the argumentative experience in comparison with other figure 1 valid categorical syllogisms (see Faiciuc, 2009). That is why, an inference schema for this syllogism would be familiar only for those persons that have an extensive argumentative experience and a higher ability to learn from that experience, whereas those with a low ability in this respect and a poor argumentative experience, would fail to

recognize the appropriate inference schema for this syllogism, giving an incorrect answer. Finally, the remaining category of valid categorical syllogisms that were indentified to be able to discriminate between the two syllogistic performance groups, but in a lower degree, are those having a universal negative conclusion, for which the pattern of the correct and incorrect answers for the two syllogistic performance groups was different than the one obtained for the five categorical syllogisms with the highest discriminative potential. In their case, the participants from the group with the highest performance at the categorical syllogistic task gave predominantly the correct answer, but the participants from the group with the lowest performance at the categorical syllogistic task did not give predominantly incorrect answers, as it was the case for the five categorical syllogisms with the highest discriminative potential, the percentage of the correct and incorrect answers in their case being rather balanced. It may be that the inference schemata for these categorical valid syllogisms that have presumably a higher cognitive function (see Faiciuc, 2009), even though not the highest, are well known to those persons with a high ability for the categorical syllogistic reasoning, but are somewhat familiar also for the persons with a low ability for the categorical syllogistic reasoning, as they may be frequently encountered in the day to day argumentative experience.

A first partial confirmation of the results obtained in the present research comes from a study of Galotti, Baron and Sabini (1986). They also compared the performance in a categorical syllogistic task of 16 valid and invalid syllogisms of the Figure 4 type (different than the ones used in this research) with a concrete nonsense content for two groups of undergraduate students, with a high vs. low syllogistic performance, named good (16 participants) vs. poor (15 participants) reasoners, selected based on a pre-test comprising eight categorical syllogisms. To those two groups, they added a so-called group of seven experts in logical thinking, i.e. students in Psychology and Computer Science that had some general training in logical thinking, but not an extensive specific training in the solving of categorical syllogisms. In their study, they report the percentage of correct answers for each of the 16 categorical syllogisms in the three groups, but the authors' analysis was oriented mainly on a more general performance comparison, i.e. on broader types of syllogisms. Their results indicated that the three groups differ largely in the performance for the invalid categorical syllogisms when the syllogistic task required specifically that the derived conclusion to be logically necessary, but not when it required only the inference of a possible conclusion. The interpretation given by the cited authors for this result was that the obtained difference in the first case cannot be completely explained by considering possible differences in the comprehension of the task between the three groups. Using the think aloud technique in order to find the sources of this difference discovered between the three groups, they "found some evidence for the existence of some deduction rules, apparently spontaneously discovered by some of our subjects" (p. 19), in the good reasoners group, especially the rules that lead to the "no valid conclusion" correct answer, as it is the rule corresponding to the logical principle that nothing follows

from two particular premises. Galotti *et al*. (1986) note that "novice reasoners sometimes spontaneously articulate such rules after only brief practice at the task" (p. 23). Analyzing the performance difference between the three groups from their research on three categories of invalid categorical syllogisms: those for which holds the logical principle that nothing follows from two negative premises (NEGATIVE rule), those for which holds the logical principle that nothing follows from two particular premises (SOME rule), and the remaining ones (OTHER), taking into consideration also the response times and the results obtained when additional response time was allowed, they conclude that: "Experts have good command of the SOME and NEGATIVE Rules, in particular the former and apply them quickly and effectively. Poor reasoners show little evidence of familiarity with either rule." (p. 22) and that "both good reasoners and expert benefit from additional time": good reasoners mostly for the invalid syllogisms with double negation and double particular premises, and experts for other invalid syllogisms, "suggesting that the expert superiority does not reside merely in the possession and facility with the two familiar rules." (Galotti *et al*., 1986, p. 22). Generalizing to other syllogistic figures than the one used by the cited authors, from their results it is to be inferred that the invalid categorical syllogisms with two negative premises should have a discriminative potential between the good and poor reasoners, as it was supported in the present research. In fact, in their study, Galotti *et al*. (1986) showed that for the syllogism EE4N, there was a performance difference between the three groups from their research: the percentage of the correct answers was for the expert group 100%, for the good reasoners, 94%, and for the poor reasoners, 33%.

A second partial confirmation of the data from the present research comes from a study elaborated by Shikishima, Yamagata, Hiraishi, Sugimoto, Murayama and Ando (2011), in which the performance in a syllogism-solving test (the BAROCO Short) was correlated with an intelligence test from Japan (Kyoto New NX) on a sample of 487 pairs of twins of several kinds, with ages between 13 and 18 years (with a mean age of 15.3 years). Of the five syllogisms of that syllogistic test, only one was a categorical syllogism with an abstract and belief-neutral content as in the present research: EI1O, and the performance in solving this syllogism correlated positively, in a statistically significant manner, with the intelligence test Kyoto New NX: $r = .39$. It can be noted that that categorical syllogism is the only valid categorical syllogism that was found in the present research to have a high discriminative potential between those with a high ability for the categorical syllogistic performance and those with a low ability in this respect.

