



OPEN ACCESS

EDITED BY
Gavin Duffy,
Technological University Dublin, Ireland

REVIEWED BY
Cucuk Wawan Budiyanto,
Sebelas Maret University, Indonesia
Mariana Rocha,
Technological University Dublin, Ireland

*CORRESPONDENCE
Zoe Falomir
✉ zoe.falomir@umu.se

RECEIVED 28 September 2023
ACCEPTED 03 January 2024
PUBLISHED 23 February 2024

CITATION
Garcia-Segarra P, Santamarta V and Falomir Z
(2024) Educating on spatial skills using a
paper-folding-and-punched-hole
videogame: gameplay data analysis.
Front. Educ. 9:1303932.
doi: 10.3389/educ.2024.1303932

COPYRIGHT
© 2024 Garcia-Segarra, Santamarta and
Falomir. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Educating on spatial skills using a paper-folding-and-punched-hole videogame: gameplay data analysis

Pablo Garcia-Segarra¹, Vicent Santamarta¹ and Zoe Falomir^{2*}

¹Languages and Computer Systems Department, Universitat Jaume I, Castelló de la Plana, Valencian Community, Spain, ²Computing Science Department, Umeå University, Umeå, Sweden

Introduction: Paper folding and punched hole tests are used to measure spatial abilities in humans. These abilities are relevant since they are associated with success in STEM (Science, Technology, Engineering, and Mathematics). This study addresses the challenge of teaching spatial reasoning skills using an educational videogame, the Paper Folding Reasoning Game.

Methods: The Paper Folding Reasoning Game is an interactive game which presents activities intended to help users train and understand how to fold a paper to get a specific shape (Part I) and the consequence of punching a hole on a folded paper (Part II). This educational videogame can automatically generate paper-folding-and-punched-hole questions with varying degrees of difficulty depending on the number of folds and holes made, thus producing additional levels for training due to its embedded reasoning mechanisms (Part III).

Results: This manuscript presents the results of analyzing the gameplay data gathered by the Paper Folding Reasoning Game in its three parts. For Parts I and II, the data provided by 225 anonymous unique players are analyzed. For Part III (Mastermode), the data obtained from 894 gameplays by 311 anonymous unique players are analyzed. In our analysis, we found out a significant difference in performance regarding the players who trained (i.e., played Parts I and II) before playing the Mastermode (Part III) vs. the group of players who did not train. We also found a significant difference in players' performance who used the visual help (i.e., re-watch the animated sequence of paper folds) vs. the group of players who did not use it, confirming the effectiveness of the Paper Folding Reasoning Game to train paper-folding-and-punched-hole reasoning skills. Statistically significant gender differences in performance were also found.

KEYWORDS

paper folding, education, gameplay analysis, videogames, qualitative descriptors, skill training, spatial skills, spatial cognition

1 Introduction

Spatial cognition studies have shown that there is a strong link between success in Science, Technology, Engineering, and Mathematics (STEM) disciplines and spatial abilities (Newcombe, 2010).

As Hegarty (2010) defined, spatial thinking “involves thinking about the shapes and arrangements of objects in space and about spatial processes, such as the deformation of objects, and the movement of objects and other entities through space. It can also involve thinking with spatial representations of non-spatial entities, for example, when we use an organizational chart to think about the structure of a company”.

Research studies by [Verdine et al. \(2014\)](#) showed that students with low spatial abilities under-perform in STEM tasks and then they avoid STEM disciplines when selecting college. Children in families with low socioeconomic status are also disadvantaged in spatial ability development ([Wai et al., 2009](#)). [Lippa et al. \(2010\)](#) reported a gender gap in spatial ability (i.e., mental rotation and line-angle judgment) in a worldwide study and concluded that gender equality and economic development were significantly associated, across nations, with larger sex differences. The European Parliament resolution of 10 June 2021 on promoting gender equality in STEM education and careers [2019/2164(INI)] ([European Parliament, 2021](#)) highlighted that *in view of the rising demand for STEM practitioners and the importance of STEM-related careers for the future of the European economy, increasing the share of women in the STEM sector is critical to building a more sustainable and inclusive economy and society through scientific, digital, and technological innovation*.

According to studies in the literature, spatial skills can be trained. The studies carried out by [Sorby \(2009\)](#) showed that spatial skills can be developed through practice: students who attended an engineering graphics gateway course at university to improve their ability to visualize in three dimensions improved also their success and retention significantly, particularly female students. Moreover, spatial thinking can be taught using visual and kinetic interactions offered by new digital technologies ([Highfield and Mulligan, 2007](#)), and research has demonstrated that video game training enhances cognitive control ([Spence and Feng, 2010](#)), especially when aging ([Anguera et al., 2013](#)).

During the COVID-19 pandemic, there was an urgent need to fulfill urgent distance learning requirements, so massive open online courses, open course ware, and other educational resources, such as virtual avatar-based platforms and educational videogames became popular. The European Parliament promoted a Digital Education Action Plan for 2021–2027 stating that *learning can happen in a fully online or blended mode, at a time, place, and pace suited to the needs of the individual learner*.¹ According to [Qian and Clark \(2016\)](#), “*Game-based learning (GBL) describes an environment where game content and game play enhance knowledge and skill acquisition where game activities involve problem solving spaces and challenges that provide players/learners with a sense of achievement*”.

Thus, this study faces the challenge of teaching spatial reasoning skills using a videogame. The following research question is addressed here: can the *Paper Folding Reasoning videogame* help players to improve their paper-folding-and-hole-punching reasoning skills? This study answers this question by analyzing players’ gameplay data.

In the literature, videogames have been used to analyze skills, motivation, and/or way of playing of the players. [Kirschner and Williams \(2014\)](#) analyzed gameplay reviews to assess videogame engagement of players. [Nicolaidou et al. \(2021\)](#) showed that the more hours the young adults in their study played digital puzzle games, the higher their spatial reasoning skills. [Kim et al. \(2023\)](#)

concluded that game-based assessments that incorporate learning analytics can be used as an alternative to pencil-and-paper tests to measure cognitive skills, such as spatial reasoning. [Clark et al. \(2023\)](#) analyzed three games—*Transformation Quest*, *NCTM’s Flip-N-Slide*, and *Mangahigh’s transtar*—which are designed to practice geometric transformations (spatial rotations) to find out if they provided an academically meaningful play to the users. [Baki et al. \(2011\)](#) compared the effects of using a dynamic geometry software (DGS) vs. physical manipulatives on the spatial visualization skills gathered by pre-service mathematics teachers, obtaining that the DGS-based group performed better than the physical manipulative-based group in the views section of the Purdue Spatial Visualization Test. [Yavuz et al. \(2023\)](#) analyzed the *Sea Hero Quest* gameplay data, and their findings showed that video gaming is associated with spatial navigation performance, not influenced by gender, but associated with weekly hours of video gaming. In another analysis of the gameplay data gathered by *Sea Hero Quest* videogame, [Coutrot et al. \(2023\)](#) found out that education level was positively associated with wayfinding ability, and that, this difference was stronger in older participants and increased with task difficulty.

These related studies provide evidence of the interest of using videogames and smart applications for social good. However, as far as we are concerned, there are no studies in the literature that report the use of a videogame to train paper folding reasoning skills of players.

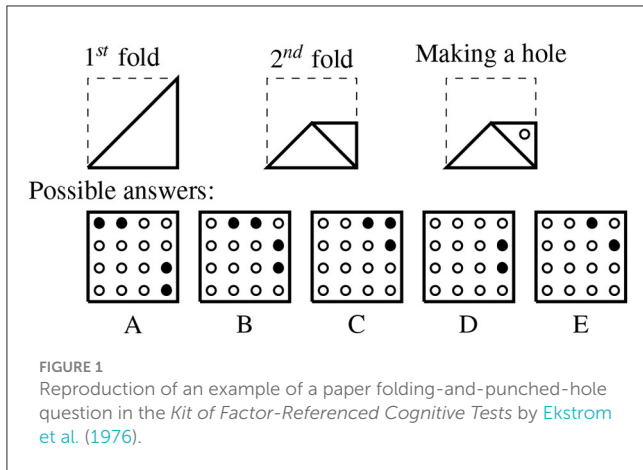
The rest of the study is organized as follows. Section 2 presents the paper-folding-and-punched-hole test. Section 3 describes the computer game developed to train paper folding: *Paper Folding Reasoning Videogame*. Section 4 presents our analysis of the gameplay data. Section 5 links our findings with research studies on spatial cognition. Section 6 presents conclusions and future studies.

2 The paper folding and punched holes test (PFT)

Spatial visualization was defined by [Ekstrom et al. \(1976\)](#) as the *ability to manipulate or transform the image of spatial patterns into other arrangements*. The paper folding and punched holes test (PFT for short) measures of spatial visualization skills of participants, and it is included in the *Kit of Factor-Referenced Cognitive Tests* by [Ekstrom et al. \(1976\)](#).

An example of a question in the punched holes test is presented in [Figure 1](#). The instructions provided are as follows: *A flat square is folded one or more times. The broken lines indicate the original position of the paper. The solid lines indicate the position of the folded paper. The paper is never turned or twisted. The folded paper always remains within the edges of the original square. There may be from 1 to 3 folds in each item. After the last fold, a hole is punched in the paper. Your task is to mentally unfold the paper and determine the position of the holes in the original square. Choose the pattern of black circles that indicates the position of the holes on the original square.*

¹ European Union’s Digital Education Action Plan 2021–2027.



3 The Paper Folding Reasoning Videogame

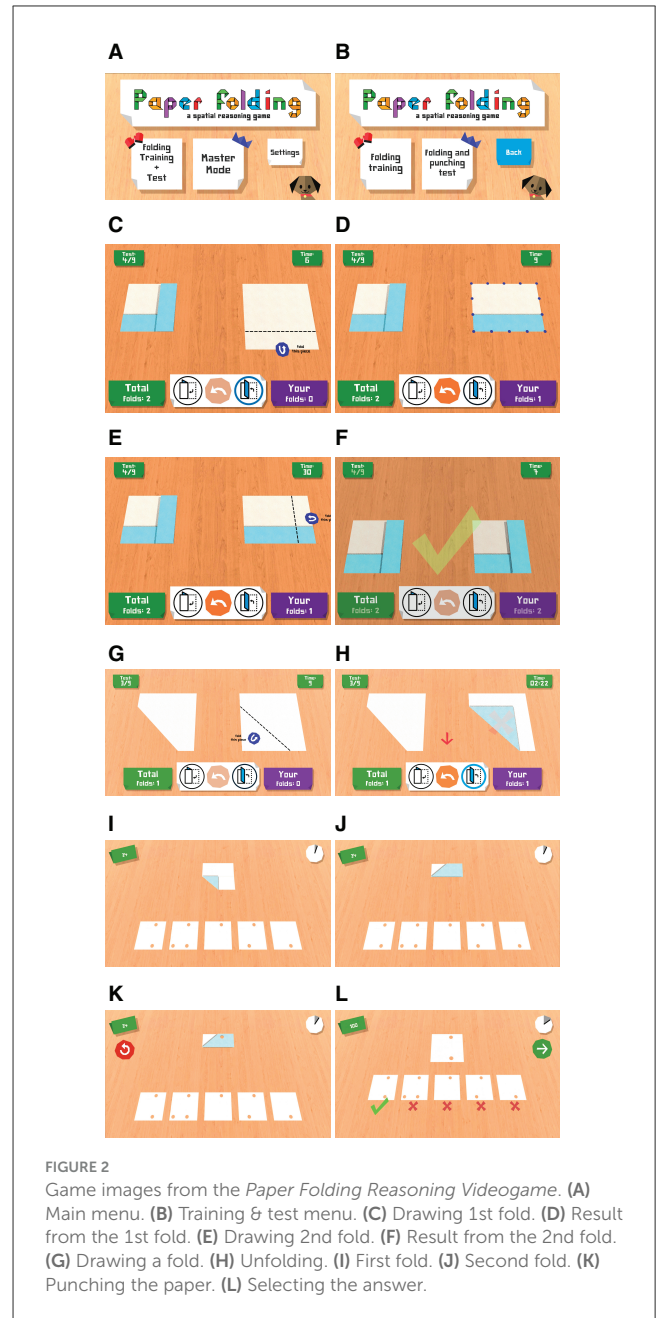
The *Paper Folding Reasoning Videogame* is an interactive game which presents activities intended to help users train and understand how to fold a paper to get a specific shape and the consequence of punching a folded paper. It can automatically generate paper-folding-and-punched-hole questions with varying degrees of difficulty depending on the number of folds and holes made, thus producing additional levels for training. For that, it uses the Qualitative model for Paper Folding (QPF) and its reasoning logics (Falomir et al., 2021) (see an overview in Section 3.2) which allows it: to infer the right answer to each paper-folding-and-punched-hole question; to provide feedback to the players when they are wrong; and to create other plausible answers automatically so that random question-answers are shown to the players in the Mastermode.

