Open Education Studies, 2021; 3: 212–225 DE GRUYTER

Research Article

Eduard V. Pavlysh*, Sviatlana V. Astapchuk, Alecia Adelaide May Reid, Carlos Rioja del Rio, Tiina Mäkelä, Kristof Fenyvesi, Dimitris Pnevmatikos, Panagiota Christodoulou, Juho Mäkiö

Developing a Method for Measuring Science and Technology Oriented Creativity (STOC)

https://doi.org/10.1515/edu-2020-0155
received March 31, 2021; accepted July 14, 2021.

Abstract: The article contains the results of a research within the STIMEY (Science, Technology, Innovation, Mathematics, Engineering for the Young) project funded by the European Union's Horizon-2020 research and innovation program (2016-2021). In the project, a hybrid learning environment (LE) was developed for both on-site and online learning suitable for the learning conditions in the COVID-19 era and beyond. The purpose of the research segment presented in this paper was to develop an instrument for assessment of the learner's progress in creativity as one of the key targets of STIMEY. COVID-19 has shown that creativity is also needed to allow education systems the flexibility for unexpected changes and circumstantial challenges. The article presents a definition of Science and Technology Oriented Creativity (STOC) based on the existing theories in the field, a short outline of the STOC measurement method developed, procedures, evaluation algorithms and an overview of the experimental STOC testing results. The relevance of the method developed to its purpose and suggested tentative positive influences of the STIMEY LE on learners' creativity are discussed. Argumentations of strengths, limitations, applications of STOC testing outside the STIMEY project and the key directions of further improvements of the method developed are provided.

Keywords: STEM; aggregate group creativity; problem-solving; learning environment; assessment.

***Corresponding author: Eduard V. Pavlysh,** Vitebsk State University named after P.M. Masherov, Vitebsk, Belarus, E-mail: eduard.pavlysh@gmail.com
Sviatlana V. Astapchuk, Polotsk State University, Belarus
Alecia Adelaide May Reid, Carlos Rioja del Rio, University of Cadiz
Tiina Mäkelä, Kristof Fenyvesi, University of Jyväskylä, Finland
Dimitris Pnevmatikos, Panagiota Christodoulou, University of Western Macedonia, Greece
Juho Mäkiö, University of Applied Sciences Emden / Leer, Germany

1 Introduction

The need for boosting creativity (Szabó et al., 2019) in teaching and learning science, technology, engineering, and mathematics (STEM) has been internationally recognized. It has also been one of the objectives the European Union has defined in its research and innovation actions[1]. Fostering technological (inventions), economic (entrepreneurship), and artistic (cultural) creativity in learning requires supporting learners' ability to generate ideas, experiment and solve problems in novel ways. Fostering innovation, in turn, implies guiding the implementation of creative ideas in order to create economic or social value (Edwards-Schachter et al., 2015). In addition to identifying ways to boost creativity in STEM, there is a need to develop tools for evaluating creativity related to STEM studies.

In this paper, we will present the process of developing a method for measuring particularly science- and technology-oriented creativity. This study was part of a broader European research project named STIMEY (Science, Technology, Innovation, Mathematics, Education for the Young) funded by the European Union's Horizon 2020 research and innovation program (2016-2019) conducted in Belarus, Finland, Germany, Greece, and Spain. The project researched and developed a hybrid STEM Learning Environment (LE) for young people of 10 to 18 years of age. A hybrid LE was designed for both on-site and online formal and informal learning. It was piloted before and during the COVID-19 pandemic in different settings: in traditional classroom learning, hands-on on workshops, in fully remote learning situations, and in hybrid settings in which sessions were facilitated by the researchers remotely while teachers supported learners at school.

1 See e.g., https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/seac-1-2014

The developed STIMEY LE[2] consists of components of a digital learning environment such as a social web platform, e-portfolio, serious games, entrepreneurial tools, a digital radio as well as physical socially assistive robots. The STIMEY LE connects various stakeholders in shared efforts to increase both female and male students' interest and motivation in STEM education, innovations, and engineering, technology, and science careers from the young age. In addition to STEM subjects, the STIMEY project focused on cross-curricular skills (also named in the literature as transversal skills or competencies, 21st-century skills, or key competences), and particularly creativity and innovation.

One of the objectives of the STIMEY project was to improve learners' Science- and Technology-Oriented Creativity (STOC). To achieve that goal, a theoretical background research was carried out, which resulted in defining the STOC. Following the definition suggested, a STOC measurement method was developed and complemented with a description of testing procedures. The article also describes the design of the evaluation algorithm and criteria needed for the test results assessment. Finally, the article provides a short overview of the validity and experimental STOC testing results. The discussion and conclusion sections describe the strengths and limitations of the method developed. They also offer improvements of the STOC evaluation method and possible applications of the method outside the STIMEY project framework.

2 Phases of the method development

2.1 Theoretical background

Defining and theorizing creativity has been a challenge in all areas where creativity is studied (Wyse & Ferrari, 2015). The definitional challenge is due to creativity's "multifaceted and complex nature," the "broadness" and "lack of specificity" of the term (Henriksen et al., 2018). Creativity's multidimensional interrelations (Hollanders & Van Cruysen, 2009) in various disciplinary, historical, and cultural contexts (Banaji, Burn & Buckingham, 2006) contribute to its inconsistent understanding across multiple domains.

2 See https://stimey.eu/home

Despite the definitional challenge, there is a general agreement about creativity's critical role in 21st-century education: creativity has been regarded essential in most of the current teaching and learning trends (OECD, 2019). Simultaneously, the competing interpretations of creativity in the educational discourse have several common characteristics. Creative ideas, solutions, and approaches are usually identified with novelty, originality, effectiveness, inventiveness, value, and productivity (Corazza, 2016). Creative learning strategies and teaching methods in education are also frequently associated with flexibility, discovery, inquiry, enhanced capacity for imagination, and cognitive and embodied abilities to restructure the information in problem-solving situations (Henriksen et al., 2018).

In addition to teaching and learning practices, creativity has been widely discussed in educational assessment and policy contexts and studied from the perspective of everyday schooling realities as well. The intention behind conceptualizing standard definitions for creativity (Runco & Jaeger, 2012) has been significantly motivated by the policy level and the need for assessing creativity and creative assessment.

The first research task was related to outline the existing definitions of creativity by suggesting a definition for STOC within the STIMEY project domain. Table 1 summarises the definitions of creativity proposed by various authors that we used for systematizing various definitions of creativity.

Combining the definitions and approaches summarised in Table 1, we concluded that creativity is a phenomenon that involves three main domains: ability, process and capability. It is simultaneously:

- a mental ability to generate new ideas or concepts;
- a process of generation of new ideas or concepts;
- a capability for creation of original material or non-material artifacts.

