THE TEACHING OF MATHEMATICS

2022, Vol. XXV, 1, pp. 1–12

GEOMETRY UNDERSTANDING ASSESSMENT BASED ON VAN HIELE THEORY USING COMPARATIVE JUDGMENT

Aleksandar Balašev-Samarski, Manuela Muzika Dizdarević and Selma Zeković

Abstract. The importance of improving and raising the understanding of core concepts in mathematics is well known. Geometry, along with sets, algebra, data and probability, is one of the basic mathematics fields taught in primary and high schools. We have noticed that lately, students' results in geometry are much worse than the results in other mathematical fields. One of the proven methods for improving teaching geometry is harmonising the learning process and content presented in school with the level of understanding on which students are. The mentioned method is based on Van Hiele theory. In order to monitor the effectiveness of the method, it is necessary to assess the level of understanding of students periodically. On the other hand, Comparative Judgment as an assessment method is efficient, fast, and has good outcomes. Our research aimed at investigating whether the Comparative Judgement method can be used to predict the level of understanding of geometric concepts according to the Van Hiele theory.

MathEduc Subject Classification: D44

AMS Subject Classification: 97G40; 97D60

Key words and phrases: Assessment; van Hiele theory; Comparative Judgement.

1. Introduction

Mathematical competencies are considered as critical competencies for personal development and employment in 21st-century society. PISA 2018 results show that Bosnia and Herzegovina's learning outcomes are below average [7]. Unsatisfactory results have been achieved in reading, mathematics and science. Only 42% of students in Bosnia and Herzegovina attained Level 2 or higher in mathematics (OECD average: 76%). This, among other reasons, results in a need for a detailed review of the education system and the launch of its comprehensive reform.

Geometry is one of the essential mathematics areas taught in elementary school. In the older grades (6th-9th grade) in primary schools in Bosnia and Herzegovina, geometry is represented by 45% of the mathematics curriculum. In recent years, we have noticed a decline in the level of understanding of geometric concepts, even among mathematics students. Learning mathematics in Bosnia and Herzegovina is based on the adoption of procedures that enable successful problem solving, but the understanding of the concepts is neglected. The shortcomings of such a teaching process are evident in geometry. Such approach certainly requires change.

Van Hiele theory of levels of understanding geometry was created in 1956. In 1970's and early 1980's, it was already implemented in mathematics curricula in the

Soviet Union, the United States, and many other countries. The implementation of the theory resulted in good outcomes. The essence of the application of *Van Hiele theory* is in harmonising the level at which geometry is taught with the level of understanding of students. In this way, there is a gradual desired progress in conceptual understanding. Monitoring the application of *Van Hiele's theory* requires periodic assessments of students' understanding. Determining Van Hiele level is not an easy task. It requires trained examiners and a series of tests and interviews performed on students.

There is growing interest in using *Comparative Judgement* to assess student work as an alternative to traditional marking [4]. This method has achieved an extremely high level of reliability, often much higher than traditional assessment methods [2]. *Comparative Judgment* has great potential adjustments to become a method applicable to the evaluation in different educational contexts. In the paper [1] it is shown that *Comparative Judgement method* is a quick and efficient alternative method of measuring conceptual understanding in mathematics and could therefore be particularly useful in *Experimental studies*.

During the scientific conference *Research in Mathematics Education* (`https://dms.rs/conference-research-in-mathematical-education/`), which was held in the spring of 2019 in Belgrade, we participated in a workshop "Research on monitoring and evaluation of knowledge in mathematics teaching – from quantitative to qualitative research methods" (see [6]) The workshop leader was plenary speaker Patrick Barmby. He introduced us to the basics of *Comparative Judgement* methodology. Then we came up with the idea to conduct this research. We are investigating whether the *Comparative Judgement* method, as quick and easy to apply, can be used to estimate Van Hiele's levels of understanding geometry. Here we provide results of the study we conducted in two high schools in Sarajevo. We created a test by which we ranked students according to Van Hiele scale, then applied the *Comparative Judgement* method on the same tests and compared the obtained results.

2. Comparative Judgement

Comparative judgement is a research method that originates in the academic discipline of psychophysics. People make better evaluations by comparing objects than by evaluating them according to a certain criterion. The law of comparative judgement is a psychological phenomenon that Thurstone wrote about and showed some special applications in measuring psychological values [10]. Thurstone established that human beings are consistent with one another when asked to compare one object with another, but are inconsistent when asked to judge the absolute location of an object on a scale [4].