The *Paper Folding Reasoning Videogame* offers two playing modalities: *Folding training & Test* and *Mastermode* (Figure 2A). These options focus on new players or on experienced players, respectively. The *Folding training & Test* part (Figure 2B) consists of exercises of *Folding training* and a PFT.

The *Paper Folding Reasoning Videogame* runs on any tablet/mobile with an operative system Android 4.1 “Jelly Bean” (API 16) or upper version. This game is available to download from GooglePlay,² and a demo video is also available in our website.³

3.1 The gameplay at Paper Folding Reasoning Videogame

In the *Folding training & Test* mode (Figure 2B), a brief tutorial shows: (i) how to draw a folding line, (ii) how to select the part of the paper to fold, (iii) how to undo a fold, and (iv) how to determine the direction of a fold: upwards or downwards.



After the tutorial, the *Paper Folding Reasoning Videogame* asks players to imitate folded papers using the folding options and directions previously learnt. Players are presented a short set of folded papers. The folded paper to imitate is located on the left, and the player has to fold the paper located on the right of the display. The objective paper is yellow, if the front is visible, or blue, if the back of the paper is showed (see Figures 2C–F). There is also an *Undo button* in the bottom-middle of the screen, which can be used to revert the paper to the previous state.

In each case, the number of folds done in the objective paper are indicated: they go from 1 to 3 folds. When a player has reached the maximum number of folds without success, an arrow will point out the *Undo button*, as shown in Figures 2G, H. In this part, as users are training, they cannot fail, this means that they can press

² <https://play.google.com/store/apps/details?id=com.spatialreasoninggames.PaperFolding>
³ Spatial reasoning games: <https://spatialreasoninggames.weebly.com/>.

the *Undo button* as many times as they need until the folded paper is correct. There is a *Skip button* available that appears after the player has failed 3 times so that the players do not get totally demotivated.

The main aim in this part is that players get familiar with the dynamics of paper folding in this game so that they got the chance to practice consecutive foldings and realize how the paper changes depending on the fold.

After players finish the *Folding training* part (Figure 2B), the *Paper Folding Reasoning Videogame* takes players to the next level, the *Folding and punching test*, whose objective is to test their performance in a set of 15 paper-folding-and-punched-hole questions. Figures 2I–L shows a paper-folding-and-punched-hole question. First of all a paper is being folded one or multiple times, then a hole is punched through it, and finally five possible answers appear. Players' aim is to select the correct answer before the time ends (represented by the clock on the right-up corner).

For these tests, players have two buttons available: the reload button (red arrow on the left in Figure 2K), which repeats the folding animation when pressed; and the help button, which discards two possible answers. When players provide an incorrect answer, Doggo provides a feedback by explaining why the given answer was wrong (see Falomir et al., 2021 for a description of all types of feedback provided by the game). This part ends after players answer the 15 selected questions.

After the *Folding training + Test* part, players can continue training using the *Mastermode* (Figure 2A). In this mode, all the questions and answers are randomly and automatically generated using the QPF model by applying reasoning logics (see an overview in next section). The logics behind *Paper Folding Reasoning Videogame* ensure a different experience each time is the *Mastermode* played.

3.2 Overview of the Qualitative Model for Paper Folding

This section provides an overview of the Qualitative Model for Paper Folding (QPF) which describes qualitatively: hole locations and folding actions. Inference tables bring paper-folding actions and areas into correspondence, and they are used for reasoning (Falomir et al., 2021). The Qualitative Descriptor for Paper Folding (QPF) describes qualitative areas on a paper, the possible locations of holes in these areas, and the actions that can be done on that paper using the reference systems (RS) HoleLocation_{RS} and FoldActions_{RS}, which are illustrated in Figures 3A, B, respectively.

Sequences of paper-folding actions are related to hole punching locations, and the QPF can be used to find out these locations. For example, the QPF can solve paper-folding-and-punched-hole questions presented in the spatial reasoning tests described in the *Kit of Factor-Referenced Cognitive Tests* by Ekstrom et al. (1976). Let us exemplify this by considering the sequence of foldings as shown in Figure 1 at the Introduction section:

So, when punching a hole at the location *rdm* (or right-down-middle) after the folding sequence {*diagonal-upward*, *in-middle*}, which other areas were punched? This problem is solved by propagating the hole locations using the inference tables in the inverse folding sequence. As the inference table related to the action

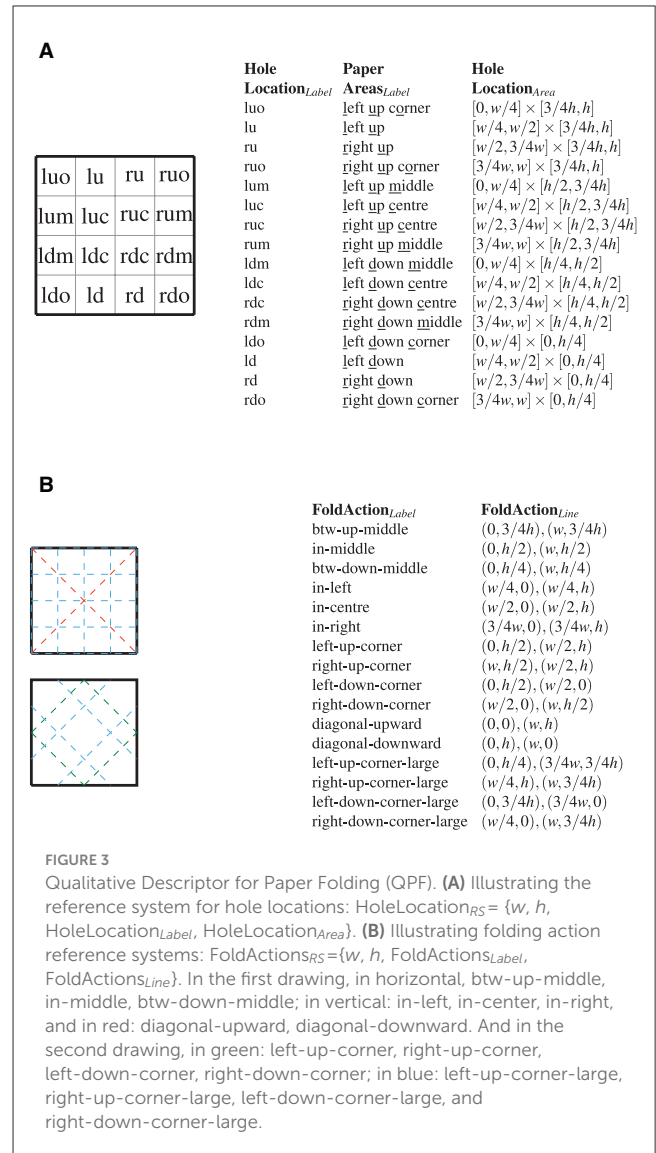


FIGURE 3 Qualitative Descriptor for Paper Folding (QPF). (A) Illustrating the reference system for hole locations: HoleLocation_{RS} = {w, h, HoleLocation_{Label}, HoleLocation_{Area}}. (B) Illustrating folding action reference systems: FoldActions_{RS} = {w, h, FoldActions_{Label}, FoldActions_{Line}}. In the first drawing, in horizontal: btw-up-middle, in-middle, btw-down-middle; in vertical: in-left, in-center, in-right, and in red: diagonal-upward, diagonal-downward. And in the second drawing, in green: left-up-corner, right-up-corner, left-down-corner, right-down-corner; in blue: left-up-corner-large, right-up-corner-large, left-down-corner-large, and right-down-corner-large.

in-middle shows, *rdm* is connected to *rum*, and then, it is inferred that two holes are punched located at *rdm* and *rum*, respectively.

Folding	luo	lu	ru	ruo	lum	luc	ruc	rum
in middle	ldo	ld	rd	rdo	ldm	ldc	rdc	rdm

Then, as the inference table related to the action *diagonal-upward* shows, *rum* is connected to *ru* and *rdm* is connected to *lu* by unfolding. Thus, finally four holes are obtained in total and located at: *rdm*, *rum*, *ru*, and *lu*. So, the correct answer is B.

Folding	ldm	lum	luo	luc	lu	ru
diagonal upward	ld	rd	rdo	rdc	rdm	rum

3.3 Storing data in Paper Folding Reasoning Videogame

The *Paper Folding Reasoning Videogame* records gameplaying data for further analysis. Most of the information come implicitly

TABLE 1 Player's demographics from Parts I, II, and III.

Age (years)	Part I & II	Part III	Studies level	Part I & II	Part III	Area of studies	Part I & II	Part III	Nationality	Part I & II	Nationality	Part III
	% of Total	% of Total		% of Total	% of Total		% of Total	% of Total		% of Total		
0–10	8.4	3.2	Elementary School	14.2	6.1	Engineering & Technology	20.9	7.1	Spain	16.9	India	2.9
11–20	21.3	9.0	High School	17.3	7.1	Social Sciences	5.3	1.3	India	4.4	Spain	2.6
21–30	12.4	5.5	Bachelor	14.7	5.1	Natural Sciences	2.7	1.6	United Kingdom	2.2	Germany	1.0
31–40	10.2	3.5	Master	9.3	3.9	Not provided	71.1	90.0	North America	2.2	United Kingdom	1.0
41–50	2.7	1.0	PhD	4.9	1.0				Thailand	1.8	North America	0.6
51–60	5.3	1.0	Not provided	39.6	76.8				Germany	1.8	Bangladesh	0.6
Not provided	39.6	76.8							Mexico	1.3	Pakistan	0.6
									Netherlands	1.3	Other (<0.6%)	3.9
									Italy	1.3		
									Canada	1.3		
									Other (<1%)	5.1		
									Not provided	59.6	Not provided	86.8

Other countries (Part I and II) (<1%): Australia, Brazil, Ecuador, France, New Zealand, Philippines, Romania, Scotland, South Africa, South Korea, and Sweden.

