Out of the Maze: Investigating Fluid Intelligence and Numeracy as Predictive Factors of Planning Skills Using Video Games

Gianluca Guglielmo[1(✉)], Elisabeth Huis in't Veld[1,2], Michał Klincewicz[1,3], and Pieter Spronk[1]

[1] Department of Cognitive Science and Artificial Intelligence, Tilburg University, Warandelaan 2, 5037 AB Tilburg, The Netherlands
G.Guglielmo@tilburguniversity.edu
[2] Sanquin, Department of Donor Medicine Research, Plesmanlaan 125, 1066 CX Amsterdam, The Netherlands
[3] Institute of Philosophy, Department of Cognitive Science, Jagiellonian University, Grodzka 52, 31-048 Krakow, Poland

Abstract. The aim of this study was to test whether an online video game can be used to investigate planning ability and whether fluid intelligence, objective numeracy, and subjective numeracy are predictive of game performance. Our results demonstrate that fluid intelligence is particularly important, which is in line with previous non-game-based studies that show a relationship between classical planning tests and fluid intelligence. Video games have been previously used for research into cognitive processes and taking them online facilitates data collection on a larger scale. Online video games also afford data collection without the expense and stress of a laboratory environment. For these reasons, using online video games to investigate human cognition is a promising alternative to the classic cognitive paradigms used in laboratories.

Keywords: Cognitive skills · Numeracy · Fluid intelligence · Planning capabilities · Video games

1 Introduction

The ability to plan is useful in all areas of life. Individuals that are especially good at planning are typically faster in coming to a decision. Planning, defined as "the predetermination of a course of action aimed at achieving some goals" [1], is often operationalized as sequential decision-making and prediction of an optimal, or at least satisfying, course of action. According to previous research, planning skills are associated with better occupational outcomes, including income [2], and additionally, planning performance is also likely to be associated with more foundational skills, such as numeracy and fluid intelligence [2, 3]. Fluid intelligence and objective numeracy are involved in problem-solving, including calculations and mental rotation of objects. Subjective numeracy is a

K. Kiili et al. (Eds.): GALA 2022, LNCS 13647, pp. 202–211, 2022.
https://doi.org/10.1007/978-3-031-22124-8_20

measure of confidence in one's ability to solve problems that involve numbers, including calculations. Subjective numeracy also measures the preference that one has to deal either with the information presented as numbers or in prose [4, 5].

However, pen and paper versions of psychometric tests are more difficult to use in experimentation. Digital versions of these tests could facilitate the automatic, large-scale assessment of these skills in an easy and remote way. Additionally, video games could be suited for this purpose, as they have been shown to be effective for cognitive skills measurement [6] and could therefore provide a useful tool able to facilitate the study of individual differences in planning skills outside the lab, for non-academics and businesses. Video games have already successfully been used to test cultural differences [7], introduce new business concepts [8], and hire new employees [9], just to name a few. The idea of introducing a planning test as a sort of game was already introduced by [3]. However, no online video game was developed yet to directly measure planning tasks in an online fashion. The aim of this study is to develop an online game able to measure planning skills; in doing so we would like to suggest that online video games can be used to track specific skills and their relationship with more fundamental skills such as fluid intelligence or numeracy We do this by using a game specifically created to engage the planning process. Additionally, we hypothesize that fluid intelligence and numeracy are significant predictors of performance in the game.

2 Related Works

2.1 Planning Tests and Gamification

Several tasks have been developed and used to measure planning capabilities, the most famous of which are the Tower of London (TOL) [10] and the Tower of Hanoi (TOH) [11]. In these tasks, participants have to reorder either balls (in case of the TOL) or disks (in case of the TOH) given certain constraints. Both have been extensively used in different domains such as in child and adult neuropsychology to test, for example, the effect of obsessive-compulsive disorders [12], ADHD [13], and autism [14] on planning capabilities. Other existing planning tasks involve grids or mazes; e.g. the Porteus Maze Test [15] in which the subject has to draw the correct path through as many mazes as possible while avoiding touching the walls of the mazes or the errands-task [1] where participants have to stop by shops, represented as points on a grid, in a fictitious town center. The goal is to find the path with the shortest distance from a starting point to the endpoint while at the same time being able to visit all the shops present on the grid.