One of the tasks within the objective of STOC measurement was to outline measurable parameters of creativity. To fulfill this task, we needed not only to define creativity, but also to outline the structure of creativity as a process.

To develop theories on creativity, research psychologists focus their emphasis and investigations on one or more central aspects of creativity, which has been labeled the six "P's.": process, product, personality, place, persuasion, and potential (Kaufman & Sternberg, 2010). Systems theory developed by Mihaly Csikszentmihalyi (2014), emphasizes big-C creativity and how it takes place interrelationally among three components: (1) the domain, that is, a body of knowledge for a particular discipline

Table 1: Systematisation of creativity definitions.

Author (Source)	Definition	Central feature
Torrance (1966)	A process of becoming sensitive to problems, deficiencies, gaps in knowledge, missing elements, disharmonies, and so on; identifying the difficulty; searching for solutions, formulating hypotheses about the deficiencies: testing and retesting these hypotheses and finally communicating the results.	**Process**
May (1975)	The process of bringing something new into being.	
Sassenberg & Moskowitz (2007)	Mental process of generation of new ideas (concepts), or new associations between existing ideas (concepts).	
	Cognitive process leading to original and appropriate outcomes.	
Simonton (2004)	Creativity is the act of turning new and imaginative ideas into reality. <…> It involves two processes: thinking, then producing.	**Capability**
Sternberg (2011)	A product is creative when it is (a) novel and (b) appropriate.	
Mumford (2003)	Creativity involves the production of novel, useful products.	
Amabile (1983)	The production of novel and useful ideas in any domain.	
Boden (1998)	Ability to come up with "new, surprising and valuable" things.	**Ability**

(math, fiction writing, etc.) containing symbolic rules and procedures; (2) the field comprising the gatekeepers of the knowledge contained within a domain deciding if a novel idea, product, or artwork (journal editors, art and book critics etc.) will be included in the domain; and (3) the person with big-C creativity (educated and knowledgeable in all important aspects of a domain) taking the domain's rules and procedures, developing a novel pattern, and persuading the gatekeepers to accept this novelty into the field.

Due to STIMEY objective and our research tasks on creativity, we chose to focus on "personality" and "potential" as (a) these are two aspects which directly correlate with teaching and learning processes, and (b) these are two aspects, which can possess measurable parameters, contrary to "process", "place", "product", and "persuasion". Moreover, a "systems theories'' and componential theory of creativity verifies the existence of domain-specific types of creativity, which (a) supports the idea of possible specific traits of STOC; (b) confirms the idea of creating new (or modifying existing) teaching and learning methodologies focused on STOC development, which is one of the central tasks of STIMEY.

Furthermore, taking into consideration both science and technology definitions, we can say that technology is a system of material (tools, machines etc.) and non-material (rules, algorithms etc.) components, aimed at creation of products (services), satisfying definite personal or (and) society needs (Dusek, 2006). Technology is also a result of applied research and (or) dispensing solutions processes. A functional meaning of science, in turn, may be described as a three-stage process of technology creation (Salmon et al., 1999):

– Pure research giving patterns for the applied research;
– Applied research producing prototypes and algorithms;
– Practitioners shaping the results of applied research into concrete forms.

In particular, Applied Science (Technology creation or improvement) is a process of problem solving.

Based on these considerations, E. Paul Torrance's (1966) process-based definition of creativity seems to be the most appropriate. Therefore, combining the definitions of Creativity, Science and Technology, we suggest the following definition of STOC:

– an ability to identify the deficiencies in existing knowledge or technology, formulate (or discover) the problem and give possible solutions to it;
– a process of new knowledge or technology creation;
– a capability to generate original solution(s) to new or existing scientific/technological problems.

A person is creative in science and technology if he/she is capable of, first, *identifying a problem(s)*, second, *finding out the causes of the problem(s)*, and third, *suggesting solutions to a problem(s)*. This was the specific domain and definition of the creativity used within this research.

Moreover, instead of focusing on evaluating an individual´s creativity, we chose to measure group creativity. The term "group creativity" is usually used to describe the collaborative creativity, or creativity of the teams (Nijstad & Stroebe, 2006; Paulus, 2000). The concept of group creativity is also in line with authors such as Vygotsky (2004), who viewed creativity as collective and accumulative. It describes the potential of the group to produce creative results when working on a common task. But in STIMEY's case, the groups are mainly nominal, chosen due to such attributes, as age, gender, residency, relation to STIMEY LE piloting. This means that in our research we defined the group creativity as the aggregate creativity of the learners, united into test groups, measured as the average creativity of the participants of the group.

With regards to assessing creativity, there is a multitude of different models already available (Barbot, Besancon, & Lubart, 2011). For instance, in J.P. Guilford (1950) model, fluency, flexibility, originality and the elaboration of the responses to a given problem are measured. Divergent thinking tests such as the Torrance (1966) Tests of Creative Thinking and the Wallach and Kogan test (1965) evaluate the capacity of the individual to generate as many ideas as possible from a simple starting point, in a limited amount of time. The Creativity Assessment Packet (Tunik, 2003) includes three sections, which are the divergent thinking test, divergent feeling test and the William scale which measures the perception of teachers regarding their learners´ creativity. We did not identify, however, specific methods for assessing aggregate progress of a certain group of learners in STOC over a certain period of time. In this aspect our research provides a novel and specific tool for this domain.

2.2 Method for STOC evaluation

Based on the theoretical considerations described above, we chose to measure the degree of STOC through the degree of a person's divergence in thinking – the ability to generate multiple solutions to given problems. In STIMEY, the main goal was to assess the effectiveness of methods stimulating learners' creativity, but not to judge their abilities. In this sense, creativity testing should not produce results, which could be used to question any learners' abilities. To avoid this, the project team proposed to assess the total group creativity (as a sum of individual creativities), and then – to evaluate the progress of the whole group. Furthermore, we decided that the originality of responses would not be measured via the level of novelty. Therefore, we decided that the response would either be original (1) or not (0).

In STOC test, a learner is proposed to perform two tasks. The first task aims to test the ability to identify the deficiencies: (a) a learner is given an object of their everyday use (a toy, a mobile phone, a pencil etc.) and is asked to give as many deficiencies (imperfections) in a given object as possible within the limited period of time. Then, (b) a learner is given an everyday process (riding a bus, brushing teeth, writing a letter etc.) and is asked to formulate what can be improved to make a process more comfortable, interesting, easy, etc. For the purpose of further data processing, the answers should be limited to one sentence. The second task aims to test the ability to generate solutions: a learner is asked to formulate possible solutions to each of the deficiencies identified. The answer should be limited to three sentences. The whole testing procedure should not last longer than sixty minutes, or preferably less, in order to assure that participants would be able to stay focused.

