

Chapter 5

Facilitating the Practice of 4C Skills in Biology Education Through Educational Escape Rooms



Georgios Villias and Mark Winterbottom

5.1 Introduction & Theoretical Background

Educational escape rooms (EERs) are game-based activities that adopted the initial concept from the escape room industry and adapted it appropriately for use in an educational context. By offering immersive experiences that promote students' active participation, EERs have emerged as promising alternative approaches to fostering students' conceptual and skill-based learning (Nicholson, 2018). A deeper look inside the design principles and the conceptual framework of an EER reveals their connection to several well-established educational methodologies (e.g., problem-based, inquiry-based, experiential, game-based, narrative-based, etc.), as well as motivational theories (e.g., self-determination theory, self-efficacy theory, etc.), justifying their learning potential. EERs have been widely used with the intention to create an active learning environment, motivate students, simulate conditions from real-life scenarios for health-care professionals, and facilitate the learning of content knowledge and the development of 21st century skills (Lathwesen & Belova, 2021; Taraldsen et al., 2022; Veldkamp et al., 2020a, b).

The popular and overarching term '21st century skills', refers to a set of skills and competencies (learning, social and cultural, life and career, literacy, etc.) that are considered to be of vital importance for both present and future, in order to confront the challenges, introduced to the new generation with the arrival of the third millennium (Bapna et al., 2017). Many attempts have been made by various initiatives and educational organisations to outline and classify the most important 21st century skills, resulting in several different skill frameworks. Among them are included frameworks from the P21: Partnership for 21st Century Learning (P21, 2009),

G. Villias (✉) · M. Winterbottom (✉)
Faculty of Education, University of Cambridge, Cambridge, UK
e-mail: gv283@cam.ac.uk; mw244@cam.ac.uk

the Assessment and Teaching of 21st Century Skills – Cisco/Intel/Microsoft (<http://www.atc21s.org/>), the World Economic Forum, and OECD initiatives (DeSeCo and PISA). Although there are many overlapping competencies in these frameworks, there is a lack of consensus between them about which 21st century skills should be regarded as the most essential. Depending on their focus (e.g. studies, work, literacy, social life etc.), they categorise skills differently (Ananiadou & Claro, 2009; Bapna et al., 2017; Bialik et al., 2015; Lai & Viering, 2012). Bialik et al. (2015, p. 3) compared these frameworks to identify their commonalities and concluded that the learning skills which are present in most of them are the following four: Critical thinking, Creativity, Collaboration and Communication (4Cs).

The 4Cs seem to have a central role in modern proposals for 21st century curriculum re-design (Fadel et al., 2015) and school networks for 21st century learning (e.g. EdLeader21). Each of these skills encompasses different performance areas (abbreviations in brackets) that should be considered when trying to measure or develop them. Critical thinking, according to Sternberg (1986, p. 3), “comprises the mental processes, strategies, and representations people use to solve problems, make decisions, and learn new concepts.” These include the ability to discover information (IFD), interpret and analyse collected data (IPA), make and support claims with valid reasoning (RES), and propose adequate, applicable solutions when needed (PRB). Regarding Creativity, Sternberg and Lubart (1998, p. 3) define it as “the ability to produce work that is both novel (i.e. original, unexpected) and appropriate (i.e. useful, adaptive concerning task constraints).” That means being able to brainstorm and generate ideas (IDG), articulate and refine these ideas (IDR), but also to effectively select and integrate materials to develop a unique product or finish a specified task (CPI). Collaboration is a “coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem” (Roschelle & Teasley, 1995, p. 70). While collaborating, individuals are expected to take initiatives or even lead the group (LDI), follow appropriate norms, avoid conflicts, and share their insights (CPF), be responsible and productive (RPR), and show responsiveness to others through the provision and acceptance of feedback (RSP). Finally, Communication is described by Qian and Clark (2016, p. 51) as “the ability to articulate thoughts and ideas in a variety of forms, communicate for a range of purposes and in diverse environments, and use multiple media and technologies.” Engaging in fruitful conversations and discussions entails the use of verbal and non-verbal language, empathy, conveying of emotion, and consensus building (ENG).

Over the past five years, educational research on EERs has been gaining momentum. Researchers’ increasing interest in this trending phenomenon is reflected in the growing number of papers being published each year (Fotaris & Mastoras, 2019). Most of these studies were conducted on university level students, coming from several different fields (STEM, medicine, nursing, computer science). Nevertheless, a great number of STEM-oriented EERs have also been developed, implemented, and studied in secondary schools (Lathwesen & Belova, 2021). Knowledge about the structure and design of EERs is accumulating fast since their first appearance, leading to the development of several design frameworks to guide the early adopters