Other countries (Part III) (<0.6%): Brazil, Chile, Costa Rica, El Salvador, France, Ireland, Italy, Mexico, Philippines, Romania, and Scotland.

from the performance of the player while playing the game: question identifier, answer given by the player, a true/false value showing if the help button was pressed, how many seconds took the player to answer, the total score and some extra technical information such as a unique identifier for the machine where the game is being played, date, and time. Players can also rate the *Paper Folding Reasoning Videogame*.

To follow the rules of the European General Data Protection Regulation,⁴ approved and adopted by the EU Parliament in April 2016, the game shows initially a pop-up message that notifies the user that data of usage will be retrieved to carry out this research and improve the game. Players can disable the sending of information by unchecking the corresponding toggle in the settings menu. If agreed by the players, players' performance is stored in a Firebase Realtime Database,⁵ that is, a cloud-hosted NoSQL database from Google, available in cross-platform apps (iOS and Android). These data are stored in JSON format, being easily converted to other formats such as ".csv" for conducting research studies. The next section analyses the data obtained in the gameplay by *Paper Folding Reasoning Videogame*.

After playing the *Paper Folding Reasoning Videogame*, a survey appears where players are asked about demographic data that are useful for dividing them into different study groups: their age, gender, nationality, level of studies, and field of studies. These data can be left unanswered, and then, the player is considered an anon-definednymous person.

4 Analysis of data obtained by *Paper Folding Reasoning Videogame*

This section analyses the data obtained by players in the three parts of the game:

- Part I: users' paper folding training;
- Part II: users' performance in the folding-and-punched-hole test; and
- Part III: users' performance in the Mastermode.

For the Parts I and II, the videogame gathered data from $N = 225$ players. All 225 participants provided gameplay data but only 136 provided demographic data (age, gender, nationality, level of studies, and field of studies). These 136 participants consisted of 74 female (mean age 22.9 years) and 62 male (mean age 26.1 years). The complete players' demographics can be seen in Table 1 under Part I & II.

For Part III (Mastermode), the videogame gathered 8,940 answers (894 players \times 10 questions) corresponding to 894 gameplays by 311 unique players. All 311 participants provided gameplay data but only 72 provided demographic data (age, gender, nationality, level of studies, and field of studies). These 72 participants consisted of 29 female (mean age 20.3 years) and 43 male (mean age 23.9 years). The complete players' demographics can be seen in Table 1 under Part III.

4 General Data Protection Regulation: <https://eugdpr.org/>.

5 Firebase Realtime Database: <https://firebase.google.com/docs/database>.

4.1 Part I: analysis of data obtained in the folding training

The folding training part consists of nine exercises which are similar to that in Figures 2C–F, where a folded paper is shown on the left, and the player must perform the folding actions on the paper located on the right of the display to obtain the same shape.

These nine training exercises increase gradually the number of foldings required: initially 3 simple exercises consisting of only 1-fold are presented, followed by three exercises consisting of 2-folds, and finally, the last three exercises involving 3-folds. Figure 4 shows the exercises used in *Paper Folding Reasoning Videogame*.

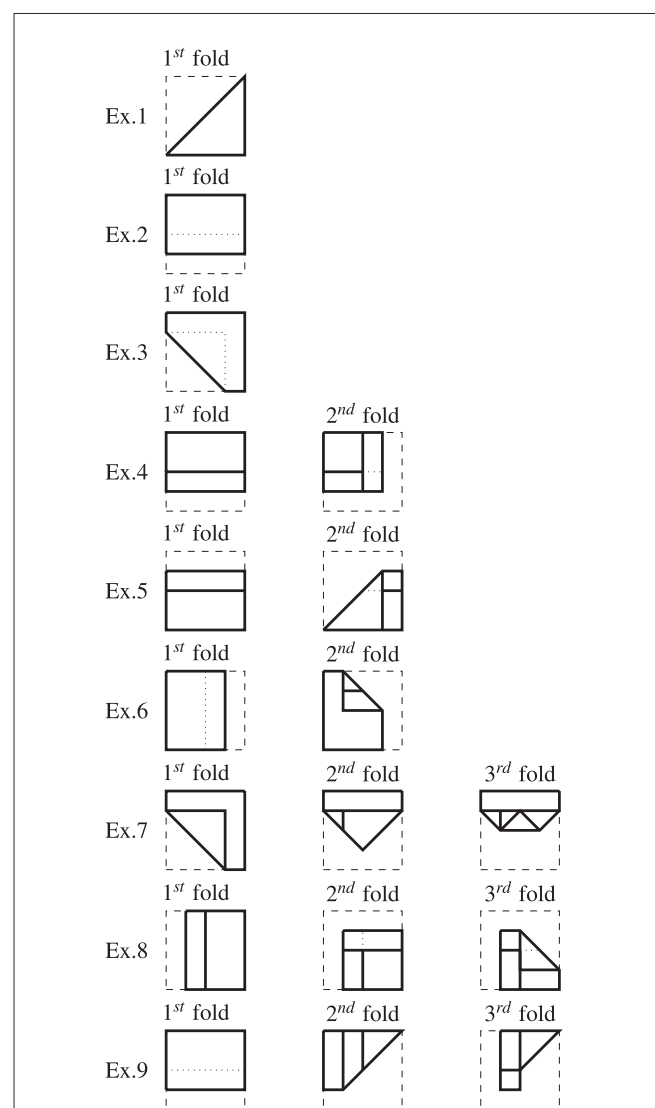
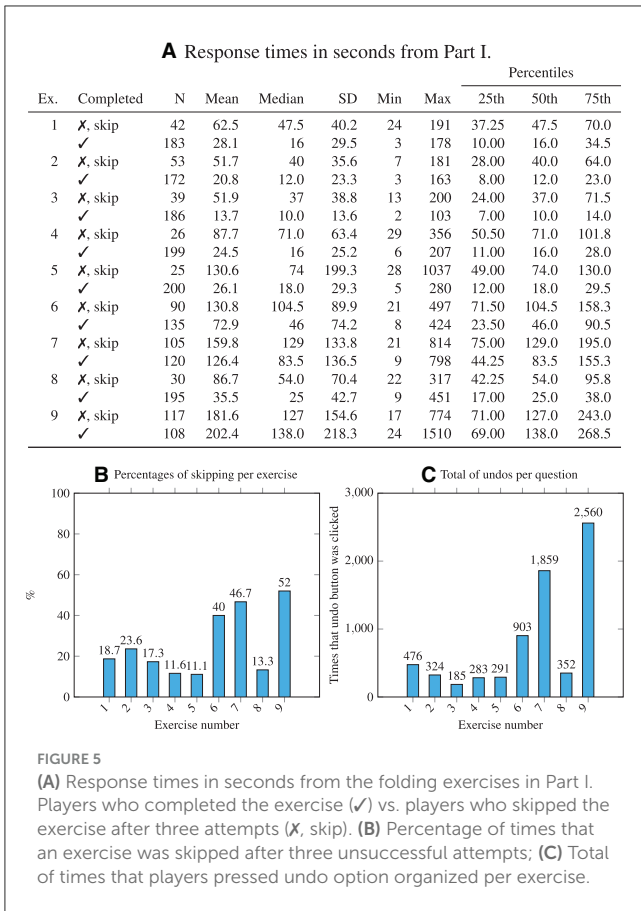


FIGURE 4

Folding training exercises in the *Paper Folding Reasoning Videogame*: Exercises 1–3 consisted of only 1-fold, Exercises 4–6 consisted of 2-folds, and Exercises 7–9 consisted of 3-folds. The direction of the folds can be forwards or backwards. The dotted lines indicate the border of the paper that are invisible to the view (when folded backwards). The dashed lines indicate the original area of the paper.



In Part I, players have no limit of time for completing the folding exercises. After three folding attempts (i.e., three times pressed the undo button), the skip button appears allowing the players to pass to the next exercise, if they click it. However, they can still try to solve the current exercise as many times as they like. The *Paper Folding Reasoning Videogame* stores their gameplay data in Google Firebase after getting their players' consent.

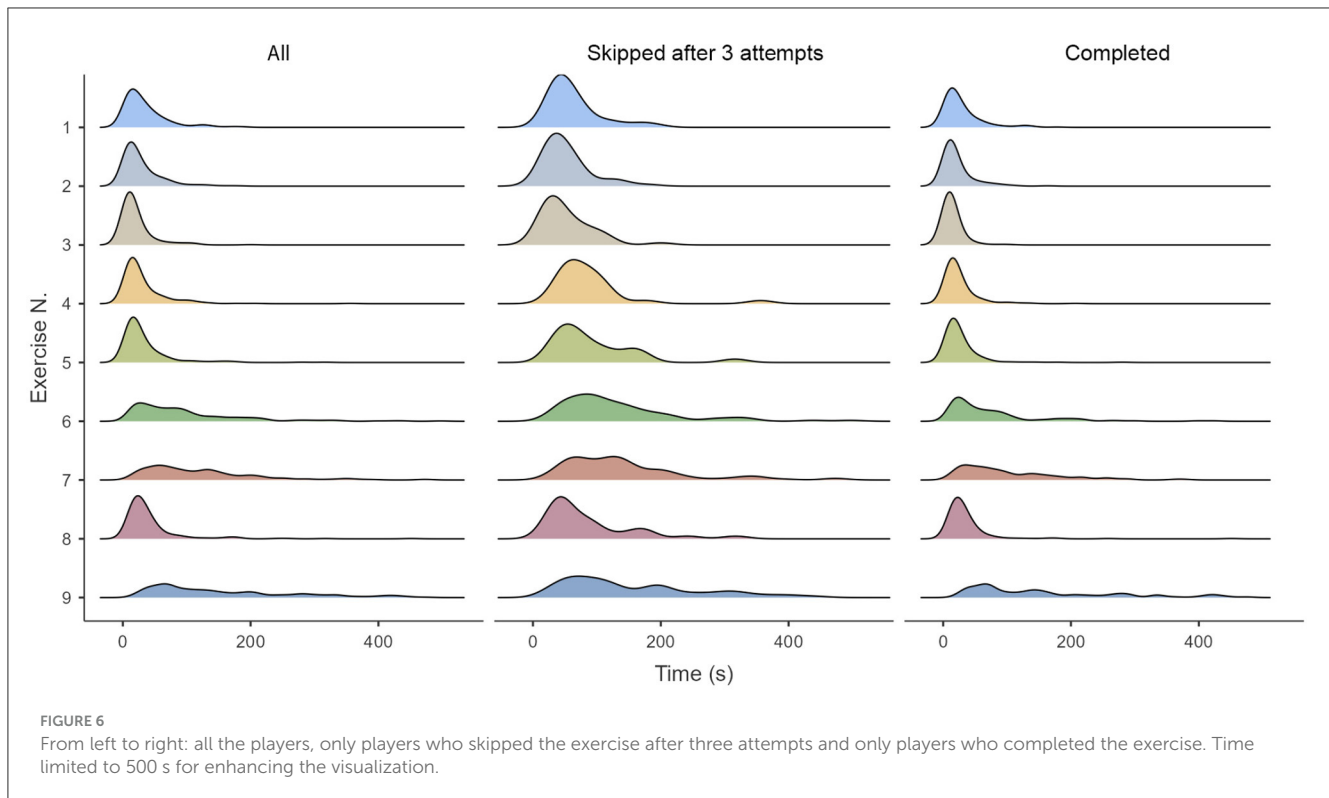
This section presents the analysis of the data gathered from 225 anonymous players who downloaded the *Paper Folding Reasoning Videogame* from GooglePlay or AppleStore. These 225 players responded to the nine folding training exercises, so 2,025 answers were stored. Among these answers, 527 responses correspond to situations where the exercise was skipped after three attempts, that is, players were not able to solve it, and that corresponds to 26% of the total answers. Players who solved the exercises provided 1,498 answers, which is 74%. Next, players' response times are analyzed.