The evaluation of answers is based on the Expert Judgment Method. As Meyer and Booker (2001, p.3) suggest, expert judgement is "data given by an expert in response to a technical problem. An expert is a person who has a background in the subject area and is recognized by his or her peers or those conducting the study as qualified to answer questions. Questions are usually posed to the experts because they cannot be answered by other means''. Benini et al. (2017, p. 13) define Expert Judgement as "expert opinion given in the context of a decision". They also note that "analysis in humanitarian settings is the structured, transparent and controlled human process of transforming raw data into actionable insights. An expert's opinion is data as well as context provided by persons with supposedly superior skills or knowledge when availability, quality, time and cost considerations rule out traditionally sourced data'' (Ibid, p. 13). After the STOC testing is complete, the responses received in each STIMEY project country from the test participants, are de-personalized, translated into English and provided to a group of experts with expertise in pedagogy and (or) psychology, who will assign the attributes to each of the responses.

When attributing each of the responses, if similar answers are found, or the answer belongs to a common practice, which is obvious for the expert, it is marked as "standard", which means it is not original. All the answers, not marked as standard, are marked as "non-standard". The experts also should mark senseless and non-functional answers. "Senseless" answers are answers, not related to the subject of testing. "Non-functional" answers – answers about deficiencies and improvements with weak substantiation. More detailed description of attributes will be given further.

A number of indicators were suggested for measurement of STOC among the learners belonging

Table 2: A shortened list of STOC indicators.

1. The absolute indicators:	2. Ratios accounting for the structure of testing results:
- The total number of:	• *Deficiencies share*
• *answers;*	• *Improvements share*
• *deficiencies outlined;*	• *Deficiencies solutions share*
• *improvements suggested;*	• *Improvements solutions share*
• *solutions for objects deficiencies;*	• *Senseless answers ratio*
• *solutions for improvements;*	• *Nonfunctional answers ratio*
• *non-standard answers;*	• *Nonstandard answers ratio*
• *non-standard deficiencies;*	• *Non-standard deficiencies ratio*
• *non-standard improvements;*	• *Non-standard improvements ratio*
• *non-standard solutions to deficiencies;*	• *Non-standard solutions for deficiencies ratio*
• *non-standard solutions to improvements;*	• *Non-standard solutions for improvements ratio*
• *senseless answers;*	
• *non-functional answers.*	

to different age groups as well as outlining the quality of solutions, originality and uniqueness of learners' responses to the problems given. In Table 2 a shortened list of indicators is listed.

Although the method developed can be used to evaluate the individual performance, rating each learner's individual creativity would not give the results expected. Each learner possesses unique mental characteristics, individual perception and might have specific results. Moreover, some learners might not show the progress in creativity, while others – show an outstanding increase in the indicators applied. In the case of STIMEY, evaluation of individual creativity would have to deal with the problem of choosing, which of the individual results are more relevant, and the problem of substantiation of such choice. To avoid such issues, we chose to measure the total group creativity.

The hypothesis was that if the influence of pedagogical approaches, methods, tools gives the impact, the results of such impact conform to normal distribution. In other words, if such an influence exists, the larger part of the group will show the impact, although some individuals might show no impact at all, while some might show outstanding results.

Before launching the testing procedures, it was considered as essential to make sure that the objects and processes used in the test are those which are familiar to the age groups involved in testing. For this purpose, a questionnaire was launched by STIMEY consortium during "the STIMEY days" conducted in all participant countries in spring 2018. Learners of the three age groups were asked to rate a set of objects and processes. The rating procedure was performed using a 7-point Likert scale (Ankur et al., 2015), with scores from -3 (totally unfamiliar with object or process) to +3 (absolutely familiar object or process). The questionnaire was done by a total of 375 learners from each project country. They represented three age groups, namely 10-12 years (n = 209), 13-15 years (n = 107), and 16-18 years (n = 52) old learners.

The final choice of the objects and processes for the testing was done due to the following assumptions:

1. The object should have the average rating of 2,0 points or higher in the age group.
2. The process should have the average rating of 1,0 points or higher in the age group.
3. The objects and processes in each of the age groups must be unique. No objects or processes should belong to sets of different age groups.

Based on the results of the survey (see Table 3), a set of objects and processes had been prepared to be used for validation and preliminary experiment with STOC testing.

The selection of the objects and processes for the testing in validity test and preliminary experimental test was decided on a random basis from sets presented in Table 2. The objects employed in the test were "a backpack", "a tablet", and "a laptop". The processes used were "cleaning one's room", "taking a shower", and "using social media".

2.3 Development of STOC test answers evaluation technique and algorithm

As the next step in the STOC test development, we organized STOC testing as a part of "STIMEY days" organized in all participant countries in Spring 2019. As was mentioned earlier, all the answers in data collection were depersonalized. The set of answers received is evaluated by five experts representing the partner

Table 3: The set of selected objects and processes for the testing.

10-12 years of age		**13-15 years of age**		**16-18 years of age**	
object	**rating**	**object**	**rating**	**object**	**rating**
a ball	2,11	a bicycle	2,23	a smartphone	2,18
a backpack	2,36	a tablet	2,05	a laptop	2,17
a school desk	2,28	headphones	2,45	an electric kettle	2,45
sneakers	2,53	an exercise book	2,11	a bus stop	2,17
a spoon	2,43	a ballpoint pen	2,10	a pair of jeans	2,08
process	**rating**	**process**	**rating**	**process**	**rating**
cleaning one's room	1,85	taking a shower	2,17	using social media	1,98
making your bed	1,88	making a sandwich	1,56	waking up to an alarm clock	1,88
watching TV	1,70	listening to music	2,50	talking on the phone	2,46
buying sweets	1,66	chatting on-line	1,63	taking the bus	2,08
doing homework	2,09	reading books (paper)	1,35	reading e-books	1,23

universities in each STIMEY project country, who assigned each answer with attributes *standard/non-standard*, *senseless/making sense*, *functional/non-functional*. The attributes standard/non-standard show the originality of the answer. Alongside with the number of answers, given within a limited time period, the share of non-standard answers shows the intensity of science and technology-oriented creativity (how good a learner is at generating ideas aimed at improving the environment). To eliminate answers, which a learner might give without engaging the creative thinking (technical responses, responses not leading to any ideas etc.), each answer is checked for sense. The attributes functional/non-functional relate to quality of the answer and the potential of its utilization.