(Clarke et al., 2017; Fotaris & Mastoras, 2022; Nicholson, 2018; Nicholson & Cable, 2021; Veldkamp et al., 2020a, b). EERs' educational and design aspects have been systematically reviewed (Veldkamp et al., 2020a, b), offering insights into their learning mechanism. Nevertheless, evidence-based research on EERs' learning effectiveness has still not provided conclusive results. Several researchers have assessed students' cognitive (e.g., content knowledge, skills) or affective (e.g., motivation, interest, engagement) learning outcomes after their participation in EERs. The research designs they used range from quantitative (pre-/post-participation surveys and tests), to qualitative (interviews, informal feedback and observations), and sometimes mixed methods (Taraldsen et al., 2022). However, there are some issues in reference to the applied research methodology. Only a few of these researchers have actually adopted a control and treatment group design. Fotaris and Mastoras (2019) also highlight the importance of using larger sample sizes to avoid questioning of these studies' results. According to Lathwesen and Belova (2021, p. 10), there is not enough empirical evidence on the short and long-term effectiveness of these activities in comparison to the traditional, lecture-based approach. As they stress, "to research whether escape games have a long-lasting learning effect, multiple post-tests need to be undertaken at different time intervals." Taraldsen et al. (2022) also acknowledge the need for systematic, longitudinal studies in primary and secondary schools. They suggest that researchers should adopt more complex research designs to evaluate EERs' learning gains, emphasising 21st century skills and school subjects' content learning.

Attempting to bridge the gap in knowledge presented above, this study examined EERs' learning impact on secondary school students, focusing on the practice and development of their 21st century skills. The study's main research questions are the following:

- (a) What 4C skills are practised by students when engaging in an EER? Do they develop?
- (b) What structure and features of an EER enable students to practise and possibly develop their 4C skills?

Some of the present study's characteristics and the added value it brings in this particular field of educational research are outlined below:

- adopts a mixed-method research design; thus, it capitalises on the benefits from both quantitative and qualitative approaches.
- uses several different data collection and analysis methods; thus, it facilitates the validation of its data through triangulation.
- is applied on a large number of students from different schools; thus, it increases the findings' reliability.
- is based on a control and treatment group design; thus, it evaluates EERs' learning impact on students' 21st century skills in comparison to other didactic approaches.
- is longitudinal with multiple data collection points; thus, it offers insights into the long-term effectiveness of EERs.

- approaches knowledge holistically; thus, it combines information provided by different sources (e.g. interviews, tests, observation, questionnaires) and different perspectives (e.g. student, facilitator, researcher).
- uses a design-based research methodology; thus, it reviews and optimises the developed EERs' design during its three successive iterations.
- provides design guidelines for developing puzzles which promote the practice and development of 4C skills; thus, it offers workable solutions and ideas to education practitioners for biology teaching and learning.

5.2 Materials and Methods

5.2.1 *Design of the Study and Data Collection*

The work presented here is the result of a nine-month, longitudinal, mixed-method research study that adopted a design-based research methodology and followed an iterative process of three meso-cycles. Three EER interventions (EER1, EER2, EER3) were designed and developed according to a series of criteria that were consistent with the students' age group, expected cognitive abilities, pre-existing knowledge, but also the lessons' content, learning objectives and connection to common misconceptions. These three EER activities were embedded in the teaching of the biology science course, as part of the normally expected-to-be-taught Greek national curriculum and lasted between 45 and 60 min. Even though the content knowledge of these EER activities was different, all of them retained the same focus on developing students' 4C skills and thus they are connected. Each EER activity was built upon the previous one. Therefore, all the theoretical and practical findings that were produced during the empirical micro-cycles that preceded (Analysis/Exploration) and followed (Evaluation/Reflection) the implementation of the first EER activity acted as layers that informed and added value to the respective stages of the next meso-cycle (second EER). The same approach was applied to the third meso-cycle (third EER), as well. A total of 209 Year 10 students, from three different Greek (main, secondary, and back-up unit) schools and one Cypriot school (pilot unit) divided into ten different classes, participated in the study. However, we only conducted an in-depth analysis of 125 full datasets that were collected from students enrolled in the main and secondary school unit. Acknowledging the importance of using a control group to contradict plausible counterfactual inferences as argued by Marsden and Torgerson (2012) led us to the adoption of a control group research design. An effort was also made so that both groups were comparable in terms of student socioeconomic background, gender ratio and prior performance in science. Nevertheless, in all schools, students' allocation to classrooms was non-random, but based on criteria predefined by the schools' administrations.