Table 5 shows the analysis of time gathered for each of the exercises. For example, Exercise 5 was solved by 200 players in an average time of 26.1 s, where the maximum time needed was 280 s and the minimum time needed was only 5 s. In contrast, 25 players failed to solve the exercise after three attempts, and their average playing time was 130.6 s (the minimum time they needed was 28 s actually higher than the average time needed by the players who solved the same exercise, 26.1 s). The rest of the rows in the table are read similarly.

Analyzing the data in Figure 5A, the players who correctly solved the Ex. 1–8 were more (in quantity) and also faster (more efficient). Regarding Ex. 9, there were more players skipping the exercise than actually solving it (117 vs. 108) and those players who solved the exercise took longer than those who skipped it. This indicates that Ex. 9 was the hardest to solve by most of the players.

Figure 5B shows the % of skipped exercises after three unsuccessful attempts. Notably, Ex. 6, 7, and 9 were skipped by 40, 46.7, and 52% of times, respectively. Figure 5C shows that the players pressed more the undo button in the Ex. 6, 7, and 9 than in the other exercises. The more times the *undo* button was pressed, the more actions the players needed to solve the exercise. Figure 6 plots the players' response time (in seconds) to solve the folding Ex. 1–9. Comparing the times spent to answer the exercises to the quantity of skips per exercise (Figure 5B) and to the quantity of undone actions per exercise (Figure 5C), we can conclude that Ex. 1–5 and 8 were relatively easy for the players, and that Ex. 6, 7, and 9 were the most challenging. Let us analyze these folding exercises:

- Ex. 6 requires to mix folding directions: folding backwards the right side of the paper, so when folding the right corner forwards, the paper on the top was the front face and not the back.
- Ex. 7 was the first exercise involving three folds, and it got the second highest percentage of skip: 46.7%. The difficult part of this exercise was that several players tried to fold the lower right and left corners to the center of the paper to make the first two folds, so the result is a symmetrical arrow, with the tip on the bottom center of the paper, which is not the correct answer. Figure 6 shows that the group of players, that finally did skip the exercise, actually spent more time trying to solve the exercise than those who did not skip it. It was an exercise which got similar solving and skipping time average (126.4 vs. 159.8 s, respectively). That is, after ~2 min, some players got the answer and others got discouraged and skipped the exercise. The fact that those who skipped the exercise took more time in their attempts indicate that at least players were engaged with the game, taking it seriously.
- Ex. 9 was the hardest to solve with a minimum solving time of 24 s and an average solving time of ~3 min (202.4 s in Figure 5A). The difficulty is similar to Ex. 6, players must fold backwards a piece of paper, so when it is folded forwards again, the paper facing up is the front and not the back. Ex. 9 was skipped 52% of the times (Figure 5B) and generated 2,560 undone actions (Figure 5C). As Figure 5 shows, usually players who completed the exercises were quicker than those who skipped them, who tried folds for longer time. However, Ex. 9 is the exception, where players who completed the exercise took longer than those who skipped it (202.4 vs. 181.6 s, respectively). This indicates that Ex. 9 was elaborated to solve, maybe because players thought about many possible intermediate steps which they had to discard. The most attempted first and second folds in Ex. 9 was stored in the game data and they are shown in Supplementary Figure 1, the correct answer is highlighted in bold. For the first fold, 12.6% of the players successfully selected the fold *between-down-middle_backwards*. The most tried folds are *in-left forwards* and *diagonal-upwards_forwards*



(with 17.7 and 15.3%, respectively). The fold *in-left_forwards* is needed to complete the exercise, but its correct location in the folding sequence is the last fold. The fold *diagonal-upwards_forwards* is also one of the folds required for solving Ex. 9, but its correct location in the sequence is the second folding, not the first. In conclusion, the three most used folds are part of the solution in the Ex. 9, which indicates that the players were carrying attempts in the right solving direction most of the times.

Table 2A shows a further analysis regarding the more difficult exercises, that is, Ex. 6, 7, and 9. We have observed that 38.2% of the players completed all the exercises, whereas 29.3% of the players skipped them. This table also shows all the performance options regarding these exercises. Among all the options, 52.8% of the players completed at least two of these hard exercises, and that 70.7% solved at least one.

As the final analysis regarding Part I, Table 2B shows that 34.7% of the players completed all the exercises, and that 42.7% of the players skipped less than the half of them. This shows that the adequacy of the tasks was acceptable for 77.4% of the players (that is ~ 174 people) who completed at least five of nine of the exercises.

4.2 Part II: users' performance in the folding-and-punched-hole test

The Part II in the *Paper Folding Reasoning Videogame* presents 15 folding-and-punched-hole questions extracted from a sample of the *Perceptual Ability Test* (PAT) from the admission testing

program at the *North American Dental Association* (ADA).⁶ Specifically, the test number is 56, and the questions selected are PART/4, question numbers 46–60 (see Appendix 1 for more detail).

This section analyzes the performance corresponding to 225 players who answered 15 questions in this Part II (225 tests \times 15 questions/test = 3,375 answers). From these, 1,768 answers were correct (52.39%) and 1,607 answers were wrong (47.61%).

Table 3 shows players' performance per question. A total of 225 responses were gathered per question from where we extracted the % of success, % of times the help button was pressed, and % of times that the reload/repeat button was clicked. The questions that got a higher % of success were PAT4_48, PAT4_50, and PAT4_56 which got 87.6, 69.3, and 69.8, respectively, whereas the exercise that got a lower % of success was PAT4_60 with only 13.3% of success.

Table 3 shows the questions and the players' selected options. For example, when solving the easiest question, PAT4_48, 87.6% of the players selected option C, and the rest of the options got $<5\%$ of responses. In contrast, when solving the PAT4_60, only 13.3% of the players selected the correct answer, option D, the preferred option was E, although it was not the right one. Notably, that in the PAT4_60, players who solved correctly used 2.7% of times the help button and 33.8% of times the reload button. Moreover, we might affirm that players preferred the reload button (visual repetition of the spatial transformation) to the help button (discarding answers) when looking for assistance. Notably, the last three questions are those where the highest % of responses do not correspond to the correct answer. Moreover, the question PAT4_58 was the most confusing for the players since 43.1% chose option B, whereas the correct answer was option E, which was only chosen 28.4% of times.

⁶ Dental Admission Testing Program example: <http://www.ada.org/>.

TABLE 2 (A) Players' performance in the three more challenging exercises of the training Part I: exercises 6, 7, and 9. (B) Amount of skipping in the training Part I grouping the players according to their performance.

(A) Players' performance in ex 6, 7, and 9				
Exercise			Players	% of Total
6	7	9		
✓	✓	✓	86	38.2
✓	✓	✗	16	7.1
✓	✗	✓	12	5.3
✗	✓	✓	4	1.8
✗	✗	✓	6	2.7
✗	✓	✗	14	6.2
✓	✗	✗	21	9.3
✗	✗	✗	66	29.3
Total			225	100.0

(B) Players' skipping percentage		
Group	Players	% of Total
Completed all	78	34.7
Skipped < 50%	96	42.7
Skipped > 50%	47	20.9
Skipped all	4	1.8
Total	225	100.0

TABLE 3 (Left) Answers to the questions corresponding to Part II, folding-and-punched-hole test, % of success (the highest in bold), % of times when help was used, and % of times that the repeat option was used by players per question. (Right) Percentages of selected options per exercise (correct answers in bold).

Exercise	Solved			Help used			Repeat used			Answers					
	✓	✗	✗	□	☑	□	□	□	□	A	B	C	D	E	Timeout
PAT4_46	68.4	31.6	0.9	99.1	5.3	94.7	68.4	2.2	15.6	10.2	3.6	0			
PAT4_47	54.7	45.3	0.0	100.0	8.4	91.6	20.0	15.1	54.7	4.4	5.8	0			
PAT4_48	87.6	12.4	0.4	99.6	4.0	96.0	2.2	2.7	87.6	4.4	3.1	0			
PAT4_49	46.2	53.8	1.8	98.2	16.0	84.0	28.9	46.2	14.2	2.7	8.0	0			
PAT4_50	69.3	30.7	0.0	100.0	2.2	97.8	0.9	21.3	7.1	69.3	1.3	0			
PAT4_51	49.8	50.2	0.4	99.6	8.4	91.6	49.8	20.9	7.1	19.1	3.1	0			
PAT4_52	65.8	34.2	2.2	97.8	11.1	88.9	8.4	65.8	20.9	3.6	1.3	0			
PAT4_53	42.7	57.3	0.4	99.6	22.7	77.3	42.7	20.0	25.8	10.7	0.4	0.4			
PAT4_54	52.9	47.1	1.3	98.7	16.0	84.0	7.6	52.9	29.8	3.1	6.2	0.4			
PAT4_55	50.7	49.3	1.3	98.7	8.4	91.6	50.7	20.4	17.3	7.6	4.0	0			
PAT4_56	69.8	30.2	0.4	99.6	12.4	87.6	4.0	69.8	2.2	4.0	20.0	0			
PAT4_57	60.0	40.0	0.4	99.6	6.7	93.3	60.0	23.6	3.6	7.6	4.9	0.4			
PAT4_58	28.4	71.6	2.2	97.8	26.7	73.3	7.1	43.1	15.6	5.3	28.4	0.4			
PAT4_59	26.2	73.8	1.3	98.7	19.6	80.4	18.2	28.0	16.0	10.2	26.2	1.3			
PAT4_60	13.3	86.7	2.7	97.3	33.8	66.2	17.8	18.7	19.6	13.3	28.9	1.8			

Appendix 1 shows a geometrical description of each option.

Table 4B summarizes players' success in Part II: 39.6% of players answered correctly more than 70% of the questions; 24% of players answered correctly between 50%–70% of the questions; 36% of players were successful less than 50% of times; and only 1 player answered correctly all questions.