Assigning attributes to STOC test answers is so far done manually using the expert judgement method. A small example of assigning attributes to answers on object deficiencies is given. The list of chosen deficiencies from the answers, produced by participant learners during the validity test, is presented in the Table 4.

Answers (1) and (2) were marked as *standard* and *functional*. Answer (2) was marked *senseless* while answer (1) marked as *making sense*. Both answers marked *standard* with argumentation that both answers show that two learners independently have noticed that the object does not possess the quality of flying or levitating, which means that notifying such deficiency is typical for learners of such age. Both answers marked *functional* with the argument that modern technologies allow constructing objects with properties of flying or properties of fluff. Possessing the quality of flying in the answer (1) would make the ball an interesting object for the games and leisure, bringing such quality would make *sense*. Answer (2) is *senseless:* the ball is an object for games and the fluff is a material, so the deficiency is too discrete.

Answer (3) was marked *standard*, *making sense* and *functional*. Marked *standard* with comment that although the fact that the ball can cause injury cannot be denied, it is obvious that the ball is not the only object which can cause damage to a person in certain conditions and the probability of injury by the ball in normal conditions is low. Answer (3) *makes sense*, because even though the possibility of injuring the person with the ball is not high, still it exists, and if the ball was "injury-proof" it would definitely be an improvement for the object. Marked as *functional* with the comment that being injury-proof doesn't bring any new features to the ball, but makes using it safer. Answer (4) was marked as *senseless, non-standard* and *non-functional*. Marked *senseless* with comments: "The answer is hard to understand"; "the ball is the object of use; the mother is a person with specific relation to the child". Marked as *non-standard* with explanations that even not understanding what specific features of the object made the learner to answer in a manner he or she did, one can admit the originality of the learner's point of view. Marked as *non-functional* with the argument that the features, which a ball lacks compared to mother, are too many and are not specified in the answer.

Answer (5) marked as *non-standard*, *functional* and *making sense*, suggesting that the answer does not repeat, describes the valid attribute of the ball and can be implemented technically, and assuming that if the ball

Table 4: The examples of the deficiencies of the object.

	The answer No.	The deficiency	Senseless (senseless=1, makes sense =0)	Non-standard (original=1, standard=0)	Functional (functional=1, non-functional=0)
a ball	1	it doesn't fly	0	0	1
	2	it is not like fluff	1	0	1
	3	you can hit somebody in the eye with the ball	0	0	1
	4	ball is not like my mommy	1	1	0
	5	it is the same size all the time	0	1	1
	6	argsytckjgavk	1	0	0
	7	it is boring	1	0	0

could change the size, this might broaden the range of the use of the ball. Answer (6) was marked as *senseless*, *standard* and *non-functional*. Marked is *senseless* because it is a random set of letters. Considered as *standard*, because although it is original, it is not a deficiency and has no meaning at all. Answer (7) was marked as *senseless*, *standard and non-functional*. Marked *senseless* with the argument that it was a subjective judgment. Marked *standard* claiming, the level of boredom is based on each person's personality and ability to be bored by an object. Therefore, it was a generalized deficiency that could be applied to any object. Marked *non-functional* suggesting, although the answer could be understood, it wouldn't give any leads to elimination of the deficiency, it is discrete.

Based on the judgement of 5 experts from consortium partner universities, a set of criteria to assign the attributes was developed. The attribute Standard is assigned, when:

– similar answers are given by two or more participants;
– the deficiency or solution is already widely known;
– the deficiency is a feature of an object or process, possessed by definition (e.g. the shape, the taste);
– the solution suggested is already implemented and the fact is widely known.

The attribute senseless is assigned, when:

– the answer is a random set of letters;
– the deficiency noted or solution suggested is a meaningless phrase;
– the deficiency noted or solution suggested is overly subjective (e.g. "I just don't like it");
– the deficiency noted relates to properties of other class of objects or processes (e.g. "The carpet is not a dog");
– the deficiency noted relates to useless properties of the object or process (e.g. "pencils cannot cook dinner").

If the answer is neither standard, nor senseless, it should be marked as non-standard. Based on the criteria suggested, a special algorithm was designed - the STOC Test Answers Evaluation Algorithm.

3 Validity testing

Validity of the STOC test suggested was evaluated in two stages. For the ***content validity***, in spring 2019, a group of four experts external to STIMEY project were asked to analyze the theoretical background and methodology of STOC testing to conclude, whether the STOC test suggested is a relevant tool regarding the purposes it was designed for. The expert panel included an expert in engineering, two experts in pedagogy (primary and secondary education) and an expert in psychology, three of four possessing Ph.D. degrees and all working in academic sphere. They represented respectively Spain, Greece, Finland, and Poland.

The expert panel concluded that the STOC testing methodology developed, with certain constraints, can serve as an instrument for measuring the quantitative parameters of creativity related to divergent thinking within the science and technology domain. In opinion of one expert, "the proposed method assesses in a clear and efficient way the personal parameters that the authors relate to a creative process of a scientific or technological type. For this purpose, it elaborates a set of indicators that reflect the creative activity of a group of students, within a defined procedural activity. Within this environment of development-creativity definition, the parameters and the way of assessing them seem correct and effective". The other remark was that "overall, it could be suggested that

Table 5: The comparison of the results of STOC test and Williams test.

Age group (years of age)	Country 1		Country 2		Country 3		Country 4	
	STOC non-standard answers ratio	Williams test average score	STOC non-standard answers ratio	Williams test average score	STOC non-standard answers ratio	Williams test average score	STOC non-standard answers ratio	Williams test average score
10-12	0,44	51	0,44	51	0,47	59	0,44	51
13-15	0,49	57	0,46	55	0,47	56	0,43	53
16-18	0,51	58	0,47	61	0,50	69	0,52	60

the procedures suggested are suitable to verify Science and Technology Oriented Creativity … with the establishment of some more indicators, specific answers related with analogies and assimilations, as strategies employed in creative thinking, could be also highlighted. Nonetheless, since the most important variables of creativity are considered for measurement, it can be assumed that the established measurements can indeed measure an increase in STOC creativity”.

To evaluate the ***construct validity*** of the results produced by the testing, we decided to perform STOC testing in small groups and follow up these tests with Williams creativity tests within the same groups. The scores, obtained for each of the tests in each of the age groups in 4 countries, were compared to calculate the correlation between them and to prove or to deny the hypothesis that the STOC test measures what it is intended to measure – the creative behavior. To prove the hypothesis, the correlation between the scores should be high, which means the correlation coefficient should be 0,5 or higher.