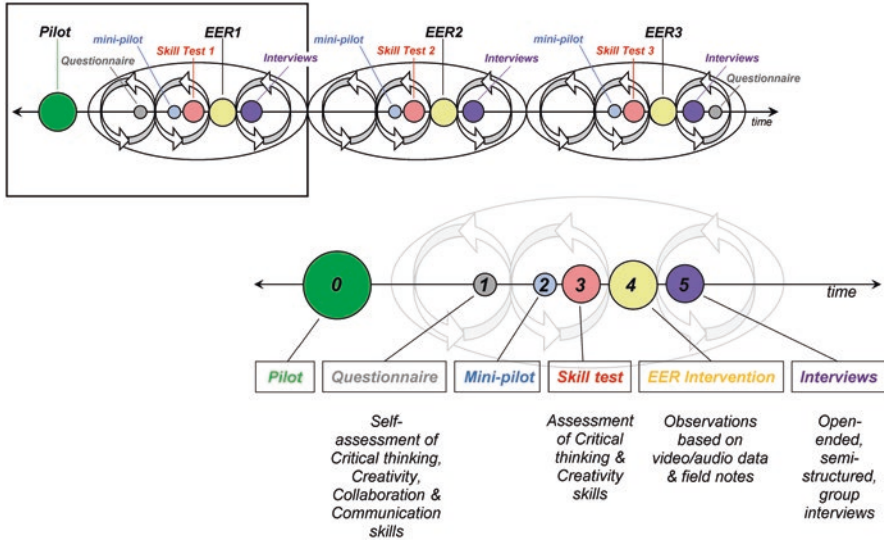


Fig. 5.1 The overall research plan of this study and data collection methods as applied on the first meso-cycle

Several different data collection methods were employed, including questionnaires, skill tests, observations, as well as interviews (Fig. 5.1). Considering that each of these methods has inherent weaknesses that may affect the data's validity and provide inaccurate results (e.g. observer bias, lack of recording equipment, omission of participants' less noticeable behaviour, difficulty in analysing collected data, interviewees' introversion or reluctance to answer, skills being overrated, poor self-awareness, miscomprehension of questions), we did not rely exclusively on one of the aforementioned data collection methods, but we triangulated the findings of many, so as to have an objective and unbiased assessment of participants' skills. The main data collection method that we applied during the implementation of the EERs was the use of video/audio recordings. For each implementation, there were four groups of students, consisting of between three and six, dependent on the group, working on large benches simultaneously.

Students were usually standing or sitting around the bench, with all the resources needed in the middle. Using 360-degree cameras, placed on the students' benches, we collected 88 video/audio recordings (each video's average duration: 40 min). Field observations were also collected, assessing participants' performance from the facilitator's perspective. Regarding interviews, we conducted six informal, exploratory interviews to assess and optimise the EERs' design (after the mini-pilots), as well as 15 semi-structured, in-depth, group or small-group interviews to gain insights into the structure, design, and learning impact of EERs, from the perspective of 33 students (after the interventions). In reference to the study's written

assessment tools (tests, questionnaires), we constructed and validated our own. Three skill assessment tests that contained different but equivalent items were designed to measure and compare objectively students' critical thinking and creativity skills. Critical thinking test items were based on the format of standardised situational judgement tests (SJTs), while for creativity we used items adapted from a standardised scientific creativity test, developed for secondary school students (Hu & Adey, 2002). Finally, at the beginning and at the end of the research study, we administered a 7-point Likert scale and self-assessment questionnaires, in order to measure students' self-perception of all their 4C skills.

5.2.2 *Qualitative Data Analysis*

A major part of the analysis was devoted to observational data from video/audio recordings. Despite initially collecting a vast amount of data, less than half of them were analysed in-depth. Nevertheless, the student groups whose recordings were not analysed exhibited the same patterns of behaviour in the 4C skills practice. In order to analyse these rich data in a methodical and efficient way, we used a set of rubrics designed for a performance-based assessment of the 4Cs. These rubrics were revised and customised according to the EERs' special design features. Their two-dimensional design consisted of a vertical axis which evaluated the sub-categories of different performance areas of each skill, and a horizontal, 4-level, performance rating scale that evaluated more accurately the mastery level of each sub-category (Fig. 5.2).

Based on the revised 4C skills rubrics' design presented above, we developed an analytical coding framework of 12 double indicators. Each of them corresponded to one of the 4C skills performance areas mentioned in the introduction section (i.e. IFD, IPA, RES, PRB, IDG, IDR, CPI, LDI, CPF, RPR, RSP, ENG). The aforementioned indicators were used to code students' practice of the 4C skills during the observational analysis of the video/audio recordings. Adopting this strategy of double indicators allowed us to have a coding scheme that was flexible enough to describe observational variability. At the same time, we restrained the number of its basic codes to an operational degree (Table 5.1). This double system also allowed us to conduct a 'two-axes' observational analysis. The first 'axis' analysed the skills' frequency of appearance, while the second one described the observed progress of the skills' level of practice. Considering qualitative data analysis tools, we used version 22 of ATLAS.ti software. Its powerful analysis tools helped us to calculate each code's absolute frequency, and focus on the co-occurrence of codes for specific skills and puzzle designs. Coloured heat maps were used to visualise these quantified observational data for each EER's puzzle and showcase their ability to facilitate the practice of certain aspects of the 4C skills (Table 5.2).

As regards observational data from field notes and interview data, they were also systematically analysed after firstly being classified into categories, based on their content (structure of the game, teams' operation, acquisition of knowledge).