The method of *Comparative judgement* has great potential for adaptation. ACJ (*Adaptive Comparative Judgment*) is a modification of Thurston's CJ method and an adaptation of education. Pollitt [8] was the first to write about using the CJ method in education as an alternative for scoring tests pointing out the possibility of adaptive use.

The application of the CJ method is simple and does not impose additional requirements on students. Before applying this assessment method, it is necessary to form a team of judges and determine the criteria on the basis of which the tests will be evaluated. The process of applying this method is performed as follows: After testing, the students' papers were converted to digital form and uploaded to the software that allows implementation of the method. Students' tests are presented on the screen in pairs, and for each pair of students' responses the judges decide which is "better". Judges make assessments relatively quickly (a few minutes at the most). The results are displayed as a ranking list after all judges have completed the evaluation process [1].

The comparative assessment method does not use detailed grading criteria or scoring rubrics, and the final ranking is based on the collective expertise of the judges [1]. This method has been shown to have an extremely high level of reliability, often higher than traditional grading methods have ever reached [9].

Advances in technology have had a positive impact on the development and improvement of the method. Several platforms have been developed to enable the implementation of the method, while the use of statistical methods has reduced the number of comparisons, which directly reduces the time required for evaluation, and thus the cost of implementation.

The ACJ model can be used in a variety of situations such as assessment: writing, art, reports and complex portfolios, and a variety of other contexts. This model provides an assessment of the quality of each student, and offers teachers a new approach to summative and formative assessment [9].

3. Van Hiele theory

The Van Hiele level theory, developed by a married couple Dina Van Hiele-Geldof and Pierre Van Hiele, at Utrecht University in the 1950's, is an elegant theory regarding the acquisition of an understanding of geometry as a mathematical system [11]. The theory is based on the existence of 5 levels of understanding geometry, characterisation of each level and the principles of transition from one level to the next. The levels with their basic properties are:

Level 0 (Visualisation)

At the visualisation level, students can recognise shapes (for example, triangles, squares) but not their exact properties. Students can draw, write and identify the parts of a geometric figure. Students do not recognise the square as a rectangle with all equal sides, nor they recognise the rhombus as a parallelogram.

Level 1 (Analysis)

The properties of geometric figures are identified and categorised at this level, and students can describe the figures according to their properties. Students begin to notice shapes' properties but cannot see the relationships between the properties.

Level 2 (Abstraction)

At an Abstraction level, students understand the relationships between the figures and identify the relationships between geometric properties. Students can reason with simple arguments about geometric figures. They understand necessary and sufficient conditions and can write concise definitions. At this level, students are incapable of understanding the meaning of deduction.

Level 3 (Deduction)

Students recognise the importance of definitions, axioms, theorems and the significance of deduction. Students believe that the system of axioms is fixed and can not understand non-Euclidean geometry.

Level 4 (Rigour)

Students can understand systems based on different axioms at a Rigour level and can study non-Euclidean geometries with understanding. Students can compare systems based on different axioms and study various geometries in the absence of concrete models.

Four important characteristics of the theory are summarised as follows in [11] and [3].

–*Fixed order*– The order in which students progress through the levels is invariant. In other words, a student can not be at level n without having passed through level $n-1$.

–*Adjacency*– At each level of thought, that was intrinsic in the preceding level becomes extrinsic in the current level.

–*Distinction*– Each level has its own linguistic symbols and its own network of relationships connecting those symbols.

–*Separation*– Two persons who reason at different levels cannot understand each other.

Each Van Hiele level is characterised by corresponding associated thought processes of the students at that level. If the teacher uses the language and reasoning characteristic for the higher level, he/she becomes incomprehensible to the students. This results in stopping the learning process and prevents further progress.

4. The research

4.1. Participants

Participants in the research (N=93) were students from two high schools in Sarajevo. The standards of achievements in mathematics at these schools were above national average. Testing was performed in October 2020. In that period first and fourth grade students attended the school. Second and third graders attended classes online from home. Classes were held according to a combined method due to the pandemic.

Total of 55 students of the first grade (aged 15) and 38 students of the fourth grade (aged 18) were included in the study. Students took part in the study voluntarily during their regular mathematics lesson. The study was approved by the Ministry of Education. Information sheets with opt-out option were sent to parents prior to the study.