The Creativity Assessment Packet (Tunik, 2003) includes three sections, which are the divergent thinking test, divergent feeling test and the Williams scale which measures the perception of teachers regarding a learner’s creativity. In the validity test, we chose to use only the Williams scale. The validity tests were held in four countries of the consortium – Greece, Belarus, Germany, and Spain in spring 2019. For ethical reasons, the results of validity tests are coded, not relating the figures achieved with the exact countries where the validity testing took place.

All responses of the STOC test in each age group were grouped according to the following characteristics: standard, non-standard, functional, and senseless. The replies on the Williams’ test were evaluated according to the available rating scale. Each student in each age group scored their own, different from each other’s number of points. Both indicators are positive, which means the higher the score – the better the result. In the case of the STOC test and Williams test the dependence between the results is proven by the direct correlation.

According to the STOC test, in each age group, the percentage of non-standard answers (non-standard answers ratio) in the group was determined. For the Williams test, the average score for each age group was calculated. Table 5 contains the data on the Williams test average score and the STOC non-standard answers ratio of each group tested.

It can be seen that in almost every partner country, with an increase in the age group, the results both of average percentages and the average score of answers increase, which means that with increase in age, creativity is also increasing. Such dynamics are observed in both tests of creativity.

The statistical analysis, studying the correlation between the results of the STOC test and the Williams test performed, had shown multiple correlation coefficients R of 0,72 and R squared of 0,52, which means a rather tight connection between test outcomes. The graphic representation of data of two tests is shown in Figure 1.

The average scores for each age group of the Williams test are not significantly deviated from the non-standard answers ratio of the STOC test, which means: the results can be considered valid and consistent with the level of creativity of the students.

4 Experimental testing

The preliminary experimental testing was held during STIMEY LE piloting in 5 consortium countries in autumn 2019. To achieve enough data, some learner groups participated in experimental testing during the school year 2020-21, in the middle of COVID-19 pandemic. For this reason, the settings varied from traditional classroom learning to fully remote learning situations, and to hybrid settings in which sessions were facilitated remotely by the researcher, while teachers supported learners at school.

No personal data was collected in any of the countries where piloting took place. Each partner university for each age group (10-12, 13-15, 16-18 years of age) had formed

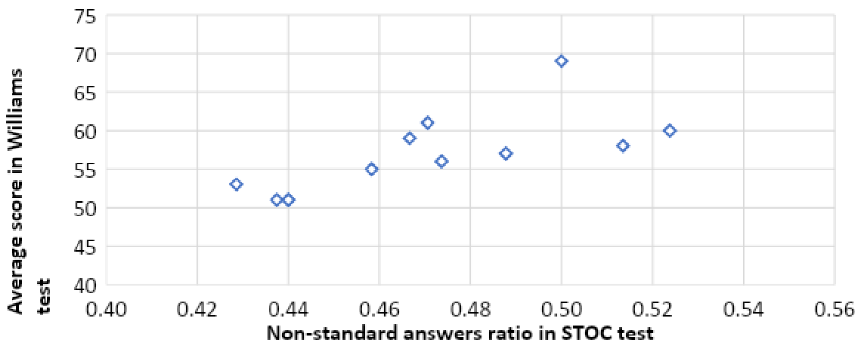

Figure 1: The validity tests data graphic representation.

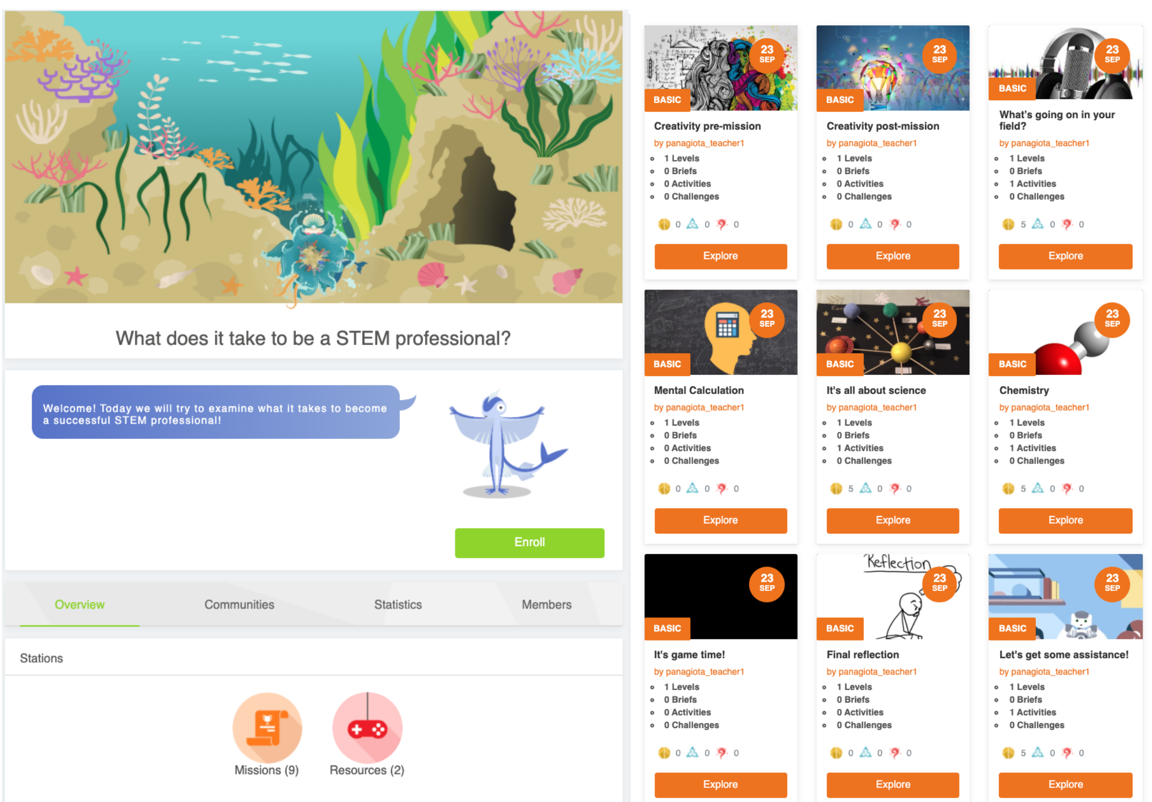

Figure 2: STIMEY world and its missions used in piloting.

two testing groups of learners. The first subgroup in each age group consisted of learners, who participated in STIMEY piloting, the second one – of learners, who did not participate in piloting. Learners in both subgroups possess similar or compatible levels of academic progress, which was assured by their teachers when assisting in forming the groups at schools during STIMEY piloting sessions. This was made for the elimination of maturation effect on the group creativity. Total number of participants in testing was 381, representing learners of the three age groups: 10-12 years of age (n = 123), 13-15 years of age (n = 110), and 16-18 years of age (n = 148) old learners.