COMMUNICATION rubric for EER activities (long version)				
Engag	1	2	3	4
Performance Area				
COLLABORATION rubric for EER activities (long version)				
Conte	1	2	3	4
Performance Area				
CREATIVITY rubric for EER activities (long version)				
Leader	1	2	3	4
Performance Area				
CRITICAL THINKING rubric for EER activities (long version)				
Speak	1	2	3	4
Performance Area				
Information Discovery (IFD)				
Facilitator	1	2	3	4
Performance Area				
Cooper	1	2	3	4
Performance Area				
Contrib	1	2	3	4
Performance Area				
Conflict	1	2	3	4
Performance Area				
Using & Resour	1	2	3	4
Performance Area				
Problem Solving (PRB)	1	2	3	4
Performance Area				
Proposing solutions (Quality or Quantity)	1	2	3	4
Performance Area				
Evaluating solutions (Application)	1	2	3	4
Performance Area				

IFD - 4

In-depth understanding. Able to describe it and examine it through multiple perspectives. Generates thought-provoking inquiry questions. Good phrasing that provides quality and value. Gathers information of adequate depth, high quality and great value for the specific inquiry questions.

Fig. 5.2 Connection of a 4C skills rubrics' design to the analytical coding framework for EER activities (IFD-4 corresponds to Information Discovery - Level 4)

Table 5.1 Double indicators for observational data analysis on 4C skills

			Indicators' second part (xxx-X)				
			Level of Skill mastery				
Performance area			1	2	3	4	
			Poor level	Average level	High level	Expert level	
Indicators' first part (XXX-x) Performance area	CT	IFD	Information discovery	IFD-1	IFD-2	IFD-3	IFD-4
		IPA	Interpretation & Analysis	IPA-1	IPA-2	IPA-3	IPA-4
		RES	Reasoning	RES-1	RES-2	RES-3	RES-4
		PRB	Problem-solving	PRB-1	PRB-2	PRB-3	PRB-4
	CR	IDG	Idea generation	IDG-1	IDG-2	IDG-3	IDG-4
		IDR	Idea refinement	IDR-1	IDR-2	IDR-3	IDR-4
		CPI	Creative production & Innovation	CPI-1	CPI-2	CPI-3	CPI-4
	CB	LDI	Leadership & Initiative	LDI-1	LDI-2	LDI-3	LDI-4
		CPF	Cooperation & Flexibility	CPF-1	CPF-2	CPF-3	CPF-4
		RPR	Responsibility & Productivity	RPR-1	RPR-2	RPR-3	RPR-4
		RSP	Responsiveness	RSP-1	RSP-2	RSP-3	RSP-4
	CM	ENG	Engaging in discussions	ENG-1	ENG-2	ENG-3	ENG-4

CTCR critical thinking & creativity, *CB* collaboration, *CM* communication, *Dark colour* initial questionnaire, *Light colour* final questionnaire, *Purple* experimental group, *Green* control group

Table 5.2 Cross tabulation of 4Cs' performance area codes per puzzle in EER1

	P1	P2	P3	P4	P5	P6
IFD	(15)	(15)	(36)	(12)	(31)	(33)
IPA	(6)	(2)	(5)	(5)	(20)	(12)
RES	(2)	(21)	(8)	(8)	(23)	(36)
PRB	(8)	(8)	(16)	(6)	(21)	(16)
IDG	(8)	(4)	(19)	(4)	(10)	(19)
IDR	(5)	(7)	(8)	(1)	(5)	(8)
CPI	(8)		(11)	(9)	(5)	(22)
LDI	(8)	(9)	(9)	(6)	(9)	(13)
CPF	(16)	(11)	(22)	(7)	(18)	(28)
RPR	(11)	(2)	(4)	(8)	(6)	(7)
RSP	(3)	(13)	(7)	(1)	(12)	(9)
ENG	(15)	(15)	(20)	(11)	(26)	(27)

IFD information discovery, *IPA* interpretation & analysis, *RES* reasoning, *PRB* problem solving, *IDG* idea generation, *IDR* idea refinement, *CPI* creative production & innovation, *LDI* leadership & initiative, *CPF* cooperation & flexibility, *RPR* responsibility & productivity, *RSP* responsiveness, *ENG* engagement in discussions, *P1-P6* puzzles

5.2.3 *Quantitative Data Analysis*

After calculating the numerical scores of the study's written assessments (skill tests, questionnaires), we analysed them using descriptive statistics in SPSS statistical software (version 28). Regarding the skill assessment tests, for each single test item we calculated the responses' mean score and standard deviation. These measures were also calculated for the tests' sub-scores of critical thinking, creativity and originality. Apart from the two basic groups, i.e. the experimental and the control group, during our quantitative data analysis, we divided students into more groups, selecting them based on several other criteria (by school, by gender, by game performance) and compared their responses. Depending on the responses' distribution, parametric or non-parametric tests were applied so as to check for statistically significant changes between the items' ratings. In reference to the Likert-scale questionnaires, we followed the same analytical process, using different measures of central tendency and dispersion (median, mode, interquartile range).