4.2. Stimuli

The tests for assessing the level of the Van Hiele scale usually contain 25 questions, five questions for each level [11]. A student is considered to be at the n-th level if he/she has met the requirements for all levels lower than the n-th level and has correctly answered three of the five tasks provided for the n-th level.

The mathematics curriculum of the education system in Bosnia and Herzegovina for primary and secondary school covers only Euclidean geometry. Non-Euclidean geometry and rigorously axiomatic foundation of geometry are not taught in schools. For this reason, we did not include questions for the fourth level of the Van Hiele scale in the test.

The tasks we have chosen for the test belong to the thematic units covered in detail in primary school and relate to crucial geometric concepts such as congruence of triangles, classification of quadrilaterals, the relationship of the line and the circle and the Pythagorean theorem.

We intended to create a test that would be suitable for comparative judgment and assess Van Hiele levels. Therefore, the test included tasks that had different acceptable answers for different levels.

The test we used is presented in Appendix A.

4.3. Procedure

Participants were given 25 min to complete the test. The anonymised student papers were uploaded to the online comparative judgement engine at No More Marking (`https://www.nomoremarking.com/`) platform for assessment.

The student papers were judged by six judges. Two of them were university teachers, two judges were PhD students in Mathematics in education, and one judge was a high school teacher. All judges were sent a link by email to access the mentioned judging website, as well as a short instruction manual on how to use the website. The judges were allocated 155 pairwise comparisons. This resulted in 930 judgements in total, with each test being compared between 19 and 22 times.

Inter-rater reliability[1] was strong, $r = 0.91$.

Let us now consider the criterion by which we performed the assignment of students to the appropriate levels of the Van Hiele scale.

In the first part of the first task, we offered eight polygons, and parallelograms had to be chosen among them. We considered that students who correctly solved this task could recognise and classify appropriate geometric figures. So these students are placed on level one of the Van Hiele scale.

The second level of the Van Hiele scale characterises recognising the properties of geometric figures and the connections among them and the ability to define concepts precisely and meaningfully. The characteristic of students at this level is that they combine empirically obtained conclusions with a deduction because

[1] In statistics, inter-rater reliability is the degree of agreement among independent observers who rate, code, or assess the same phenomenon.

they do not fully understand the importance of deductive reasoning. Therefore, we placed students on the second level of the Van Hiele scale if they did at least two of the following three tasks correctly:

1) The second part of the first task, i.e. they gave the correct definition of a parallelogram.
2) The first part of the second task. We also considered the answers in which two pairs of congruent triangles (triangles 2) and 5)) and (triangles 3) and 4)) were recognised, as the correct answer, although the triangles in the second pair are not congruent. The reason lies in the fact that students at the second level intuitively understand the congruence and triangles 3) and 4) "look" as congruent.
3) The first part of the third task. We considered that the answer was correct if the offered drawing indicated that the student understood the required relationship between the line and the circle.

The third level of the Van Hiele scale is characterised by the ability to deduce by deduction. Therefore, at this level, students understand the meaning of proof of a mathematical statement and can make a sequence of implications from already proven statements to the desired statement. Therefore, we placed students on the third level of the Van Hiele scale if they did at least two of the following three tasks correctly:

1) Second task. The correct answer to the second task was a recognised congruent pair of triangles 2) and 5) followed by an appropriate explanation.
2) The second part of the third task. The correct answer to the second part of the third task directly indicates a deeper understanding of the properties of circles and their tangents and leads to the conclusion that the required locus is a union of two lines, i.e. angle bisectors.
3) The second part of the fourth task. Correct proof of Pythagorean theorem in any of the known ways, including the use of trigonometry or vectors.

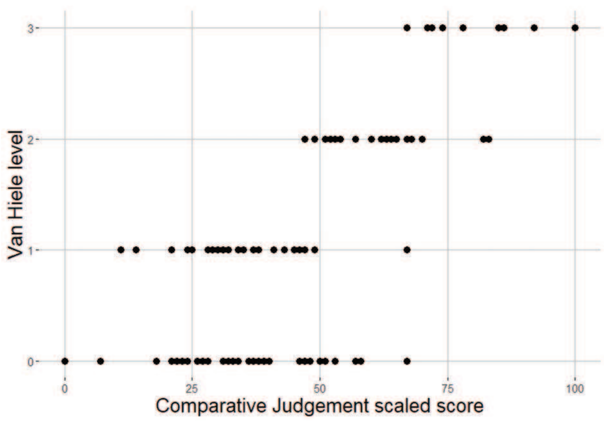


Fig. 1. Scatter plot diagram of scaled scores vs Van Hiele level

4.4. Results

There were total of 93 participants tested. After the testing phase, comparative judgement results were downloaded from the judging website, in the form of CSV file.