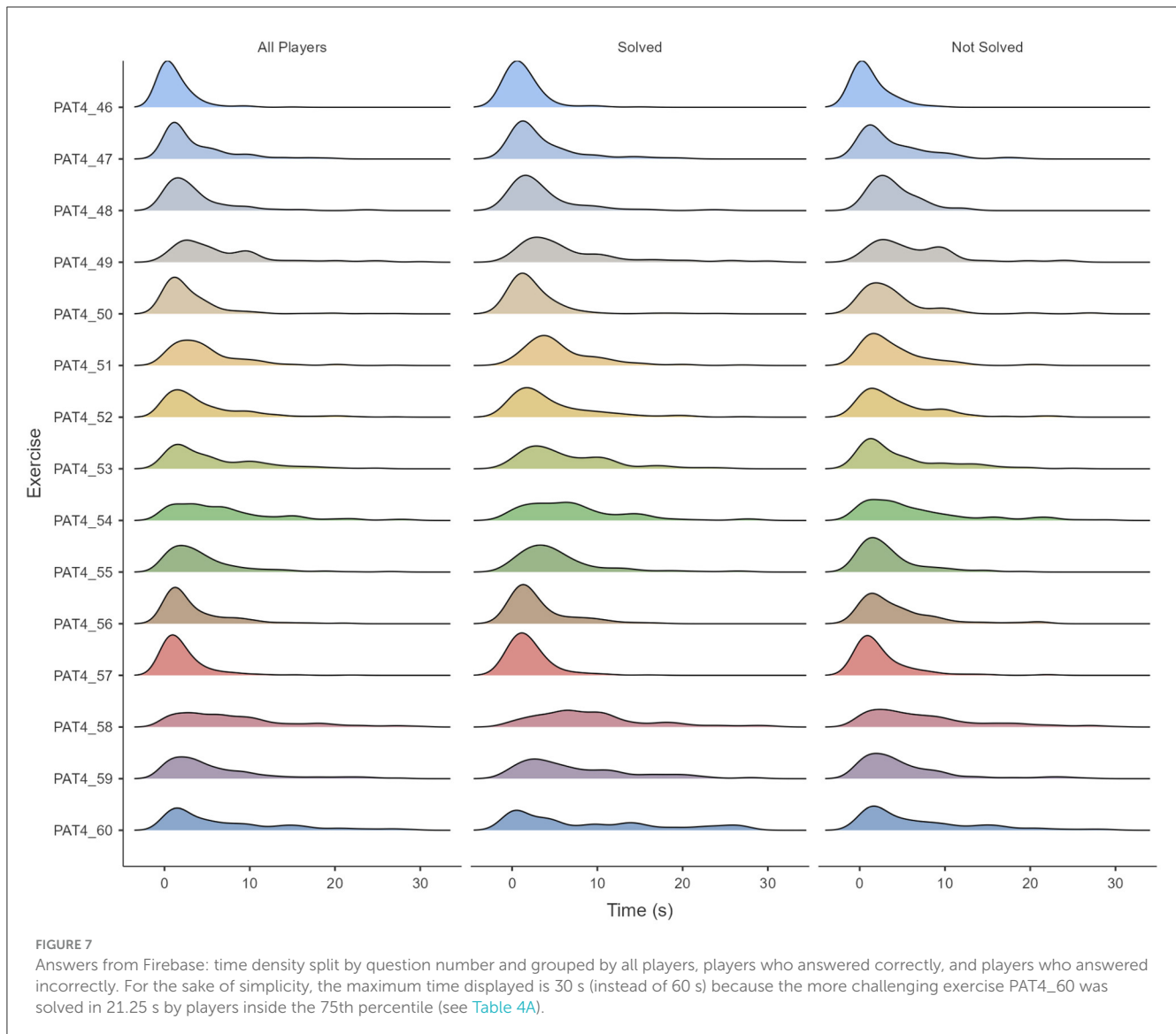
TABLE 4 (A) Folding-and-punched-hole test in Part II: time statistics per exercise. (B) Solved % in the training Part II grouping the players according to their performance.

(A) Time to answer (in seconds) in Part II									
Ex.	N	Mean	Median	SD	Min	Max	Percentiles		
							25th	50th	75th
PAT4_46	✓ 154	1.55	0	3.93	0	42	0	0	2
	✗ 71	1.21	0	1.96	0	9	0	0	2
PAT4_47	✓ 123	4.16	1	6.60	0	51	1	1	5
	✗ 102	3.79	2	4.01	0	19	1	2	5.75
PAT4_48	✓ 197	3.46	2	4.18	0	24	1	2	4
	✗ 28	3.93	3	2.65	1	12	2	3	5.25
PAT4_49	✓ 104	7.11	5	7.08	0	31	2	5	10
	✗ 121	6.40	5	5.32	0	25	3	5	9
PAT4_50	✓ 156	2.76	1	4.34	0	35	1	1	3
	✗ 69	3.96	3	4.53	0	27	1	3	5
PAT4_51	✓ 112	5.72	4	4.48	0	27	3	4	7.25
	✗ 113	3.76	2	3.80	0	21	1	2	5
PAT4_52	✓ 148	5.53	3	7.90	0	55	1	3	6.25
	✗ 77	4.42	3	4.73	0	23	1	3	6
PAT4_53	✓ 96	7.18	5	7.01	0	43	2	5	10
	✗ 129	5.96	3	8.54	0	60	1	3	8
PAT4_54	✓ 119	8.53	7	8.70	0	51	3	7	10
	✗ 106	7.15	4	8.68	0	60	1.25	4	9
PAT4_55	✓ 114	5.48	4	5.16	0	27	2	4	7
	✗ 111	3.34	2	3.68	0	19	1	2	4
PAT4_56	✓ 157	3.06	2	3.49	0	17	1	2	4
	✗ 68	4.37	3	4.73	0	21	1	3	6
PAT4_57	✓ 135	2.02	1	2.50	0	16	0	1	3
	✗ 90	3.22	1	7.04	0	60	0.25	1	3
PAT4_58	✓ 64	10.20	8.50	8.12	0	37	5	8.50	12.25
	✗ 161	9.60	6	11.14	0	60	3	6	12
PAT4_59	✓ 59	8.95	6	9.82	0	53	2.50	6	12
	✗ 166	6.22	4	9.23	0	60	1	4	7
PAT4_60	✓ 30	12.87	9	13.55	0	43	1.25	9	21.25
	✗ 195	10.25	5	13.01	0	60	2	5	14

(B) Groups by % of solved questions		
Group	Players	% of Total
Solved all	1	0.4
Solved > 70%	89	39.6
Solved between 50%–70%	54	24.0
Solved < 50%	81	36.0
Failed all	0	0.0

Table 4A provides the players' response time per exercise and Figure 7 illustrates that. Three groups are displayed: (a) all players, (b) players' time who solved the exercise, and (c) players' time who did not solve the exercise. *Paper Folding Reasoning Videogame* provides 1-min maximum for answering each exercise. By observing the density time distribution, it seems that the given time is acceptable, taking into account that users are given 3 min to answer 10 questions in the original paper test. By Analyzing Figure 7 and Table 4A in more detail, we observe that:

- The questions involving only one fold got quicker answers: PAT4_46 (with an average response time of 1.55 s for right answers and of 1.21 s for wrong answers) and PAT4_50 (with



an average response time of 2.76 s for right answers and of 3.96 s for wrong answers). They got also quite high % of success: 68.4 and 69.3%, respectively.

- The questions that got higher % of success, that is, PAT4_48, PAT4_50, and PAT4_56, also got quicker correct answers, that is, players who solved the exercise correctly took shorter time to answer than those players who did not solve it.
- The question PAT4_57 was solved correctly 60% of times and was the second quicker responding exercise: the correctly solved average time was 2.02 s while the not solved average time was 3.22 s, indicating that the solution was quite obvious for some players.
- The three last questions (PAT4_58, PAT4_59, PAT4_60) were the most challenging for the players, who only succeeded 28.4, 26.2, and 13.3 of times, respectively, and took the longest response average time (10.2, 9, and 12 s, respectively).
- In the rest of the questions, players who solved the question correctly took longer to answer than those players who did not solve it. Moreover, we can affirm that more thoughtful players were gathering more right answers.

Notably, Appendix 1 shows the geometry of all the folding-and-punched-hole questions used in Part II.

Figure 8 shows the success rate in each exercise (numbered from 1 to 15 in x axis) and it compares it with the total rate of success (in % shown in y axis). The blue line shows the rate success tendency, and it can be observed that it is negative, which indicates that the questions at the end were found more challenging by the users. The red dashed line indicates the average of success rate per exercise, that is, 52.4%.

The Part II in the *Paper Folding Reasoning Videogame* provides the players with two extra options: (i) a reload button which repeats the folding actions in the paper and the punching location, and (ii) a help button which discards two possibilities in the set of answers. Analyzing the 3,375 answers obtained, we found that the reload/repeat button was pressed 454 times, that is 13.45% of the times. Table 6 shows that 6.61% of times that the repeat option was used, the players got the correct answer, whereas 6.84 % they got the wrong answer.

Table 5A shows a comparison in the amount of correct answers in Part II obtained by those players who used the repeat/reload

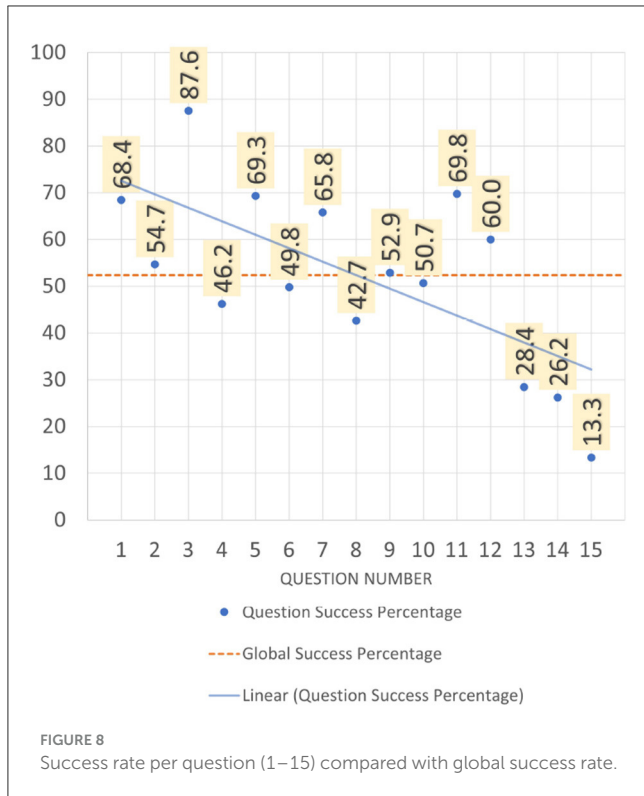


FIGURE 8 Success rate per question (1–15) compared with global success rate.

button and by those players who did not use it. Those players who used repeat obtained a higher number of correct solved questions (56.5%), in comparison to the group who did not use it (47%). Table 5B shows a Mann–Whitney *U*-test comparing players who answered correctly and used the reload/repeat option (Group 1) vs. players who answered correctly and did not use the reload/repeat option (Group 2). The alternative hypothesis is that the Group 1 got a greater amount of correct answers in Part II than the Group 2. The *p*-value in the test is very low (0.002), indicating that there is a significant difference in performance when using the repeat/reload function, thus the repeat/reload function is helpful for the players.

In contrast, Analyzing the 3,375 answers, we found that the help button was only used 1.07% of the times (see Table 6), so obviously it was not the preferred option by the players. To find out a relation between using the help option and the success selecting the correct answer, a Mann–Whitney *U*-test was used (Table 7) to compare users who answered correctly and did not use the help option (Group 1) with users who answered correctly and used the help option (Group 2), and *p* = 0.755 was obtained, indicating that there is no significant dependence between groups. Thus, the help option was not influencing the players’ success.