5.3 Results

5.3.1 *Observational Data*

In order to explore which of the 4C skills were practised more by participating students during each puzzle of the three developed EER activities, we applied a code co-occurrence analysis tool to their coded data. Using data from several implementations each time (9 for EER1, 10 for EER2, and 8 for EER3), we calculated the absolute frequencies of all the skills' codes per puzzle and EER, i.e. the aggregating scores of the codes' combined presence. Visualising these quantified data with three coloured heatmaps facilitated us to detect similarities, differences or repeating patterns among these data, allowing us to associate the practice of specific 4Cs skills with the puzzles' properties (i.e. duration, type, difficulty) and their unique design features (i.e. provided resources, required tasks, complexity).

Analysing these heatmaps (Table 5.2), it became clear that some puzzles favoured more the practice of specific 4C skills compared to others (the higher the frequency of practice, the darker the heatmap's boxes). *Critical thinking* skills were mostly promoted by puzzles that contained several items or offered multiple options that students had to select from. The resources' complexity, i.e. having different information in one place that needed to be combined, also activated students' analytical thinking. By linking puzzles to challenging parts of the syllabus, where students already had some prior knowledge, it motivated them to make claims and use argumentation to support them. Last but not least, puzzles that did not offer straightforward solutions troubled students and forced them to apply different problem-solving strategies in order to come up with a solution. *Creativity* skills were practised less by students compared to the other 4C skills, either because of students' difficulty to express them in such time-pressuring activities, or because of the inappropriate

design of the developed EERs' puzzles. What our analysis showed is that puzzles that included hidden elements (e.g. symbols, letters), or coded messages that were difficult to decipher, ignited students' divergent thinking and encouraged their brainstorming. *Collaboration* and *Communication* skills, as expected, were dominant and inextricably linked with each other during these team-based activities. Easy puzzles or puzzles with a very limited number of items were often solved by a single individual; thus they did not favour these skills. On the contrary, puzzles of medium to high difficulty, that consisted of a considerable number of resources, involved the majority of students in solving them. Students collaborated actively by sharing their ideas, undertaking certain tasks, and contributing to the team through independent and team work. In general, puzzles of higher difficulty that troubled students for a greater amount of time resulted in a broader practice of all 4C skills.

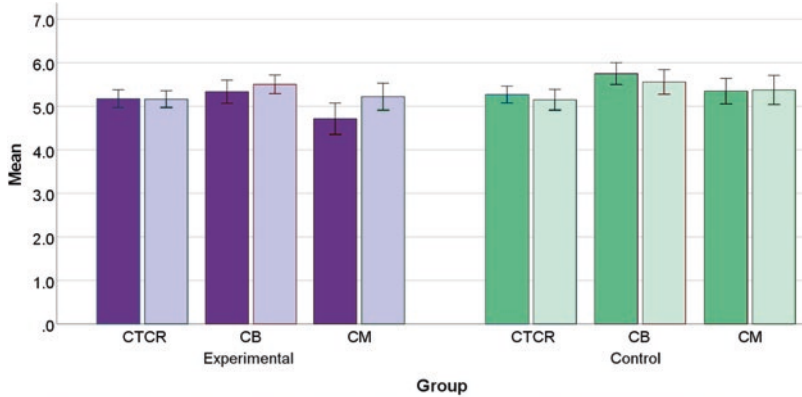
Another interesting finding emerged after comparing students' game performance between the first two EER activities. The relative frequency of codes' labels that corresponded to higher levels of demonstrated skill mastery, slightly increased in activity EER2 compared to EER1; the communication, collaboration, and critical thinking skills in particular. That was also inferred from descriptive evaluations that showed improved performance both at a team level (8 out of 12 teams), as well as at an individual level (15 out of 62).

5.3.2 *Self-Assessment Questionnaires*

Two identical self-assessment questionnaires were administered to the study's participants, one at the beginning and the other at the end of a six-month period. 100 students answered both questionnaires. A comparison of their scores revealed that the experimental group students, compared to the control group students, improved significantly in terms of their communication skills (CM: 4.715 \rightarrow 5.221, $t(47) = -3.157, p = 0.003$) (Fig. 5.3).

5.3.3 *Skill Assessment Tests*

Considering that the experimental group students participated in all three EERs, while the control group students did not participate in EER1, we focused our test scores' comparison mainly on tests ST1 (pre-test) and ST2 (post-test), and less on test ST3 (delayed post-test). For this comparison, we selected only those students that had participated in at least two EER activities and had fully completed all three skill assessment tests (80 students in total). In both groups, students' test scores fluctuated in a similar manner (Table 5.3). Critical thinking scores initially decreased (ST1 \rightarrow ST2) and increased afterwards (ST2 \rightarrow ST3). The exact opposite happened with their creativity scores. While analysing different student groups' test scores, we observed the same pattern occurring repeatedly, irrespective of the applied



CTCR critical thinking & creativity, CB collaboration, CM communication, Dark colour initial questionnaire, Light colour final questionnaire, Purple experimental group, Green control group