Finally, a third partial confirmation for the results from this study is offered by some data obtained by me in an unpublished research in which the syllogistic task from this study in its traditional linguistic format was administered to a sample of 34 high school students (mean age: 16 years, 24 girls and 10 boys) together with the Raven Progressive Matrices IQ test in its standard version. Using the Mann-Whitney test, for each of the 24 syllogisms, a comparison was made between those who solve that syllogism correctly and those who solve it incorrectly in what

regards their score for the Raven test. Statistically significantly higher scores at the Raven test for those who had correct answers than the ones for those who had incorrect answers were obtained for the following eight syllogisms: EA1E ($U = 42.5$, $p = .002$), EE1N ($U = 45.5$, $p = .001$), AE2E ($U = 45.5$, $p = .009$), EA2E ($U = 10$, $p = .002$), EE2N ($U = 42.5$, $p = .002$), AA2N ($U = 50$, $p = .014$), AO2O ($U = 74.5$, $p = .021$), and AA3I ($U = 19.5$, $p = .029$). The rank-biserial correlation coefficients (one-tailed) between the Raven test score and the performance for each syllogism were statistically significant for the following syllogisms EA1E ($\rho = .541$, $p = .000$), EE1N ($\rho = .5597$, $p = .000$), AE2E ($\rho = .458$, $p = .003$), EA2E ($\rho = .532$, $p = .001$), EE2N ($\rho = .367$, $p = .016$), AA2N ($\rho = .427$, $p = .006$), AO2O ($\rho = .401$, $p = .009$), AA3I ($\rho = .379$, $p = .014$), AA1A ($\rho = .332$, $p = .027$), EI3O ($\rho = .294$, $p = .045$), and relatively close to the chosen statistical significance threshold ($p = .05$) for the syllogisms EO1N ($\rho = .275$, $p = .058$), EE3N ($\rho = .254$, $p = .074$). In all cases, the obtained relationship was a positive one. After dividing the sample into three relatively equal groups based on the performance at the Raven test, a comparison was made between the two extreme groups (one with high Raven test scores, with 12 participants, and one with low Raven test scores, having 9 participants) in what respects their performance for each syllogism from the syllogistic task. Using the Fisher's exact test (one-tailed), statistically significant results were obtained for EA1E ($p = .029$), EE1N ($p = .000$), AE2E ($p = .029$), EE2N ($p = .029$), AA2N ($p = .037$), and relatively close to the chosen statistical significance threshold for EA2E ($p = .063$) and EO1N ($p = .078$), the group with high Raven test scores solving correctly more frequently than the group with low Raven test scores the above-mentioned syllogisms. It can be noted that five of the syllogisms that were found to have a discriminative power in the present research in what regards the ability for the categorical syllogistic reasoning had significant relationships with the Raven scores or had the power to discriminate between those with high and, respectively, low Raven scores (with the exception of the syllogism EA2E, for which only a trend in this respect was obtained): EA1E, EE1N, EE2N, AE2E, EA2E. The remaining five syllogisms from those that were identified here as having discriminative power in what concerns the syllogistic reasoning were, in general, with only one exception, invalid syllogisms with two negative premises, especially those in which one premise is universal (E type) and the other one is a particular premise (O type): EE3N, EO3N, EI1O, EO1N, EO2N. For these syllogisms, no relationship with the Raven scores was obtained or only one close to the chosen statistical significance threshold (for EO1N and EE3N). The same syllogisms did not have discriminative potential in what respects the performance for the Raven test (with the exception of a trend in this respect for the EO1N syllogism). These results indicate, on the one hand, that it is less likely that the data from the present research to be obtained by chance, as long as some of the syllogisms for which there was a significant correctness difference between the two groups with a different level of syllogistic performance may have also a relationship

with the general intelligence factor, and that the syllogisms EE1N and EE2N may have the highest discriminative potential in both respects: syllogistic ability and general intelligence. On the other hand, they indicate also that the ability for the categorical syllogistic reasoning may involve a specific factor, besides the general intelligence factor, because a particular category of syllogisms (see above) had a discriminative potential in what regards the syllogistic ability level, but not in what regards the general intelligence level and, also, had no relationship with the general intelligence factor.