Tests, implemented using grids or mazes, and tasks that involve the optimization of the distance traveled, can be used to create a video game to investigate planning skills. In this study, we designed and implemented a video game incorporating the aforementioned features with the intention to investigate the players' capability to plan ahead.

2.2 Effect of Numeracy and Fluid Intelligence

Individual differences in planning skills are sometimes due to differences in the level of (fluid) intelligence, which is defined as the ability to solve novel problems by reasoning and is typically measured by the performance of mental operations, such as solving geometrical puzzles. This may be why fluid intelligence has been identified as a strong predictor of planning performance [3] and more specifically on the TOL and the TOH [16] tasks. Differences in planning ability are also sometimes due to individual differences in numeracy, which is sub-divided into subjective numeracy and objective numeracy. Objective numeracy, often referred to as statistical numeracy, refers to the ability to deal with numerical problems involving probability and statistics. Subjective numeracy refers to the motivational aspect of decision-making and the willingness of the subject to engage in tasks concerning risk appraisal and in general numerical problems. Participants with high scores in subjective numeracy tend to have a more positive attitude toward numerical problems and adopt more complex strategies and fewer heuristics while making decisions [17]. Moreover, when combined, numeracy skills and fluid intelligence are predictors of avoiding decisions with negative consequences and optimizing monetary outcomes [17].

3 Methods

3.1 Sample and Procedure

A total of N = 78 participants were recruited from the University of Tilburg (Netherlands) and received course credits. The sample consisted of 23% (n = 18) men, 75.6% (n = 59) women, and n = 1 refused to provide their biological sex. The participants were $M = 21.27$ ($SD = 3.23$) years old and had $M = 13.64$ ($SD = 4.10$) years of education. The sample size obtained was a posteriori analyzed to evaluate the level of power reached considering the number of predictors and the outcomes used in this study (both conveyed in Sect. 4.3). Such analysis was performed considering a medium Cohen's f-squared effect size of 0.29 calculated using the correlation among predictors and between predictors the outcome [18]. The analysis, performed with the G*Power software yielded an approximated total power of 0.95.

The experiment was run during the Covid-19 pandemic, for this reason, data were collected online avoiding direct contact between participants and experimenters. Participants started the experiment with an online questionnaire in Qualtrics filling in demographic information, after which they completed the cognitive skills tasks. After having completed these tests, participants were directed to play the video game using the link specified in Appendix A. To enter the game, all participants had to first enter a unique identifier and a password and then read instructions and complete a practice level of the game. Afterward, their scores were directly sent to the experimenters via an email created for the purpose.

3.2 Cognitive Skills

Fluid Intelligence. Raven's standard progressive matrices (RPM) was used to measure fluid intelligence. The test consists of 60 patterns (and thus a maximum score of 60), where the participants are asked to complete the pattern by choosing the missing tile. The participants had 20 s to solve each of the patterns, with a total of 20 min for the whole task [19]. This resulted in a good internal consistency of Cronbach's alpha of 0.89 whereas other studies suggested an expected alpha between 0.86 and 0.92 [20].

Numeracy Skills. Objective numeracy was tested using a digitized version of the Berlin Numeracy test (BNT). Participants were asked four numeracy-related questions (with a maximum score of 4), to be completed in 5 min, requiring the participants to be able to work with percentages and proportions (they were allowed to use pencil and paper for calculations). An internal consistency of $\alpha = 0.54$ was obtained (slightly lower than the expected 0.60 [17]). Subjective numeracy was investigated using the Subjective Numeracy Scale (SNS), assessing the attitude that a subject has towards numerical problems. The SNS is composed of 8 questions, scoring on a Likert scale ranging from 1 to 6. The first 4 questions of the SNS are used to measure how confident a participant is in dealing with numerical information (self-efficacy or confidence) [4]. The last 4 are used to measure the participants' preference for information to be presented numerically or in prose. Cronbach's α in this sample was 0.74 (a bit lower than the expected 0.86) [17]. The time it took to complete all cognitive skills tasks was also recorded.