Fig. 5.3 Bar charts of the responses’ means (7-point Likert scale) for all scale items that derived from students’ initial and final questionnaires (by group)

Table 5.3 Descriptive statistics and statistically significant differences (SSD) between the main overall scores of students’ skill assessment tests (by group)

All 3 tests	Experimental group School A + B (n = 40)			* (p ≤ 0.05) ** (p ≤ 0.01)		
	ST1	ST2	ST3			
	Mean & Observed changes			ST1-ST2	ST2-ST3	ST1-ST3
Critical Thinking	49.90 ⁿ	42.20 ⁿ	42.60 ⁿ	**	No SSD p = 0.784	**
	↓ 7.7 / 15.4%		↑ 0.4 / 0.9%	t(39) = 7.240 p < 0.001		t(39) = 5.730 p < 0.001
Creativity	42.40	58.00	50.53 ⁿ	**	*	**
	↑ 15.6 / 36.8%		↓ 7.5 / 12.8%	Z = - 4.286 p < 0.001	Z = - 2.307 p = 0.021	Z = - 2.877 p = 0.004

All 3 tests	Control group School A + B (n = 40)			* (p ≤ 0.05) ** (p ≤ 0.01)		
	ST1	ST2	ST3			
	Mean & Observed changes			ST1-ST2	ST2-ST3	ST1-ST3
Critical Thinking	51.80 ⁿ	40.20	44.00 ⁿ	**	*	**
	↓ 11.6 / 22.4%		↑ 3.8 / 9.5%	Z = - 5.316 p < 0.001	Z = - 2.536 p = 0.011	t(39) = 6.488 p < 0.001
Creativity	53.43 ⁿ	62.75	53.03 ⁿ	**	**	No SSD
	↑ 9.3 / 17.4%		↓ 9.7 / 15.5%	Z = - 3.009 p = 0.003	Z = - 3.159 p = 0.002	p = 0.887

n normally distributed data, ST1 1st skill assessment test, ST2 2nd skill assessment test, ST3 3rd skill assessment test, No SSD no statistically significant differences

criteria (e.g. by school, by group, by gender, by performance). Therefore, we concluded that the selected skill tests' items were not as equivalent in terms of difficulty as we had initially thought. In order to practically cancel the items' inequality and overcome this problem, we did not focus on the exact scores, but we calculated and compared the tests' score difference and the percentage of the observed change. We observed that the experimental group's students had a much greater average score improvement compared to their peers from the control group (+15.6 vs +9.3 points, or + 36.8% vs +17.4%, respectively). Students' participation in activity EER1 was the only independent variable that changed in that case, suggesting that the observed improvement in their ST2 creativity scores is somehow related.

5.3.4 Interviews

Based on their personal experience, interviewed students stated that participating in EER activities required them to have a cooperative disposition, to show empathy, to respect others' opinions and avoid conflicts (collaboration skills), but also to establish good communication for the sharing of information and the exchange of views and ideas (communication skills). They also acknowledged the importance of being observant, having ingenuity (critical thinking skills) and using divergent thinking (creativity). When asked if their participation in the EERs facilitated them to develop any of these skills, they claimed that they did not notice any significant change. However, it is worth noting that several of them found themselves to be more alert, more observant, and able to work more efficiently with their teammates after the first EER activity.

5.4 Discussion

5.4.1 Connection Between Puzzle Types and the Practice of 4C Skills

During this study's three EER activities, we tested in total fifteen different puzzles that were developed based on ten different puzzle designs. Our study's observational data analysis revealed that each of these designs facilitated the practice of 4C skills, to a greater or lesser extent, dependent on their unique features. Trying to make inferences based on the broader categories that these puzzle designs belonged to, we associated them with specific puzzle types (word, observation, logic, deduction, cryptography and meta-puzzles), or combinations of them, as presented in the work of Nicholson and Cable (2021). Word puzzles (crossword), combined with text-based resources, facilitated students' text data-mining and collaboration skills. Observation puzzles (pattern recognition, image datasets) were usually enriched with multiple items or other resources, favouring the active involvement and

collaboration of most team members. Logic puzzles (matching up items) encouraged the practice of several critical thinking skills, with an emphasis on reasoning. Deduction puzzles (narrative-based questions) and cryptography (encrypted or hidden information) ignited learners' creativity and divergent thinking. Finally, meta-puzzles usually required and fostered the practice of all the 4Cs.