Comparing the categorical syllogisms identified in this research as having the highest discrimination index between the participants with a high syllogistic performance and those with a low syllogistic performance with the categorical syllogisms included in, to my knowledge, the only Romanian instrument currently used for the investigation of the ability for categorical syllogistic reasoning (CAS++), it can be noted that they are not the same, with the exception of the EA1E categorical syllogism, which, in this research, was not among those categorical syllogisms with the highest and stable discriminative potential. The three categorical syllogisms from CAS++ (see the introductory section) are exclusively valid syllogisms, which do not have in their answer options at least a suggestion that they might be invalid, by including the "no valid conclusion" as an answer option (increasing in this manner their difficulty, by decreasing the guessing likelihood), whereas in this research the categorical syllogisms with the highest discriminative potential were mostly invalid syllogisms. Based on the data from this research, it might be possible to improve such instruments that measure the ability for categorical syllogistic reasoning by including in them invalid categorical syllogisms with an abstract content of the type that was found in the present research to have the highest discriminative potential between those with a high ability for categorical syllogistic reasoning and those with a low ability in this respect.

Future studies would be needed, using other criteria than ones from this research, in order to investigate the measure in which the categorical syllogisms identified here to have a high discriminative potential can be really used to differentiate between the persons with a high vs. low ability for the categorical syllogistic reasoning, in particular, or the measure in which their solving would have a predictive value for the formal thinking ability, in general. In these future studies, the samples should be larger and more balanced in what respects the gender composition than in this research. The investigation should be extended also to participants of other ages and from various domains of activity, and to the remaining valid and invalid categorical syllogisms that were excluded from the present research (the figure 4 valid categorical syllogisms and the invalid syllogisms with two particular premises). It might be possible that the acknowledgment of the logical principle that from two particular premises it is not possible to infer a valid conclusion could have also a discriminative value in what respects the ability to reason syllogistically.

REFERENCES

Albu, M. (1998). *Construirea şi utilizarea testelor psihologice* [The construction and use of the psychological tests] Cluj: Clusium.

Faiciuc, L. E. (2009). *The dynamics of the mental representations in the deductive reasoning processes* (Unpublished PhD dissertation). Babeş-Bolyai University, Cluj-Napoca. Retrieved from http://www.history-cluj.ro/SU/cercet/Faiciuc_lucrari/PhD%20Thesis%20Lucia%20Faiciuc%20(extended%20version%20-%20full%20text).pdf

French, J. W., Ekstrom, R. B., & Price, L. A. (1963). *Manual for kit of reference tests for cognitive factors (revised 1963)*. Educational Testing Service Princeton, NJ.

Galotti, K. M., Baron, J., & Sabini, J. P. (1986). Individual differences in syllogistic reasoning: Deduction rules or mental models?. *Journal of Experimental Psychology: General, 115*(1), 16–25.

Lohman, D. F. (2005). Reasoning abilities. In R. J. Sternberg, J. Davidson, & J. Pretz (Eds.), *Cognition and intelligence: Identifying mechanisms of the mind* (pp. 225–250). New York: Cambridge University Press.

McCoy, F. K. (1966). *A revision of the Morgan Test of Logical Reasoning*. (master thesis). University of Richmond, Richmond. Retrieved from http://scholarship.richmond.edu/cgi/viewcontent.cgi?article=1726&context=masters-theses

Miclea, M. (Coord.), Porumb, M. (Coord.), Cotârlea, P. (Coord.), Albu, M. (Coord.), Bălaj, A., Benga, O., … Visu, G. (2009). CAS++. *Aptitudini Cognitive* [Cognitive Aptitudes]. Cluj-Napoca: Cognitrom & ASCR.

Okamoto, C., Leighton, J., and Cor, M. (2008). *The Role of Academic Confidence and Epistemological Beliefs on Syllogistic Reasoning Performance*. Centre for Research in Applied Measurement and Evaluation, University of Alberta. Retrieved from http://www.crame.ualberta.ca/files/LogicStudy.pdf

Rayside, D., & Kontogiannis, K. (2001, July). On the syllogistic structure of object-oriented programming. *In Proceedings of the 23rd international conference on Software engineering* (pp. 113–122). IEEE Computer Society.

Shikishima, C., Hiraishi, K., Yamagata, S., Sugimoto, Y., Takemura, R., Ozaki, K., ... & Ando, J. (2009). Is g an entity? A Japanese twin study using syllogisms and intelligence tests. *Intelligence*, *37*(3), 256–267.

Shikishima, C., Yamagata, S., Hiraishi, K., Sugimoto, Y., Murayama, K., & Ando, J. (2011). A simple syllogism-solving test: Empirical findings and implications for g research. *Intelligence, 39*(2), 89–99.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2016). *The Rationality Quotient: Toward a Test of Rational Thinking*. MIT Press.

APPENDIX 1

The syllogistic task in the **traditional linguistic format**

Response options:
A. All **S** are **P**.
B. All **S** are **not P**.
C. Some **S** are **P**.
D. Some **S** are **not P**.
E. It is **not** possible to derive a necessary logical conclusion.