The mentioned judging website uses variant of Bradley-Terry algorithm, with progressive method applied [12].

Each participant was given scaled score, in the possible range 0–100. Actual values were (Min=0, Max=100, Median=46, Mean=45.88).

We assigned an appropriate level to each test according to the Van Hiele scale in the manner already described. Number of participants for Van Hiele levels are: $n_0 = 36$, $n_1 = 27$, $n_2 = 21$, $n_3 = 9$.

According to Van Hiele theory, hierarchy must be satisfied. For example, at level three, one cannot find a student who did not meet the criteria for levels 0, 1 and 2.

We used *Guttman's scalogram* (`https://www.questionpro.com/blog/guttman-scale/`) to determine whether Van Hiele levels tested by our test form a hierarchy. We assigned a four-digit binary number to each student's work, which indicates the satisfaction or non-satisfaction of the conditions for each of the four levels of the Van Hiele scale. The numbers 0000, 1000, 1100 and 1111 denote student's papers in which the hierarchy is satisfied. We consider all other combinations errors. In our research, seven errors appeared. The calculated value of Guttman's coefficient [5] is 0.9812, which confirms the reliability of the test.

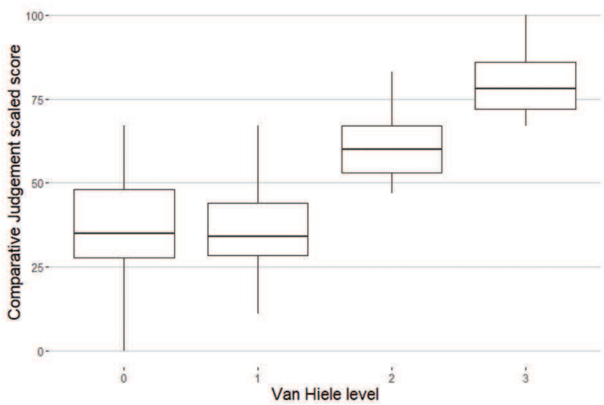


Fig. 2. Box plot diagram of scaled scores vs Van Hiele level

We noticed that level zero and level one were not sufficiently separated. Figures 1 and 2 show that students at level zero have a better score than students at level one. The reason lies in the fact that we set only one question that referred to the first level of the Van Hiele scale in the test.

The scores for the group of participants marked with Van Hiele Level 0 have mean (M), median and standard deviation (SD) as folows: $M = 36.67$, Median $= 35.00$, $SD = 14.67$. The scores for the group of participants marked with Van Hiele Level 1, have $M = 35.00$, Median $= 34.00$, $SD = 11.76$. Pearson's correlation coefficient between two respective scores is $r = -0.0619$ and Spearman's coefficient is $\rho = -0.0812$. That fact and diagrams above show that there is clearly no correlation between scores and Van Hiele levels, if we concentrate solely on Van Hiele levels 0 and 1.

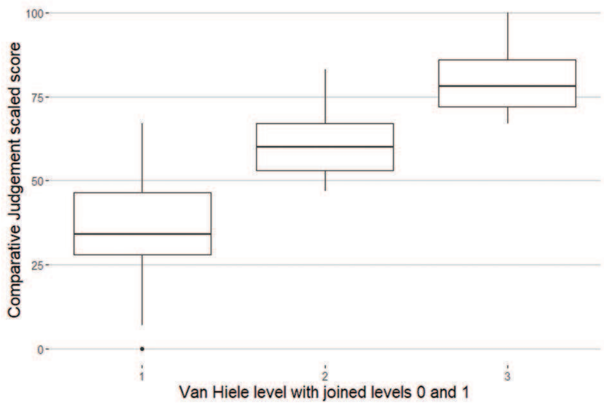


Fig. 3. Box plot of scaled score categorised according to Van Hiele level, with joined Van Hiele levels 0 and 1 presented under 1