5.4.2 *Design Guidelines for EERs that Foster the 4Cs*

The learning outcome that an EER activity is capable of delivering depends greatly on its overall design. Since education practitioners develop EERs with different intended goals (e.g. content knowledge, skills, motivation, interest, engagement), applying on each of these occasions a specialised design framework, appropriately adapted to their distinct learning objectives, could benefit more of their learners. Taraldsen et al. (2022, p. 9) stressed in their review article that “researchers and educators have started to look for frameworks for designing escape rooms for educational purposes and for evaluating both 21st century skills and subject matter competence on an individual level.” According to one of the latest and most complete design frameworks proposed by Fotaris and Mastoras (2022), there are several elements and parameters that need to be considered when designing an EER activity (e.g. demographic information about the participants' background, skill level, needs, and motivation; the activity's goal, learning objectives, constraints, required knowledge, group size, game type, playtime length, curriculum position, theme, setting, narrative, characters, puzzle types, puzzle designs, puzzle path, game flow, game assets, room layout, hint system, scoring system, introduction, rules, and reflection). Exploring thoroughly some of the features mentioned above, reflecting on the insights which we gained from the findings of the present research study, and building upon design guidelines provided by previous studies (Clarke et al., 2017; Fotaris & Mastoras, 2019, 2022; Nicholson, 2018; Nicholson & Cable, 2021; Veldkamp et al., 2020a, b, 2022), we recommend the following set of practical guidelines for the design of EERs that focus both on the development of participants' 4C skills and content learning:

1. *forming teams of four or five members.* Teams of that size collaborate more effectively during gameplay and offer their members adequate opportunities to access the puzzle resources, engage actively and learn.
2. *using escape boxes.* This particular game type has proven to be ideal for use in schools, in terms of practicality, time efficiency, cost, and facilitation. Regardless of the educational environment in which the activities are implemented (class, science lab, auditorium), escape boxes and the existence of a fixed working space encourage the team members to gather all together, brainstorm, discuss, and solve the game puzzles.
3. *including physical objects as puzzle components.* By increasing students' engagement and visualising theoretical concepts, these items arouse students' curiosity, allow them to learn by doing, and facilitate them to practise their 4C skills while using them.

4. *adopting a combination of linear and multi-linear puzzle pathways.* While linear puzzle pathways ensure that all team members are exposed to the same amount of knowledge, multi-linear puzzle pathways force the less engaged or introvert students to take action and participate more actively.
5. *designing self-guided puzzles that align well with the EERs' learning objectives, the curriculum, and the game narrative.* Appropriately designed puzzles that immerse learners in the storytelling experience and do not require (substantial) scaffolding to be solved have a good chance of increasing their cognitive, behavioural and affective engagement, and foster learning.
6. *adding layers to puzzles.* Instead of using puzzle components that provide all needed information in a direct and clear way, encrypting part of that information or cleverly 'hiding' it by making it seem trivial, can offer more depth, increase the puzzle's difficulty, and boost players' creativity and analytical thinking.
7. *increasing the number and the complexity of puzzle components.* Apart from making the puzzle more difficult, the increased number of available items enables, and sometimes requires, the engagement of more team members, creating some sort of social interdependence. Effective collaboration and communication become a prerequisite for solving the puzzle. Selecting, sorting, or matching items also foster the learners' observational, analytical and reasoning skills.
8. *limiting the provision of scaffolding and hints to a minimum.* Most players' observation, creativity and critical thinking usually sharpen when they reach an impasse, as long as they remain at a state of flow. Reducing the amount of available information makes the puzzle's solution less straightforward and increases the time players spend on it, thereby extending the practice of 4C skills.
9. *incorporating meta-puzzles.* Meta-puzzles are usually placed on the convergence point of complex or multi-linear puzzle paths. They offer an excellent opportunity for synthesising findings from previous puzzles, but also for re-examining information more carefully and discover something new that might have been overlooked. Their advanced complexity requires higher order thinking skills and effective collaboration between the members of a team.
10. *challenging learners' pre-existing knowledge.* Puzzles that deal with part of the syllabus that students find challenging are very useful in revealing students' weaknesses and misconceptions. At the same time, they can easily ignite discussion among students, fostering the practice of their analytical thinking and reasoning skills.

5.5 Conclusions

Practitioners and entrepreneurs alike have been claiming for a long time now that students utilise the 4C skills while engaging in EER activities. Educational researchers have also investigated this matter, considering that EERs have been widely used

with the learning intention of practising and developing several types of skills, including the 4Cs. Our longitudinal study provided strong evidence that verified these claims. The application of a meticulous observational analysis on rich qualitative data that derived from numerous video/audio recordings and a control-based research design, as opposed to the methods applied in previous studies, offered a more reliable, detailed and accurate documentation of these skills. Apart from the practice of the 4C skills, indications of their development were also found. Data collected from several different methods (skill tests, questionnaires, interviews) corroborated these indications. Furthermore, we identified some connections between specific puzzle types and the practice of certain 4C skills. Based on the study's findings and informed by the existing literature, we created a list of practical design guidelines that the early adopters could utilise to develop EERs that can foster these skills combined with content learning.

Teaching biology in the 21st century is much more than a sheer transfer of content knowledge. Using this knowledge effectively requires certain skills that are equally important. Among other things, biology students are expected to: (a) critically analyse biological data in order to understand them and propose creative solutions for real-life problems; (b) care about socio-scientific issues in biology, be able to express their opinion and take action; (c) use scientific reasoning to communicate their knowledge; (d) collaborate with others towards common goals. Bearing in mind the importance that educational reforms place on developing students' 4C skills, it is necessary to investigate further the design and long-term effectiveness of appealing educational activities that can deliver this outcome, like the EERs.