Each group was tested twice: the first time before the piloting of STIMEY LE, the second time – after piloting. For the pilot, we created a “STIMEY world” (curriculum) named as “What does it take to be a STEM professional?”. In addition to pre- and post-STOC test package, it consisted of seven missions (lessons) related to all STIMEY LE components: radio (podcasts), mental calculation, chemistry and physics, activities with the STIMEY robot, serious games, and so on (See Figure 2). The duration of missions was varying from fifteen to forty-five minutes, depending on how much time teachers wanted to extend the work related to each mission. There was also some

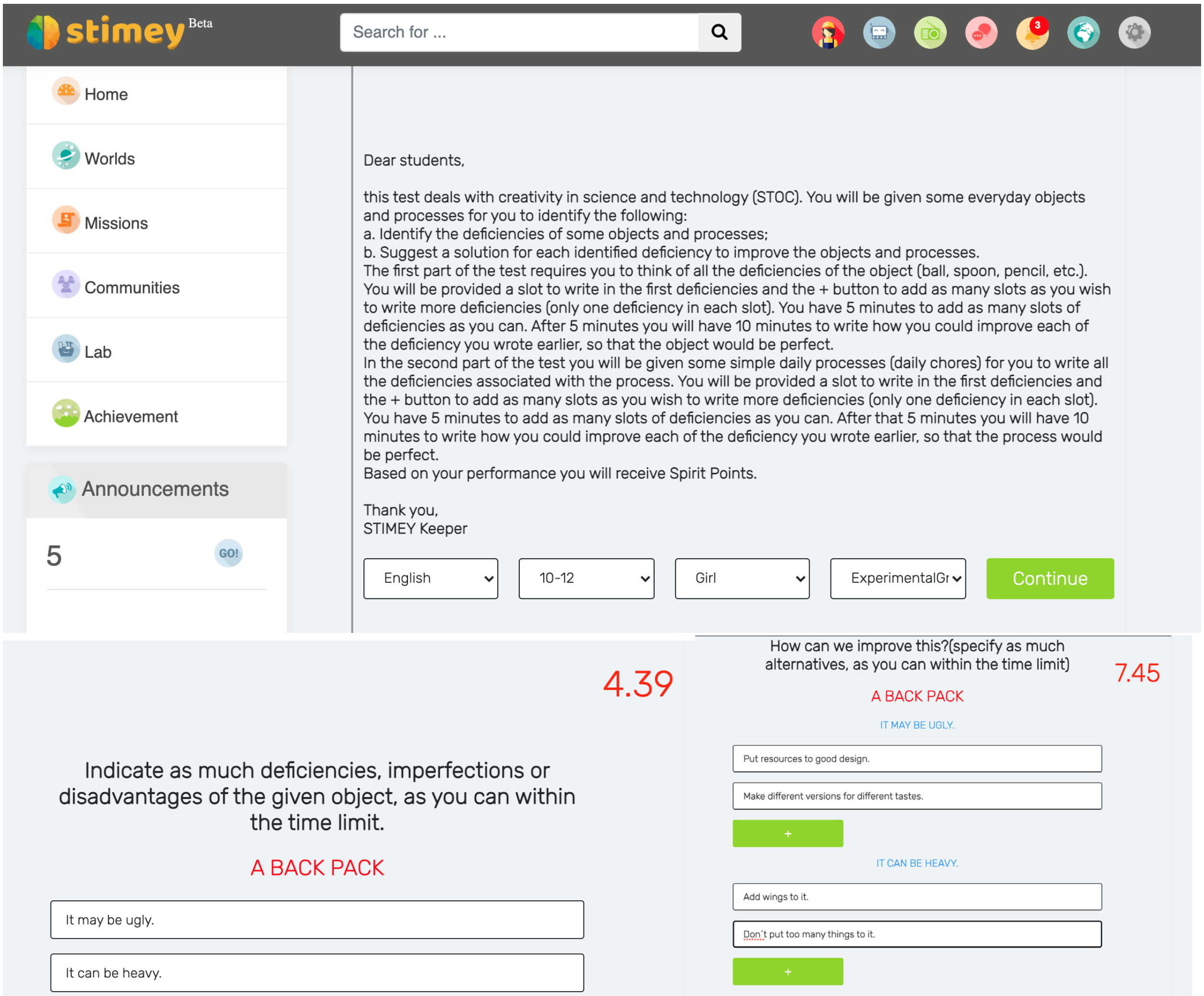

Figure 3: STOC test in STIMEY LE.

variation in how the sessions were conducted, either many activities during the same day or activities divided for various days. The control group did the pre- and post-test at the same time as the experimental group but in between they did their everyday school activities.

Participants were instructed (see Figure 3) that they would be given some everyday objects and processes for them to first identify the deficiencies of some objects and processes, and then, to suggest a solution for each identified deficiency to improve the objects and processes. It was explained that the first part of the test required them to think of all the deficiencies of the object (ball, spoon, pencil, etc.). They were provided a slot to write in the first deficiencies and the + button to add as many slots as they wished to write more deficiencies (only one deficiency in each slot). They had five minutes to add as many slots of deficiencies as they could. After the five-minute period they would have ten minutes to write how they could improve each of the deficiencies they mentioned earlier to make the object perfect. In the second part of the test, participants were given some simple daily processes (daily chores) to write down all the deficiencies associated with the process. This was followed by writing down the improvements, with the same logic and time limits as with the objects.

The results of testing were de-personalised, evaluated with the help of STOC Test Answers Evaluation Algorithm and compared. The preliminary experimental testing produced a total of 5929 responses. 3257 responses were collected in the experimental group and 2672 responses - in the control group. The results in the four types of non-standard answers ratios are presented in Table 6.

In three cases (out of 24, see Table 6) the control group in the post-test showed ratios lower, than in the pre-test. Namely, the control group of 13-15 years old learners in the post-test produced 25,8% less non-standard deficiencies

Table 6: The non-standard ratios for specific categories of responses.