4.3 Comparing players’ performance in Part I and Part II

This section analyzes players’ performance in Part II (paper-folding-and-hole-punching questions), taking into account players’ performance in Part I (folding exercises). The main aim is to

TABLE 5 (A) Analyzing players, performance: average of total correct answers during the test if they used repeat/reload option at least once vs. who did not use that option at all. (B) Mann–Whitney *U*-test for total correct answers in Part II: players that used repeat (Group 1) vs. players who did not use it (Group 2).

(A) Average correct answers when using repeat or not					
Repeat	By players	Got ✓ questions	Median	SD	SE
☑	126	8.48/15 = 56.53%	9.00	3.48	0.310
☐	99	7.06/15 = 47.07%	8.00	3.43	0.344
Total	225				
(B) Total ✓ answers: used repeat vs. no repeat					
Statistic	<i>p</i>	Mean diff.	Rank Biserial Correlation		
4817	0.002	1.42	0.228		

Hypothesis_a: $\mu_{Repeat Used} > \mu_{No Used Repeat}$

TABLE 6 (Left) Amount of times the repeat/reload button was used in Part II related to question success (3). (Right) Percentages of times that help button was pressed per exercise solved correctly (✓) or wrongly (X).

Repeat	Success			Help	Success		
	✓	X	Totals		✓	X	Totals
☑	6.61	6.84	13.45	☑	0.65	0.41	1.07
☐	45.78	40.77	86.55	☐	51.73	47.20	98.93
Totals %	52.39	47.61	100.00	Totals %	52.39	47.61	100.00

TABLE 7 Mann–Whitney *U*-test for the total correct answers in Part II: players who did not use help vs. players who use it.

Statistic	<i>p</i>	Mean diff.	Rank Biserial Correlation
2618	0.755	0.51	0.0788

Hypothesis_a: $\mu_{No Used Help} < \mu_{Used Help}$

assess whether the results obtained in Part I influence the results in Part II, that is: *do the folding exercises train players for the paper-folding-and-hole-punching questions?*

To answer this research question, the performance by players who answered at least 7/9 exercises (at least one of the three challenging exercises), that is, who completed more than 78% of the exercises in Part I (Group A) is compared with the performance by players who completed <78% of the exercises in Part I (Group B).

Table 8A shows that there are 141 players in Group A, obtaining a mean of 56.6% of correct answers during Part II, whereas Group B includes 84 players who obtained a mean of 45.3% of correct answers during Part II. Therefore, Group A obtained a better performance in Part II than Group B, with a difference in the average of 11.3%. The results by both groups are analyzed in Table 8B using a Mann–Whitney *U*-test. The result of the *t*-test gets a very low *p*-value (<0.001), and this indicates that there is a difference between the two groups of players, being the first

TABLE 8 (A) Percentages of correct answers in Part II Analyzing the performance of players in Part I. (B) Mann–Whitney *U*-test. Comparing the amount of correct exercises in Part II from the group of players that completed at least 78% of the questions in Part I against the players that completed <78%.

(A) Success in Part II					
Players who	N	Mean	Median	SD	SE
Completed ≥ 78% in Part I	141	56.6	60.0	22.4	1.89
Completed < 78% in Part I	84	45.3	46.7	23.7	2.59
Hypothesis _a : $\mu_{tG} > \mu_{mG}$					
(B) T-test Part II answers: completed ≥ 78% from Part I vs. < 78%					
Statistic	<i>p</i>	Mean diff.	Rank Biserial Correlation		
4233	<0.001	11.3	0.285		

Hypothesis_a: $\mu_{Completed \geq 78\%} > \mu_{Completed < 78\%}$.

one (the more trained group) better than the second. Thus, the players that completed seven out of nine questions during Part I (which includes at least one of the three difficult questions) have a better performance during Part II, which indicates that Part I trains players’ spatial skills regarding paper folding reasoning.

From the (*N* = 141) players who got better performance during Part II, 92 provided demographic data. These 92 participants consisted of 44 male (age range: 0–60 years, mean age: 25.5 years) and 48 female (age range: 0–60 years, mean age: 25.1 years).

The 44 male players who got better performance during Part I got a success rate of 67.6% in Part II, while the 48 female players got 53.1% success.

4.4 Part III Mastermode: gameplay data analysis

The *Paper Folding Reasoning Videogame* generates 10 random questions in the Mastermode (Part III). Each question consists of 1–4-fold actions, increasing the difficulty of the question per fold. The educational videogame gathered 894 anonymous gameplays, that is 8,940 questions in total. Table 9A shows that the number of folds is equally distributed (1/4 or 25%) for both the first gameplay and all the gameplays.

Knowing that the number of folding actions is equally distributed in all sessions, the success of the users can be analyzed in two ways: (i) regarding players’ first play, where the gameplays of 311 different players are selected (that is 3,110 answers); and (ii) regarding recurrent gameplays, where the data gathered contained 894 plays, that is 8,940 answers. Table 9B shows that the players have a success rate of 48.1% in the first gameplay, and their success rate increases to 63.1% when playing recurrently. Since the gameplay contains four amount of folding levels, the performance of each one can be analyzed independently. Questions with 1-fold action are easy to solve by new and recurrent players (83.9 and 90.1%). When encountering 3 or 4-folds, the performance during the first gameplay decreases significantly down to 31.3 and 34.6%, meaning that two out of three questions are not

TABLE 9 Analyzing questions containing folds: (A) in the Mastermode; (B) performance at the 1st gameplay vs. all gameplays.

(A) Folds	1st gameplay		All gameplays	
	Questions	%	Questions	%
1	806	25.9	2,302	25.7
2	784	25.2	2,225	24.9
3	784	25.2	2,187	24.5
4	736	23.7	2,226	24.9
(B) Folds	1st gameplay		All gameplays	
1	83.9		90.1	
2	40.7		59.7	
3	31.3		50.8	
4	34.6		50.4	
All	48.1		63.1	

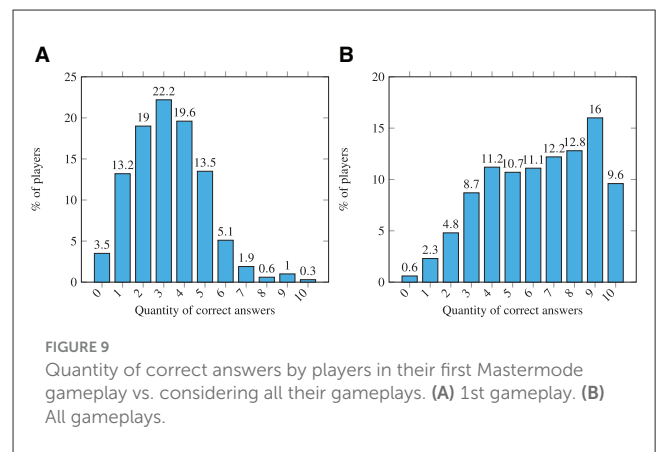


FIGURE 9 Quantity of correct answers by players in their first Mastermode gameplay vs. considering all their gameplays. (A) 1st gameplay. (B) All gameplays.

solved, whereas players who have more experience are able to solve up to 50% of those questions. This confirms that the higher the amount of folding actions, the more challenging the question.

The data in Table 9B indicate that players have higher success rate when playing multiple times, especially when facing questions with a higher amount of folds (i.e., 3 and 4). To analyze this further, Figure 9 shows the distribution of the players based on the number of questions answered correctly in the Mastermode. The majority of the players got under five correct answers on their first gameplay; notably, 22.2% of the players answered correctly only three questions. In contrast, when considering all gameplays, the data are more distributed toward the higher end, with its maximum value at nine correct questions answered by 16% of the players. Notably, the majority of the players on concurrent plays are correctly solving more than five questions, indicating that the users who played recurrently the Mastermode ended up having a better performance.

Players can also use two types of helping options in the Mastermode: (i) Help option which discards answers and (ii) Repeat option which shows again the paper folding sequence. To see their effect on players’ performance, we analyze these data split by the amount of folds per question. Table 10 shows the usage

TABLE 10 Analyzing data from Mastermode: (Left) % of help usage and % of repeat usage, split by the amount of folds per question. (Right) % of success when using help and % of success when using repeat.

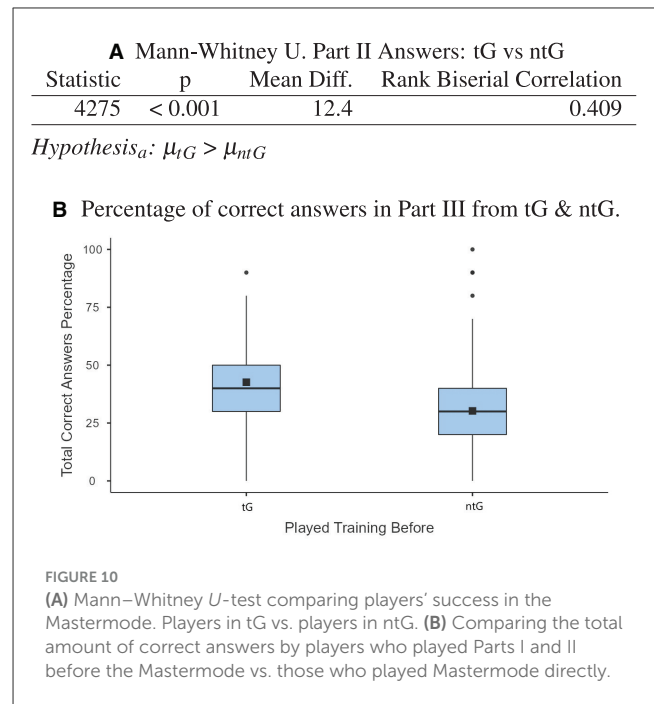
Folds	% usage of				% of success when using			
	☑ Help	☐	☑ Repeat	☐	☑ Help	☐	☑ Repeat	☐
1	0.3	99.7	3.5	96.5	50.0	90.3	88.9	90.2
2	0.7	99.3	10.7	89.3	50.0	59.8	55.6	60.2
3	3.2	96.8	19.8	80.2	64.3	50.4	61.0	48.3
4	8.7	91.3	28.8	71.2	69.1	48.7	62.3	45.6
All	3.2	96.8	15.6	84.4	66.3	33.7	62.3	37.7

of help/repeat options: (i) the Help option was not used 3.2% of time, in general, increasing to 8.7% in the case of the most difficult questions (involving four folds); (ii) the Repeat option was used in 15.6% of the questions and up to 28.8% in the most difficult ones. Moreover, the repeat option was also preferred by players in the Mastermode.