References

- Ananiadou, K., & Claro, M. (2009). *21st century skills and competences for new millennium learners in OECD countries* (OECD education working papers, 41, p. 33). <https://doi.org/10.1787/218525261154>
- Bapna, A., Sharma, N., Kaushic, A., & Kumar, A. (2017). *Handbook on measuring 21st century skills*. <https://doi.org/10.13140/RG.2.2.10020.99203>
- Bialik, M., Fadel, C., Trilling, B., Nilsson, P., & Groff, J. (2015). Skills for the 21st century: What should students learn?, 117(July 2017). Retrieved from http://curriculumredesign.org/wp-content/uploads/CCR-Skills_FINAL_June2015.pdf
- Clarke, S., Peel, D. J., Arnab, S., Morini, L., Keegan, H., & Wood, O. (2017). escapED: A framework for creating educational escape rooms and interactive games for higher/further education. *International Journal of Serious Games*, 4(3), 73–86.
- Fadel, C., Bialik, M., & Trilling, B. (2015). *Four-dimensional education: The competencies learners need to succeed*. Center for Curriculum Redesign.
- Fotaris, P., & Mastoras, T. (2019, October). Escape rooms for learning: A systematic review. In *Proceedings of the European conference on games-based learning*, pp. 235–243. <https://doi.org/10.34190/GBL.19.179>
- Fotaris, P., & Mastoras, T. (2022). Room2Educ8: A framework for creating educational escape rooms based on design thinking principles. *Educational Sciences*, 12(11), 768. <https://doi.org/10.3390/educsci12110768>
- Hu, W., & Adey, P. (2002). A scientific creativity test for secondary school students. *International Journal of Science Education*, 24, 389–403. <https://doi.org/10.1080/09500690110098912>

- Lai, E. R., & Viering, M. (2012, April). Assessing 21st century skills: Integrating research findings. In *Annual meeting of the national council on measurement in education*, p. 66.
- Lathwesen, C., & Belova, N. (2021). Escape rooms in stem teaching and learning – Prospective field or declining trend? A literature review. *Educational Sciences*, 11(6). <https://doi.org/10.3390/educsci11060308>
- Marsden, E., & Torgerson, C. J. (2012). Single group, pre- and post-test research designs: Some methodological concerns. *Oxford Review of Education*, 38(5), 583–616. <https://doi.org/10.1080/03054985.2012.731208>
- Nicholson, S. (2018). Creating engaging escape rooms for the classroom. *Childhood Education*, 94(1), 44–49. <https://doi.org/10.1080/00094056.2018.1420363>
- Nicholson, S., & Cable, L. (2021). *Unlocking the potential of puzzle-based learning: Designing escape rooms and games for the classroom* (1st ed.). SAGE.
- P21. (2009). *P21 framework definitions*. Retrieved from http://www.p21.org/documents/P21_Framework_Definitions.pdf
- Qian, M., & Clark, K. R. (2016). Game-based learning and 21st century skills: A review of recent research. *Computers in Human Behavior*, 63, 50–58. <https://doi.org/10.1016/j.chb.2016.05.023>
- Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. E. O'Malley (Ed.), *Computer-supported collaborative learning* (pp. 69–197). Springer. https://doi.org/10.1007/978-3-642-85098-1_5
- Sternberg, R. J. (1986). *Critical thinking: Its nature, measurement and improvement* (p. 37). National Institute of Education.
- Sternberg, R., & Lubart, T. (1998). The concept of creativity: Prospects and paradigms. In R. Sternberg (Ed.), *Handbook of creativity* (pp. 3–15). Cambridge University Press. <https://doi.org/10.1017/CBO9780511807916.003>
- Taraldsen, L. H., Haara, F. O., Lysne, M. S., Jensen, P. R., & Jenssen, E. S. (2022). A review on use of escape rooms in education—touching the void. *Education Inquiry*, 13(2), 169–184. <https://doi.org/10.1080/20004508.2020.1860284>
- Veldkamp, A., Daemen, J., Teekens, S., Koelewijn, S., Knippels, M. C. P. J., & van Joolingen, W. R. (2020a). Escape boxes: Bringing escape room experience into the classroom. *British Journal of Educational Technology*, 51(4), 1220–1239. <https://doi.org/10.1111/bjet.12935>
- Veldkamp, A., van de Grint, L., Knippels, M. C. P. J., & van Joolingen, W. R. (2020b). Escape education: A systematic review on escape rooms in education. *Educational Research Review*, 31(January), 100364. <https://doi.org/10.1016/j.edurev.2020.100364>
- Veldkamp, A., Rebecca Niese, J., Heuvelmans, M., Knippels, M. C. P. J., & van Joolingen, W. R. (2022). You escaped! How did you learn during gameplay? *British Journal of Educational Technology*, 53, 1430–1458. <https://doi.org/10.1111/bjet.13194>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