3.3 Videogame Design and Rules

The video game was implemented in Unity and was made available online (Appendix A). Its objective is to move a small yellow circle across a grid to the exit. Each grid contains 40 obstacles (red triangles) arranged in a different order, and 5 targets (green circles) to collect before reaching the end. All grids are drawn using tools provided at: https://www.theedkins.co.uk/jo/maze/design/index.htm.

We set up each game session with two conditions. In the first condition, participants were instructed to minimize the total distance to the end. In the second, they had to minimize the number of obstacles walked on (red triangles). In both conditions, participants were asked to finish the tasks as quickly as possible and could see the elapsed time displayed during the game. The novelty in this game, compared to other paradigms [1, 15], is the use of obstacles in the second condition used to evaluate the planning task. These details, to a certain extent, may be similar to real-life tasks where people have to plan the best course of action considering the presence of obstacles that may occur along the way (e.g. in logistics).

Each participant's session consisted of 6 games with 3 grids per condition where the same grid was never presented twice in a row. All the participants were presented with the same grids for both the triangles and the distance domain in the same order, in both conditions, lower scores represented higher performance. The game provided participants with real-time feedback about their performance in the game measured in terms of time needed and either distance or number of triangles (see Fig. 1). The participants were either asked to minimize the number of obstacles hit (the red triangles) or the total distance to reach the end while collecting the tasks (the green dots). In our game, lower scores represent better performance since in theory, proficient participants manage to reach the end, collecting the tasks, hitting fewer triangles, and walking a shorter distance. The Cronbach's alpha for both triangle and distance conditions (calculated on 78 participants) was equal to 0.70.

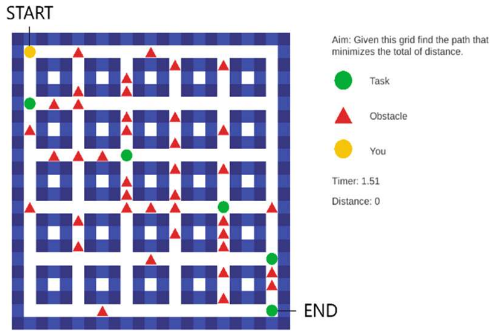

Fig. 1. An example of a game grid. The green dots are targets that the participant has to 'collect' by moving the yellow dot over them and the red triangles are obstacles to avoid (only applicable in condition 2). The instruction and performance are shown on the right of the maze (Color figure online).

3.4 Data Pre-processing

The demographic data and the results of the administered tests were extracted from .csv files obtained via Qualtrics. The pre-processing was performed using the Pandas, and NumPy libraries available in Python.

During the stage of the analysis, the participant not providing their biological sex was discarded. The biological sex information was encoded as 0 = female, and 1 = male. Scores on the cognitive skills tasks were calculated according to the scoring instructions of each task, and then standardized (standardization is a process that rescales the data between 1 and −1 where the mean is around 0). Such a process was performed in order to avoid the effect of data having a different magnitude on the outcome variable [17]. The performance in the game was measured as the total distance in the first condition, and the total number of obstacles hit in the second condition. Hence, a higher score indicates a lower performance. These scores were also standardized to a range between −1 and 1 and combined into one score to assess the overall score of the participants.

4 Results

After having performed the pre-processing, descriptive information was provided to evaluate the potential correlations between predictors. Afterward, we run analyses to evaluate if the two biological sexes differed in their score on the predictors used. Eventually, a regression analysis was run using the cognitive skill scores as predictors (adding age, biological sex, and years of education to the model as control variables) and the total score on the planning game as the dependent variable. The analyses in this section were run using the SciPy and the Statsmodels library in Python.