A further analysis of the help/repeat options is shown in Table 10 right, which shows the % of success when using help and the % of success when using repeat, categorizing the questions by amount of folds. In the case of the Help option, when encountering a difficult question (i.e., including 3 or 4 folds), the success percentage increases to 64.3 and 69.1%, respectively, whereas the players' success who did not use the help option is lower: 50.4 and 48.7%, respectively. This shows that the help button was useful in difficult questions. In the case of the Repeat option, players also benefit from using it in the difficult questions (including 3 or 4 folds), obtaining a success percentage of 61 and 62.3%, respectively. This success percentage is higher when comparing with the players who did not use the Repeat option, which is 48.3 and 45.6%, respectively. In the case of easier questions (including 1 or 2 folds), the usage of both Help and Repeat options yielded a lower success percentage than that not using it. Notably, as these questions are easier, most of the players did not use the help/repeat options (i.e., only 1% used help as shown in Table 10 left) and those who use it where successful 50% of times using the option help, but 88% of times using the option repeat in questions with only 1-fold. This indicates that the repeat option involves visually observing the folding sequence again, and it was more useful for players than the help option which consisted of discarding two answers.

4.5 Comparing players' performance in Part III with respect to previous Parts I and II

In this section, the players' performance in the Part III (Mastermode) is analyzed by selecting and comparing players' performance who have previously played Part I and Part II and the players who played directly the Mastermode. The objective is to answer whether the training in Part I and Part II influences the results in the Mastermode. For that, we selected each player's first play in the Mastermode, ending with 311 different plays. These plays were separated into two groups, depending on if: (i) they were performed by players who previously played the Part I and II (trained Group, tG), and (ii) they were performed by players who directly played the Mastermode (no trained Group, ntG). From this



two groups of players, only 18.3% (57/311) are in the tG against 81.7% (254/311) that are in ntG.

Comparing the percentage of correct answers, tG obtains a mean of 42.6% and a median of 40% against ntG that obtains a mean of 30.2% and a median of 30%. The difference between the mean of the two groups is that tG performed 12.4% better than ntG. This data are presented in Figure 10B. When comparing the interquartile range (IQR), both groups have the same range, 20%, meaning that the data are spread similarly in the middle half. In tG, the Q1 is higher than the Q1 of ntG (30 vs. 20%), which also indicates that the tG increased their success rate in 12.4%, on average.

To obtain the statistical significance of this difference of performance between the groups, a Mann-Whitney *U*-test was carried out (Figure 10A). The alternative hypothesis is that: *tG's performance in Mastermode > ntG's performance in mastermode*. The *p*-value of the *U*-test is <0.001, accepting the alternative hypothesis, indicating that there is a statistically significant difference in favor of the performance of tG vs. ntG.

From the ($N = 57$) players who completed the training (Part I & Part II) before playing the Part III-Mastermode, 20 provided demographic data. These 20 participants consisted of 12 male (age range: 10–60 years, mean age: 23 years) and eight female (age range: 10–60 years, mean age: 31.8 years). The 12 male players who completed the training got a success rate of 48.3% during Part III, while the eight female players got a success rate of 45%.

4.6 Summary of results

This section summarizes the analysis of the gameplay data gathered by the *Paper Folding Reasoning Videogame*, and it answers our research question positively, that is, the data gathered indicate

that this videogame does train players' skills on paper folding and hole punched reasoning tasks.

4.6.1 Regarding the folding exercises (Part I with 225 players)

This part consisted of nine paper folding exercises with no time limit. Comparing the times spent to answer the exercises (Figure 6) with the quantity of skips per exercise (Figure 5A) and the quantity of undone actions per exercise (Figure 5B), we can categorize exercises as *easy* for players (Ex. 1–5 and 8) vs. *challenging* exercises (Ex. 6, 7, and 9). The *easy* exercises were quicker to solve and less skipped. A further analysis of the responses to the *challenging* exercises (Table 2B) shows that 38.2% of the players completed them all, 52.8% of the players completed at least 2, and 70.7% solved at least one. Moreover, we can conclude that most of the players took seriously this training part.

4.6.2 Regarding the paper-folding-and-punched-hole questions (Part II with 225 players)

This part consisted of 15 folding-and-punched-hole questions extracted from a sample of the *Perceptual Ability Test (PAT)* by the admission program at the *North American Dental Association (ADA)*.⁷ Specifically, the questions selected are ADA I, numbers 46–60 (see Appendix 1 for a visual description).

From our analysis, we conclude that: (i) the questions that got higher % of success also got faster response times; and that (ii) the last three questions (PAT4_58, PAT4_59, and PAT4_60) were the most challenging for the players, who only succeeded 28.4, 26.2, and 13.3% of times, respectively, and took the longest response average time (10.2, 9, and 12 s, respectively).

Comparing players' performance in Part I and Part II, we found out that the players who completed seven out of nine questions during Part I (which includes at least one of the three challenging exercises) had a higher % of success in Part II, which indicates that Part I do train players' spatial skills regarding paper folding reasoning.

We found a significant difference in players' performance who used the visual help (i.e., re-watch the animated sequence of paper folds).

4.6.3 Regarding Mastermode (Part III with 311 unique players)

This part consisted of sets of 10 questions generated randomly. Analyzing the performance of the 311 unique players in the Mastermode, we found that they succeeded 48.1% of times in their first gameplay. Some users played repeatedly 894 gameplays (8,940 answers) were gathered. By Analyzing all the recurring gameplays, we obtained that the percentage of success increases to 63.1%, which indicates that users who played recurrently ended up performing better.

By Analyzing the questions depending on the folds in all the gameplays, we find out that the questions which got the highest % of

success are those including only 1-fold, followed by those including 2-folds, and the questions including 3–4-folds got ~50% of success after recurrent gameplays. Moreover, even by answering sets of different random questions, players improve their performance.

Comparing players' performance in Part III, we found out a significant difference in performance regarding the players who trained before playing the Mastermode (18.3%) with respect to those who did not train (81.7%).

4.6.4 Regarding gender differences in performance

The 65.2% of players who improved from Part I to Part II ($N = 141$) provided their gender, while 34.8% did not provide demographic data. Analyzing the data from those who provided their gender, we observed a difference in performance when training in Part I and testing in Part II: the 44 male players (mean age 25.5) who got better performance during Part I got a success rate of 67.6% in Part II, whereas the 48 female players (mean age: 25.1) with better performance in Part I got 53.1% success rate in Part II, that is a 14.5 % difference. This difference between genders is statistically significant according to the obtained p -value of < 0.01 .

With respect to the training in Part I and Part II before playing the Part III, our sample size is smaller ($N = 57$), where 37 gameplays had not provided demographic data, and only 20 gameplays had gender-associated. The difference between male and female success rate is only 3.3%. In this case, the difference is not statistically significant according to the obtained p -value of < 0.677 . Notably, the number of gameplays available in this case is very low.

4.6.5 Regarding the videogame scoring by players

In Part I and II, the videogame was rated (in a scale 1–5, being 5 the highest) as: Intuitive (4.15), Educative (4.22), Fun (4.16), and Help Feedback (3.91). In Part III, players rated the game (1–5, being 5 the highest) as: Intuitive (4.03), Educative (4.07), Fun (4.11), and Help Feedback (3.85).

5 Discussion

This section relates our findings, which are data-driven by participants' gameplay, to our results of experiments in spatial cognition literature.

5.1 Relating categorization of folds to success rate by players

The paper folding and punched holes test by Ekstrom et al. (1976) is usually used to measure spatial visualization (VZ). It has been proved by Kane et al. (2005) that it is psychometrically reliable and that is related strongly to executive functioning and working memory capacity (WMC). Jaeger (2015) investigated the role of VZ and WMC dividing the items in the PFT tasks in basic and atypical. Atypical folds were categorized as those that occlude or hide one

⁷ Dental Admission Testing Program example: <http://www.ada.org/>.

or some of the previous made folds. The findings by Jaeger (2015) indicated that participants' performance on all item types was more strongly predicted by participants' VZ skills, and at the same time, there was some evidence to suggest that the basic fold items relied more on the participants' WMC. Participants with both high VZ and high WMC had advantages solving the basic fold items but not solving the atypical fold items.

More recently, Burte et al. (2019) indicated that "the interaction of diagonal folds with other fold types also produces <<non-perceptual matches>>. These are problems where the punch location in the probe (e.g., in the top-right corner) does not match the punch location in the correct answer (e.g., in the bottom-right corner)". They also hypothesized that not having a match between probe and response punch locations also requires more effort due to memory demands.

In the *Paper Folding Reasoning Game*, the questions which got the lowest % of success are PAT4_60 (13.3%), PAT4_59 (26.2%), and PAT4_58 (28.4%), which can be categorized as *atypical* following the definition by Jaeger (2015) and having non-perceptual matches, according to Burte et al. (2019). Moreover, these specific questions (PAT4_58, PAT4_59, and PAT4_60) are challenging because they include folds which do not increase layers of paper, but they transfer to areas that had no paper (were empty) before the fold.

5.2 Relating PFT resolution strategies in the literature with our results

According to Cooper and Shepard (1973), processing the paper-folding-and-punched-hole test (PFT) involves first folding the mental image, then updating that mental image to include the hole punch, and then finally unfolding the mental image while keeping track of the newly added hole punch. In the studies by Jaeger (2015), a strategy that was significantly correlated with performance on the PFT items was the *work backwards strategy* in which participants started from the last fold and worked backwards to determine what the unfolded paper would look like.

Hegarty (2010) used a thinking aloud protocol to find out which strategies participants used to solve the PFT, and her results reported that most of the participants used (i) an *imagery strategy* (i.e., visualizing the fold noting where the holes would be, even working backward to unfold the paper and figure out where the holes would be), sometimes combined with (ii) a *spatial analytic strategy* (i.e., figuring out where one of the holes would be and then deleting answer choices that did not have a hole in that location) or even (iii) a *pure analytic strategy* (i.e., figuring out how many folds were punched to find out how many holes have the answer) and the combination of strategies improved success.

Burte et al. (2019) detected four strategies used in order to solve the PFT: (i) *folding-unfolding strategy* or imagining unfolding the paper to reveal the punch configuration (this requires high cognitive load and more folds could negatively impact accuracy); (ii) *perceptual match strategy* or matching the punch location in the probe to the punch locations in the response items; and (iii) *fold-to-punch algorithms* or applying the rule that each fold yields two holes after a punch.

The *Paper Folding Reasoning Videogame* provides a visual continuous representation of PFT by an animated sequence of fold

movements, which can be observed again using the repeat/reload button. It also provides a visual feedback/explanation after the player chooses an answer, that is, it shows an animation which unfolds the paper showing how the holes are transferred to different locations on it. This visual feedback/explanation might help players to generate their mental image of the problem, that is, to practice their *imagery strategy* (Hegarty, 2010), or their *work backwards strategy* by Jaeger (2015), or their *folding-unfolding strategy* by Burte et al. (2019). According to our data, the repeat/reload button was the option used by 56.0% of the participants in Part II (see Table 5), and our analysis also found out that there is a significant difference in players' performance who used the visual help.

Moreover, Doggo, the avatar in the *Paper Folding Reasoning Game*, provides feedback or text explanation messages whenever a participant fails to answer a PFT question (corresponding to the logic reasoning behind, see Falomir et al., 2021 for details). Some examples are as follows: Notably, this answer has less holes than layers of paper were punched; the hole was made in a piece of paper folded to the right, but more layers of paper were punched. This feedback might make the players discover the *pure analytic strategies* identified by Hegarty (2010) or the *fold-to-punch strategies* identified by Burte et al. (2019). Other feedback messages provided by Doggo the avatar are: <<Your answer has the same number of holes than the correct one, but they are in the wrong location>> and <<Note that the left-up corner paper area was not punched4>> which are messages that show or explain the *spatial analytic strategy* by Hegarty (2010) or the *perceptual match strategy* by Burte et al. (2019) to players, thus helping them acquire more strategies to face the next PFT question.