Therefore, in order to make further conclusions, we merged participants marked with Van Hiele level 0 and 1 into one unique level 1. Now, we have the following numbers: $\tilde{n}_1 = 63$, $\tilde{n}_2 = 21$, $\tilde{n}_3 = 9$. The box plot diagram categorised according to new levels is presented in Figure 3. Having merged Van Hiele levels into one with value 1, calculation of Spearman's correlation coefficient gives value of 0.7588. Values of respective coefficients suggest moderate to strong correlation between scaled score and Van Hiele level. Note that Spearman's coefficient wouldn't have changed if we recoded Van Hiele level to level 0 instead of 1. Respective means and standard deviations for adjusted Van Hiele levels 1, 2 and 3 are: $M_1 = 35.9524$, $SD_1 = 13.4217$, $M_2 = 60.8095$, $SD_2 = 10.0728$, $M_3 = 80.5556$, $SD_3 = 10.9328$.

In order to further investigate possible correlation between comparative judgement scaled score and graded Van Hiele level, we made the following ranges for scaled scores. Each test question was given specific points, which would have been given for a traditional marking, giving a total of 100 points for all questions. After that, the minimum number of points needed to give the appropriate Van Hiele level based on the scaled score was calculated, and ranges for Van Hiele levels were formed. We could make such ranges because comparative judgement scaled score and traditional marks with points given for each question show a strong correlation, with Pearson's correlation coefficient $r = 0.8905$ and Spearman's correlation coefficient $\rho = 0.8905$.

Therefore, having strong linear correlation between comparative judgement scores and traditional scores, it was natural to assume that number of points for each question was appropriately allocated.

Ranges were 0–43 for Van Hiele levels 0 and 1, range 44–85 for Van Hiele level 2 and range 86–100 for Van Hiele level 3. According to these ranges, expected Van Hiele level was calculated, based on scaled score. For newly calculated Van Hiele level 0 and 1 we have number of participants $n_1 = 43$, for Van Hiele level 2 we have $n_2 = 47$ and for Van Hiele level 3 we have $n_3 = 3$. These numbers are summarised in Table 1.

Table 1. Numbers of participants that achieved Van Hiele level

Type	Levels 0 and 1	Level 2	Level 3
Actual Van Hiele level	63	21	9
Expected Van Hiele level	41	47	3

Conducting chi-squared goodness-of-fit test of Van Hiele levels and expected Van Hiele levels based on score, we have statistical value $\chi^2 = 35.685$, while tabulated critical value for $\alpha = 0.05$ and $df = 2$ is 5.991, with p-value $= 1.783 \times 10^{-8}$. This gave unacceptable fit, so we rejected hypothesis that marked Van Hiele levels are distributed according to hypothetized score ranges.

Secondly, in search for possible connection, we conducted chi-squared Pearson's test of independence, comparing marked Van Hiele level and expected Van Hiele level according to given ranges. The test produced statistical value $\chi^2 = 63.508$, while tabulated critical value for $\alpha = 0.05$ and $df = 4$ is 9.488, with p-value $= 5.306 \times 10^{-13}$, so we can safely conclude that there is dependency between marked Van Hiele levels and expected Van Hiele levels given according to ranges.

Finally, Fisher exact test for independence showed $p = 4.842 \times 10^{-13}$, giving again strong evidence that there is dependency between marked Van Hiele levels and expected Van Hiele levels given according to ranges.

5. General Discussion

In the study reported here, we investigated the use of the comparative judgment method to assess Van Hiele level of understanding of geometry. It should be noted that the only *output* of comparative judgement assessment is scaled score. Based on that scaled score, we seek for appropriate function of scaled score, which would estimate Van Hiele level. Our assumption was based on the following:

- The better scaled score, the higher estimated Van Hiele level.
- Van Hiele levels are approximately distributed according to ranges of scaled scores.

All that can be currently said, based on the test data, is that there is some kind of moderate correlation between comparative judgement scores and Van Hiele levels, and that on average, the better score produces higher Van Hiele level. But we cannot conclude what is Van Hiele level based on comparative judgment score.

A fundamental limitation of our research was the focus on levels one, two and three of the Van Hiele scale. Namely, we did not expect that so many students in the gymnasiums where we conducted the research would be at level zero of the Van Hiele scale, so we did not create a test that could better distinguish level zero at level one.