	Pre-test		Post-test		Coefficient increment	
	Main group	**Control group**	**Main Group**	**Control group**	**Main Group**	**Control group**
Nonstandard object deficiencies ratio						
10-12 years	*0,0566*	*0,0680*	*0,1117*	*0,1076*	97,34%	58,16%
13-15 years	*0,0547*	*0,0918*	*0,0919*	*0,0682*	68,03%	-25,76%
16-18 years	*0,1463*	*0,1288*	*0,1789*	*0,1667*	22,22%	29,37%
Total	***0,0853***	***0,0980***	***0,1210***	***0,1198***	41,79%	22,19%
Non-standard process improvements ratio						
10-12 years	*0,0825*	*0,0971*	*0,1026*	*0,1196*	24,36%	23,15%
13-15 years	*0,0882*	*0,0351*	*0,1633*	*0,1111*	85,03%	216,67%
16-18 years	*0,1489*	*0,0985*	*0,2619*	*0,1452*	75,85%	47,39%
Total	***0,1058***	***0,0856***	***0,1672***	***0,1303***	58,05%	52,15%
Non-standard solutions to deficiencies ratio						
10-12 years	*0,0653*	*0,0805*	*0,0860*	*0,0769*	31,68%	-4,49%
13-15 years	*0,0698*	*0,0645*	*0,0765*	*0,1364*	9,65%	111,36%
16-18 years	*0,1229*	*0,1534*	*0,2177*	*0,1800*	77,12%	17,36%
Total	***0,0868***	***0,1062***	***0,1202***	***0,1312***	38,45%	23,60%
Non-standard solutions to improvements ratio						
10-12 years	*0,0440*	*0,0306*	*0,0803*	*0,0769*	82,66%	151,28%
13-15 years	*0,0439*	*0,2333*	*0,0693*	*0,2273*	58,02%	-2,60%
16-18 years	*0,0606*	*0,0227*	*0,0600*	*0,0465*	-1,00%	104,65%
Total	***0,0493***	***0,0690***	***0,0710***	***0,0876***	43,91%	27,09%

to the objects and 2,6% less non-standard solutions to process improvements. The control group of 10-12 years old learners produced 4,5% less non-standard solutions to object deficiencies.

At the same time, in four comparisons out of twelve the experimental group showed less progress, than the control group. Namely, the experimental group of 16-18 years old learners had -1% increment in non-standard solutions to process improvements, while the control group had 104,7% increase. The control group of 13-15 years old learners demonstrated higher non-standard answers ratios in process improvements (216,6% increment compared to 85,0% increment in experimental group) and solutions to object deficiencies (111,4% increment compared to 9,7% increment in experimental group). The control group of 10-12 years old learners was better in solutions to process improvements (151,3% increment against 82,7% increment for the experimental group). In the other eight comparisons experimental groups showed better progress. Some particularly high increments shown in post-test (151,3% or even 216,6%) can be explained, mainly, via extremely low results in pre-test and, so far, are considered as statistical randomness.

The key indicator to be evaluated was the total non-standard answers ratio. The results of its calculation are shown in Table 7.

The overall result is that this ratio was higher for learners of the older age groups. For the learners of 10-12 years of age 6-7% of answers were original in pre-test, and about 9,5% - in post-test. For the learners of 13-15 years of age 6-10% of answers were original in pre-test and 9-12% - in post-test. 16-18 years old learners have produced 10-12% original answers in pre-test and 13-18% original answers in post-test.

The results of experimental groups in post-test were in total 43,70% higher, than in pre-test; for the control groups the increment was 29,72%. The share of senseless answers is 4-6%. In the control group, the share of senseless answers was, in general, 1% higher in the control group than in the experimental group. The functionality of

Table 7: The non-standard answers ratio.

	Pre-test		Post-test		Coefficient increment	
	Main group	**Control group**	**Main Group**	**Control group**	**Main Group**	**Control group**
10-12 years	*0,0618*	*0,0704*	*0,0955*	*0,0955*	54,67%	35,67%
13-15 years	*0,0634*	*0,1006*	*0,0952*	*0,1245*	50,16%	23,72%
16-18 years	*0,1231*	*0,1051*	*0,1806*	*0,1368*	46,68%	30,17%
Total	***0,0827***	***0,0918***	***0,1189***	***0,1190***	43,70%	29,72%

answers is about 65-70%, the average functionality was higher in the experimental group, although for both experimental and control groups the functionality in post-test was better than in pre-test.

The progress in creativity in each subgroup may be explained through two main factors – maturation and the effect of teaching/learning methods applied. The difference in progress of two subgroups can be referred to as the influence of STIMEY on creativity.

In summary, the results of preliminary experimental STOC testing show that the progress of learners, having taken part in STIMEY piloting, in creativity was in general fourteen percentage points higher, than for those who did not participate in piloting activities. The key difference between experimental and control groups within the period between the pre- and post- tests in the experiment was the participation in STIMEY LE piloting, which may have resulted in producing more non-standard responses during the post-test. For improving the accuracy in measurements, further experimental testing would be crucial.

5 Discussion and conclusions

In this paper we first proposed a definition for Science and Technology Oriented Creativity (STOC) based on the previous literature (see Theoretical background) and particularly the process-based definition by Torrance (1966): a person is considered creative in science and technology if he or she is capable of, first, identifying a problem(s), second, finding out the causes of the problem(s), and, third, suggesting solutions to a problem(s). We also chose to focus on group creativity. In the case of the STIMEY project it is not the creativity of the teams (Nijstad & Stroebe, 2006; Paulus, 2000), but the aggregate creativity of group members. The STOC measurement method was then developed based on the work of authors such as Guilford (1950), Wallach and Kogan (1965), Torrance (1966), and Tunik (2003).

The strengths of the method described in this paper lie within certain criteria and specific guidelines, designed and developed in a form of evaluation algorithm based on the results from the experts in order for non-experts to execute the STOC Test effectively with very little personal interference in the evaluation of students' creativity as a group. Another strong point is the automatization of testing via a digital tool with carefully translated instructions, so that the possible misinterpretation is minimized. Finally, the fact that the very same group of experts performed the evaluation of each test brought certainty to the final results evaluation approach.

On the other hand, one of the key limitations of the STOC testing method outlined is that the originality of answers is judged via Expert Judgment Method, where each expert might have his (her) own approach, thoughts, feelings about the answer. Although, in many cases the non-originality of the answer is proven if the similar answers are given by the other test participants, a large set of answers has to be rated only due to the expert's personal approach. This is why it is important to take these results only as tentative. More iterations for developing the tool are needed to reach better accuracy of results.

Further, the answers from the five different countries were received in five different languages, none of which is English. The researchers translated the answers into English and submitted the translated data for the evaluation groups. Some patterns of the answers thus might have been lost or misinterpreted due to the translation. Further elimination of the limitations indicated is the task to be solved in further development of STOC testing methodology.

Key further direction of improvement of the STOC testing method lies within the automation of the evaluation with the help of BigData and AI instruments. The possible solution to the subjectivity of the evaluation technique is the development of an answer database with certain AI instruments for data processing. Using BigData to collect large quantities of responses to different objects and processes will allow to build a rather objective instrument, which could, first, collect an answer (deficiency or solution to deficiency of the object or process), second, correlate it with the other answers based on the key words or word combinations, third, calculate the number of similar responses, and fourth, calculate the percentage of the similar answers.