5.3 Relating our gender difference in performance with the literature

Spiers et al. (2023) analyzed the navigation ability of 3.9 million people through their gameplay data from *Sea Hero Quest* (from 18 to 99 years of age, 63 samples countries), showing a male advantage but varying considerably and could be partly predicted by gender inequality. In contrast, in another study by Yavuz et al. (2023) which analyzed the gameplay by US-based participants ($n = 822$, 280 men, 542 women, mean age = 26.3 years, range = 18–52 years) to the previous videogame *Sea Hero Quest*, no significant association was found between reliance on GPS and spatial navigation performance for either gender. They found a significant association between weekly hours of video gaming and navigation performance which was not moderated by gender.

In our analysis of players' gameplay in the *Paper Folding Reasoning Game*, we got a statistically significant difference in gender regarding their success rate from Part I to Part II in a small group (44 men vs. 48 women) since the most of the participants did not provide their gender.

6 Conclusion and future work

Paper folding-and-punched-hole tests are used to measure spatial abilities, and the *Paper Folding Reasoning Videogame* is an educational videogame developed to train these abilities. This manuscript presents the results of Analyzing the gameplay data

gathered by the *Paper Folding Reasoning Game* in its three parts. For Parts I and II, the data provided by 225 anonymous unique players are analyzed. For Part III (Mastermode), the data obtained from 894 gameplays by 311 anonymous unique players are analyzed.

Comparing players' performance in Part I and Part II, we found out that the players who completed the challenging exercises in Part I had a higher % of success in Part II, which indicates that Part I do train players' spatial skills regarding paper folding reasoning. Further Analyzing the data from the help options, we found out that players preferred visual help (reload button) vs. discarding-answers option, and that there is a significant difference in players' performance who used the visual help.

Comparing players' performance in Part III (Mastermode), we found out a significant difference in performance regarding the players who trained before playing the Mastermode (18.3%) with respect to those who did not train (81.7%). Moreover, we also observed that the 311 unique players in the Mastermode succeeded 48.1% of times in their first gameplay, increasing their percentage of success to 63.1% in recurrent gameplays.

These data results confirm the effectiveness of the *Paper Folding Reasoning Videogame* to train players' paper-folding-and-hole-punched reasoning skills. Moreover, the gameplay in this videogame might enhance players' strategies (reported in the spatial cognition literature) when facing the PFT.

The *Paper Folding Reasoning Videogame* provides a visual sequence of animated folds for each question and also an animated sequence of unfoldings, showing the solution to the question. This might train players' *imagery strategy* (Hegarty, 2010), or their *work backwards strategy* (Jaeger, 2015), or their *folding-unfolding strategy* (Burte et al., 2019). Moreover, the avatar's feedback or text explanation messages, whenever a participant fails to answer a PFT question, might make players discover the *pure analytic strategies* identified by Hegarty (2010) or the *fold-to-punch strategies* identified by Burte et al. (2019) (e.g., <<Note that this answer has less holes than layers of paper were punched>>) or the *spatial analytic strategy* observed by Hegarty (2010) or the *perceptual match strategy* by Burte et al. (2019) (e.g., <<Your answer has the same number of holes than the correct one, but they are in the wrong location>>). Moreover, according to Hegarty (2010), the more strategies applied by the participants, the higher their success.

Regarding gender differences in performance, the 34.8% of players who improved from Part I to Part II ($N = 141$) did not provide their data, whereas 65.2% of players provided their gender: 44 male players (mean age 25.5) who got better performance during Part I got a success rate of 67.6% in Part II, whereas the 48 female players (mean age: 25.1) with better performance in Part I got 53.1% success rate in Part II, resulting in a 14.5 % of difference. This difference between genders was statistically significant, according to the obtained p -value of < 0.01 .

As future study, we will extend our research to study players' spatial anxiety while training their skills using videogames.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, by request and after applying anonymization techniques to id devices.

Ethics statement

Ethical approval was not required for the studies involving humans because the collected data was from anonymous participants who downloaded a game through GooglePlay or AppleStore. Before starting the game, the participants were informed by a pop-up window that the gameplay data was stored, and that if they wished the system not to store it, they could disable the option in settings. No data regarding their gender, age, etc. were stored. No one was forced to play the game. Noone was paid for playing the game and the game was free of charge. The data collection took part between 2018 to 2022. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their consent by not disabling the storing data option.

Author contributions

PG-S: Data curation, Formal analysis, Visualization, Writing—original draft, Writing—review & editing, Software. VS: Software, Writing—review & editing, Validation. ZF: Validation, Writing—review & editing, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing—original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. ZF acknowledges the funding by the Wallenberg AI, Autonomous Systems and Software Program (WASP) awarded by the Knut and Alice Wallenberg Foundation, Sweden. PG-S acknowledges the project titled Explainable Artificial Intelligence for Healthy Aging and Social Wellbeing (XAI4SOC) funded by Spanish Ministry of Science, Innovation and Universities (MCIN/AEI/10.13039/501100011033) under project number PID2021-123152OB-C22. VS acknowledges the Ministry of Labor and Social Economy under INVESTIGO Program (Grant number: INVEST/2022/308) and the European Union and FEDER/ERDF (European Regional Development Funds).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of

their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., et al. (2013). Video game training enhances cognitive control in older adults. *Nature* 501, 97–101. doi: 10.1038/nature12486
- Baki, A., Kosa, T., and Guven, B. (2011). A comparative study of the effects of using dynamic geometry software and physical manipulatives on the spatial visualization skills of pre-service mathematics teachers. *Br. J. Educ. Technol.* 42, 291–310. doi: 10.1111/j.1467-8535.2009.01012.x
- Burte, H., Gardony, A. L., Hutton, A., and Taylor, H. A. (2019). Knowing when to fold 'em: Problem attributes and strategy differences in the paper folding test. *Pers. Individ. Dif.* 146, 171–181. doi: 10.1016/j.paid.2018.08.009
- Clark, D. B., Hernández-Zavaleta, J. E., and Becker, S. (2023). Academically meaningful play: Designing digital games for the classroom to support meaningful gameplay, meaningful learning, and meaningful access. *Comput. Educ.* 194:104704. doi: 10.1016/j.compedu.2022.104704
- Cooper, L. A., and Shepard, R. N. (1973). "Chronometric studies of the rotation of mental images," in *Visual Information Processing*, ed W. G. Chase (Brussels: Academic Press), 75–176.
- Coutrot, A., Kievit, R., Ritchie, S., Manley, E., Wiener, J., Hoelscher, C., et al. (2023). Education is positively and causally linked with spatial navigation ability across the life-span. *PsyArXiv*. doi: 10.31234/osf.io/6hfrs
- Ekstrom, R. B., French, J. W., Harman, H. H., and Dermen, D. (1976). *Manual for Kit of Factor-Referenced Cognitive Tests*. Princeton, NJ: Educational Testing Service.
- European Parliament (2021). *Promoting Gender Equality in Science, Technology, Engineering and Mathematics (STEM) Education and Careers*, P9_TA(2021)0296. Technical report, Committee on Culture and Education, Committee on Industry, Research and Energy, Committee on Women's Rights and Gender Equality. Brussels.
- Falomir, Z., Tarin, R., Puerta, A., and García-Segarra, P. (2021). An interactive game for training reasoning about paper folding. *Multim. Tools Appl.* 80, 6535–6566. doi: 10.1007/s11042-020-09830-5
- Hegarty, M. (2010). "Chapter 7 - components of spatial intelligence," in *The Psychology of Learning and Motivation*. Vol. 52 (Academic Press), 265–297.
- Highfield, K., and Mulligan, J. (2007). "The role of dynamic interactive technological tools in preschoolers' mathematical patterning," in *Proceedings of the 30th Annual Conference of the Mathematics Education Research Group of Australasia*, Vol. 1, eds J. Watson and K. Beswick (Hobart, TAS: MERGA), 372–381. Available online at: <https://researchers.mq.edu.au/en/publications/the-role-of-dynamic-interactive-technological-tools-in-preschoole>
- Jaeger, A. J. (2015). *What Does the Punched Holes Task Measure?* (PhD thesis), University of Illinois at Chicago, Chicago, IL, United States.
- Kane, M., Hambrick, D., and Conway, A. (2005). Working memory capacity and fluid intelligence are strongly related constructs. *Psychol. Bull.* 131, 66–71. doi: 10.1037/0033-2909.131.1.66
- Kim, Y. J., Knowles, M. A., Scianna, J., Lin, G., and Ruipérez-Valiente, J. A. (2023). Learning analytics application to examine validity and generalizability of game-based assessment for spatial reasoning. *Br. J. Educ. Technol.* 54, 355–372. doi: 10.1111/bjet.13286
- Kirschner, D., and Williams, J. P. (2014). Measuring video game engagement through gameplay reviews. *Simul. Gaming* 45, 593–610. doi: 10.1177/1046878114554185
- Lippa, R. A., Collaer, M. L., and Peters, M. (2010). Sex differences in mental rotation and line angle judgments are positively associated with gender equality and economic development across 53 nations. *Arch. Sex. Behav.* 39, 990–997. doi: 10.1007/s10508-008-9460-8
- Newcombe, N. (2010). Picture this: Increasing math and science learning by improving spatial thinking. *Am. Educ.* 34, 29–35.
- Nicolaidou, I., Chrysanthou, G., Georgiou, M., Savvides, C., and Toulekkis, S. (2021). "Relationship between spatial reasoning skills and digital puzzle games," in *European Conference on Games Based Learning* (Brighton), 553–560, XIX.
- Qian, M., and Clark, K. R. (2016). Game-based learning and 21st century skills: a review of recent research. *Comput. Human Behav.* 63, 50–58. doi: 10.1016/j.chb.2016.05.023
- Sorby, S. A. (2009). Educational research in developing 3D spatial skills for engineering students. *Int. J. Sci. Educ.* 31, 459–480. doi: 10.1080/09500690802595839
- Spence, I., and Feng, J. (2010). Video games and spatial cognition. *Rev. Gen. Psychol.* 14:92. doi: 10.1037/a0019491
- Spiers, H. J., Coutrot, A., and Hornberger, M. (2023). Explaining worldwide variation in navigation ability from millions of people: citizen science project sea hero quest. *Top. Cogn. Sci.* 15, 120–138. doi: 10.1111/tops.12590
- Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., and Newcombe, N. S. (2014). Finding the missing piece: blocks, puzzles, and shapes fuel school readiness. *Trends Neurosci. Educ.* 3, 7–13. doi: 10.1016/j.tine.2014.02.005
- Wai, J., Lubinski, D., and Benbow, C. P. (2009). Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. *J. Educ. Psychol.* 101, 817–835. doi: 10.1037/a0016127
- Yavuz, E., He, C., Goodroe, S., Ganstrom, C., Coutrot, A., Hornberger, M., et al. (2023). Video gaming, but not reliance on gps, is associated with spatial navigation performance. *bioRxiv*. doi: 10.1101/2023.08.10.552365

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2024.1303932/full#supplementary-material>