We planned to conduct a pilot study where we would notice this limitation, but that was not conducted due to several difficulties. Due to the COVID-19 pandemic, school classes were held in small groups at the time of testing. Soon after the first testing, students moved to the online teaching model, and we did not have the opportunity to conduct a new test that would allow better differentiation between the initial levels of the Van Hiele scale. This fact indeed confirms the need to change the way geometry is taught in our schools.

We hope that this study contributes to the growing literature that evaluates comparative judgment applying to varied ranged educational assessment activities.

Acknowledgements. The authors are very grateful to dr Patrick Barmby and No More Marking for using the platform to apply the comparative judgment method for research purposes.

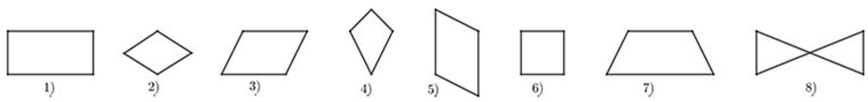
The authors are grateful to schools and students who made this study possible.

The authors would like to thank the referee for the comments and suggestions, which helped to improve the manuscript.

Appendix A

TEST

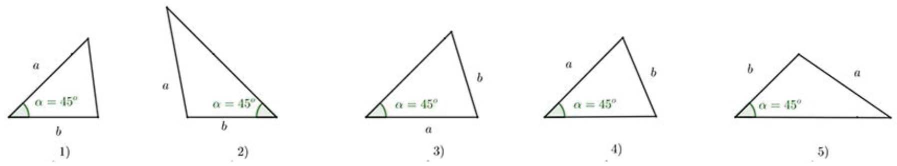
1. Find and encircle all parallelograms in this picture:



Complete the sentence:

A parallelogram is . . .

2. Find and encircle all pairs of congruent triangles on this picture:



Explain why those triangles are congruent.

3. Suppose that two intersecting lines are given. Draw a few circles that touch the given lines. Determine the locus of centers of all circles that touch the given lines. Describe that set. (Draw the picture!)

4. a) State the Pythagorean theorem.
   b) Prove the Pythagorean theorem.

REFERENCES

[1] M. J. Bisson, C. Gilmore, M. Inglis, I. Jones, *Measuring conceptual understanding using Comparative Judgement*, Int. J. Res. Undergrad. Math. Ed. **2** (2016), 141–164.

[2] M. J. Bisson, et al., *Teaching using contextualised and decontextualised representations: exeminig the case of different calculus through a comparative judgement technique*, Res. Math. Education, **22**, 3 (2020), 284–303.

[3] M. de Villiers, *Some reflections on the Van Hiele theory*, Invited plenary lecture on the 4th Congress of Teachers of Mathematics (2010).

[4] N. Marshall, K. Shaw, J. Hunter, I. Jones, *Assessment by Comparative Judgement: An application to secondary statistics and English in New Zealand*, New Zealand J. Educational Studies **55** (2020), 49–71.

[5] J. Mayberry, *The Van Hiele levels of geometric thought in undergraduate preservice teachers*, J. Res. Math. Education, **14**, 1 (1983), 58–69.

[6] J. V. Milinković, *New impulses for research in mathematics education*, The Teaching Math., **22**, 2 (2019), 77–85.

[7] OECD, *Education in the Western Balkans: Findings from PISA*, PISA, OECD Publishing, Paris, 2020. `https://doi.org/10.1787/764847ff-en.`

[8] A. Pollit, *Let's stop marking exams*, IAEA Conference, Philadelphia, (2004).

[9] A. Pollitt, C. Whitehouse, *Using Adaptive Comparative Judgement to Obtain a High Reliable Rank Order in Summative Assessment*, AQA Centre for Education Research and Policy, (2012).

[10] L. L. Thurstone, *A law of comparative judgment*, Psychological Rev., **34**, 4 (1927), 273–286.

[11] Z. Usiskin, *Van Hiele Levels and Achievement in Secondary School Geometry*, University of Chicago (1982).

[12] C. Wheadon, *Open Sourcing Our Comparative Judgement Algorithms*, `https://blog.nomoremarking.com/open-sourcing-our-comparative-judgement-algorithms-84d12d92f9c4`

Faculty of Natural Sciences and Mathematics, University of Sarajevo, Bosnia and Herzegovina

E-mail: `manuela@dizdarevic.org`