Furthermore, the evaluation mechanism could be modified: having a database with a relatively large number of responses would allow to compare each response analyzed to all other responses and verify quantitatively the level of its originality, simply by calculating the frequency of similar responses in it. This could result in an IT-tool, which could allow multipurpose testing, namely:

1. An individual could evaluate own creativity based on the answers, previously received from a large number of people;
2. Teachers could evaluate the creativity of learners and thus – the efficiency of teaching method(s) used.
3. The researcher (inventor, innovator, etc.) could examine the originality of his (her) idea.

Experimental STOC testing was conducted in varying settings (on-site, online, hybrid) suggesting that the method serves to measure STOC in different learning scenarios. At this point, we could not identify any relation between the results and the setting. This could, however, raise certain challenging questions for future research, such as: what are the differences in supporting science- and technology-oriented creativity in on-site, online and hybrid learning, how does the setting influence the development of science- and technology-oriented creativity as measured by the STOC test?

Funding information: The research, described in this article was funded by the European Union's Horizon 2020 Research and Innovation Programme within a framework of "Science, Technology, Innovation, Mathematics, Engineering for the Young (STIMEY)" project under Grant Agreement No. 709515.

Conflict of interests: Authors state no conflict of interests.

References

Amabile, T. (1983) The social psychology of creativity. Springer-Verlag New York.
Ankur, J., Saket, K., Satish, C. and Pal, D. K. (2015) Likert Scale: Explored and Explained. *British Journal of Applied Science & Technology 7(4): 396-403.*
Banaji, S., Burn, A., & Buckingham, D. (2006). *Rhetorics of creativity: a review of the literature.*
Centre for the Study of Children, Youth and Media, Institute of Education, University of London.
Barbot, B., Besancon, M., & I. Lubart, T. (2011). Assessing Creativity in the Classroom. The Open Education Journal, 4(1).
Benini, A., Chataigner, P., Noumri, N., Parham, N., Sweeney, J., & Tax, L. (2017): The Use of Expert Judgment in Humanitarian Analysis: Theory, Methods, Applications. *Geneva, Assessment Capacities Project - ACAPS*
Boden, M.A. (1998), Creativity and artificial intelligence. *Artificial Intelligence, No. 103: 347-356.*
Corazza, G.E. (2016). Potential Originality and Effectiveness: The Dynamic Definition of Creativity. *Creativity Research Journal, 28(3): 258–267.*
Csikszentmihalyi, M. (2014). The Systems Model of Creativity: The Collected Works of Mihaly Csikszentmihalyi. Dordrecht: Springer.
Dusek, V. (2006) Philosophy of technology: an introduction. Oxford: Blackwell Publishing.
Edwards-Schachter, M., García-Granero, A., Sánchez-Barrioluengo, M., Quesada-Pineda, H., & Amara, N. (2015). Disentangling competences: Interrelationships on creativity, innovation and entrepreneurship. *Thinking Skills and Creativity*, *vol. 16: 27-39*. https://doi.org/10.1016/j.tsc.2014.11.006
Guilford, J.P. (1950). Creativity. *American Psychologist vol. 5(9): 444-454.*
Henriksen, D., Henderson, M., Creely, E., Ceretkova, S., Černochová, M., Sendova, E., … Tienken, C. H. (2018). Creativity and Technology in Education: An International Perspective. *Technology, Knowledge and Learning, 23(3): 409–424.*
Hollanders, H., & van Cruysen, A. (2009). *Design, creativity and innovation: A scoreboard approach.* Maastricht, Netherlands: PRO INNO Europe
Kaufman, J.C. & Sternberg, R.J. (2010), The Cambridge Handbook of Creativity. *Cambridge University Press: 27-28.*
May, R. (1975) The Courage to Create. NY: Bantam Books.
Meyer, A.M., Booker, J.M. (2001). Eliciting and Analyzing Expert Judgment: a practical guide. *Soc. Industrial & Applied Mathematics, Washington, DC.*
Mumford, M. D. (2003). Where have we been, where are we going? Taking stock in creativity research. *Creativity Research Journal, vol.15, iss.2-3: 107–120.*
Nijstad, B.A., & Stroebe, W. (2006). How the group affects the mind: A cognitive model of idea generation in groups. *Personality and Social Psychology Review, vol. 10: 186–213.*
OECD (2019). Future of Education and Skills 2030. Learning Compass 2030. http://www.oecd.org/education/2030-project/contact/OECD_Learning_Compass_2030_Concept_Note_Series.pdf
Paulus, P.B. (2000). Groups, teams, and creativity: The creative potential of idea-generating groups. *Applied Psychology: An International Review, vol. 49: 237–262.*
Runco, M. A., & Jaeger, G. J. (2012). The Standard Definition of Creativity. *Creativity Research Journal*, *vol. 24(1): 92–96.*
Salmon, M.H. [et. al.] (1999) Introduction to the philosophy of science. Indianapolis: Hackett.
Sassenberg, K., & Moskowitz, G. (2007) Unconscious Goals and Creativity: Activating Creativity Goals Breaks Established Associations and Leads to the Generation of Original Ideas. *Advances in Consumer Research, vol.34: 270-275*
Simonton, D. K. (2004) Creativity in science: Chance, logic, genius, and zeitgeist. Cambridge University Press.
Sternberg, R.J. (2011). Cognitive Psychology (6 ed.). Cengage Learning. – Chapter 11 – Problem Solving and Creativity.

Szabó, T.P., Fenyvesi, K., Soundararaj, G. and Kangasvieri, T. (Eds.) (2019). *Everyday Creativity: Boosting Creative Resources with Finnish Models of Education. Teachers' Handbook.* University of Jyväskylä.

Torrance, P. (1966) "Verbal Tests. Forms A and B-Figural Tests, Forms A and B.". The Torrance Tests of Creative Thinking-Norms-Technical Manual Research Edition. Princeton, New Jersey: Personnel Press.

Tunik, E. (2003) Modified Williams creativity tests. *Saint Petersburg: Rech'.*

Vygotsky, L.S. (2004). Imagination and Creativity in Childhood. *Journal of Russian & East European Psychology, vol. 42(1): 7–97.*

Wallach, M., & Kogan, N. (1965) Modes of thinking in young children. New York: Holt, Rinehart & Winston.

Wyse, D., & Ferrari, A. (2015). Creativity and education: comparing the national curricula of the states of the European Union and the United Kingdom. *British Educational Research Journal, vol. 41(1): 30–47.* https://doi.org/10.1002/berj.3135