1. All **M** are **P**. All **S** are **M**.	AA1
2. All **M** are **P**. All **S** are **not M**.	AE1
3. All **M** are **P**. Some **S** are **M**.	AI1
4. All **M** are **P**. Some **S** are **not M**.	AO1
5. All **M** are **not P**. All **S** are **M**.	*EA1*
6. All **M** are **not P**. All **S** are **not M**.	**EE1**
7. All **M** are **not P**. Some **S** are **M**.	**EI1**
8. All **M** are **not P**. Some **S** are **not M**.	*EO1*
9. All **P** are **M**. All **S** are **M**.	AA2
10. All **P** are **M**. All **S** are **not M**.	*AE2*
11. All **P** are **M**. Some **S** are **M**.	AI2
12. All **P** are **M**. Some **S** are **not M**.	AO2
13. All **P** are **not M**. All **S** are **M**.	*EA2*
14. All **P** are **not M**. All **S** are **not M**.	**EE2**
15. All **P** are **not M**. Some **S** are **M**.	EI2

16. All **P** are **not M**. *EO2*
    Some **S** are **not M**.
17. All **M** are **P**. AA3
    All **M** are **S**.
18. All **M** are **P**. AE3
    All **M** are **not S**.
19. All **M** are **P**. AI3
    Some **M** are **S**.
20. All **M** are **P**. AO3
    Some **M** are **not S**.
21. All **M** are **not P**. EA3
    All **M** are **S**.
22. All **M** are **not P**. **EE3**
    All **M** are **not S**.
23. All **M** are **not P**. EI3
    Some **M** are **S**.
24. All **M** are **not P**. **EO3**
    Some **M** are **not S**.

The syllogistic task **in the new linguistic format**

Response options:
A. All members of the class **S** have the property **P**.
B. All members of the class **S** are also members of the class **P**.
C. All members of the class **S** do **not** have the property **P**.
D. All members of the class **S** are **not** also members of the class **P**.
E. Some members of the class **S** have the property **P**.
F. Some members of the class **S** are also members of the class **P**.
G. Some members of the class **S** do **not** have the property **P**.
H. Some members of the class **S** are **not** also members of the class **P**.
I. It is **not** possible to derive a necessary logical conclusion.

1. All members of the class **M** have the property **P**.
   All members of the class **S** are also members of the class **M**.
2. All members of the class **M** have the property **P**.
   All members of the class **S** are **not** also members of the class **M**.
3. All members of the class **M** have the property **P**.
   Some members of the class **S** are also members of the class **M**.
4. All members of the class **M** have the property **P**.
   Some members of the class **S** are **not** also members of the class **M**.
5. All members of the class **M** do **not** have the property **P**.
   All members of the class **S** are also members of the class **M**.
6. All members of the class **M** do **not** have the property **P**.
   All members of the class **S** are **not** also members of the class **M**.
7. All members of the class **M** do **not** have the property **P**.
   Some members of the class **S** are also members of the class **M**.
8. All members of the class **M** do **not** have the property **P**.
   Some members of the class **S** are **not** also members of the class **M**.
9. All members of the class **P** have the property **M**.
   All members of the class **S** have the property **M**.
10. All members of the class **P** have the property **M**.
    All members of the class **S** do **not** have the property **M**.
11. All members of the class **P** have the property **M**.
    Some members of the class **S** have the property **M.**
12. All members of the class **P** have the property **M**.
    Some members of the class **S** do **not** have the property **M**.
13. All members of the class **P** do **not** have the property **M**.

All members of the class **S** have the property **M**.

14. All members of the class **P** do **not** have the property **M**.
    All members of the class **S** do **not** have the property **M**.
15. All members of the class **P** do **not** have the property **M**.
    Some members of the class **S** have the property **M**.
16. All members of the class **P** do **not** have the property **M**.
    Some members of the class **S** do **not** have the property **M**.
17. All members of the class **M** have the property **P**.
    All members of the class **M** are also members of the class **S**.
18. All members of the class **M** have the property **P**.
    All members of the class **M** are **not** also members of the class **S**.
19. All members of the class **M** have the property **P**.
    Some members of the class **M** are also members of the class **S**.
20. All members of the class **M** have the property **P**.
    Some members of the class **M** are **not** also members of the class **S**.
21. All members of the class **M** do **not** have the property **P**.
    All members of the class **M** are also members of the class **S**.
22. All members of the class **M** do **not** have the property **P**.
    All members of the class **M** are **not** also members of the class **S**.
23. All members of the class **M** do **not** have the property **P**.
    Some members of the class **M** are also members of the class **S**.
24. All members of the class **M** do **not** have the property **P**.
    Some members of the class **M** are **not** also members of the class **S**.