4.1 Descriptive Information

Participants had an average fluid intelligence score of 43.52 ($SD = 4.76$). Furthermore, the participants scored $M = 1.32$ ($SD = 1.06$) on objective numeracy and $M = 33.7$

($SD = 5.12$) on subjective numeracy. Our results show moderate and weak correlations among subjective and objective numeracy ($r = 0.29$, $p = .01$), between fluid intelligence and objective numeracy ($r = 0.42$, $p < .001$), and between subjective numeracy and fluid intelligence ($r = 0.28$, $p = .01$). However, an analysis to evaluate multicollinearity in the data, showed a variance inflation factor below 2 for all the predictors, including control variables, chosen for this study. Such results suggest a lack of multicollinearity in the data [21].

4.2 Independent Variables and Differences Between Biological Sexes

No gender differences were found in age ($t(34.62) = 1.85$, $p = .07$), years of education ($t(48.76) = 0.46$, $p = .65$), objective numeracy ($t(25.09) = -0.61$, $p = .55$) or fluid intelligence ($t(25.07) = -0.55$ $p = .59$). However, women ($M = 33.02$, $SD = 5.27$) scored lower on subjective numeracy than men ($M = 35.8$, $SD = 5.27$; $t(36.01) = 2.37$, $p = .02$) and therefore, an interaction term accounting for the effect of biological sex on subjective numeracy was added in the regression model. Given the homogenous (student) sample, age and years of education were not used as a term for potential interactions.

4.3 Predictors of Planning Performance

To assess the effect of fluid intelligence, objective numeracy, and subjective numeracy on the score obtained in the game, a multiple linear regression was run. These predictors explained 23% of the variance obtained in the combined planning game score ($F(7,69) = 2.9$, $p = .01$). But only fluid intelligence significantly predicted performance in the planning game (See Table 1).

Table 1. The results obtained using a multiple linear regression.

	B	SD	β	U	L	t	*p*
Biological sex (ref: Female)	−0.04	0.08	−0.06	0.27	−0.46	−0.52	.603
Age	0.08	0.08	0.13	0.23	−0.07	1.13	.263
Years of education	−0.04	0.07	−0.06	0.10	−0.18	−0.51	.615
Fluid intelligence	−0.24	0.08	−0.38	−0.09	−0.40	−3.10	.003*
Objective numeracy	−0.04	0.08	−0.06	0.12	−0.19	−0.49	.627
Subjective numeracy	−0.06	0.08	−0.15	0.07	−0.26	−1.14	.256
Subjective numeracy x Biological sex	0.07	0.09	0.10	0.57	−0.26	0.75	.458

4.4 Planning Game Score

Figure 2 shows the distribution of the total scores of both conditions on the planning game, by fluid intelligence scores. On average, in the first condition, the mean distance was equal to 84.05 ($SD = 26.98$) and in the second condition, $M = 9.02$ ($SD = 2.00$) triangles were hit across the 3 grids. The optimal distance and number of triangles across the 3 grids were respectively 56.67 and 6.00. 26 participants managed to reach the optimal distance score while only 2 managed to obtain the optimal triangles score. No participant obtained optimal performance on both domains. Also, significant correlations between fluid intelligence and the scores ($r = -0.23$, $p = .04$ and $r = -0.29$, $p = .01$) were found across the 2 conditions.

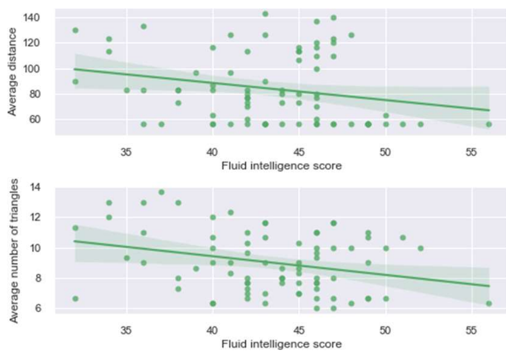

Fig. 2. Scatter plots showing the relationship between fluid intelligence and the scores obtained in condition 1 (minimizing the distance; top) and condition 2 (minimizing number of obstacles hit).

4.5 Relationship of Time Variations in Planning Game and Predictors

The participants required an average of 117.99 s to complete all the 6 proposed grids ($SD = 43.41$), where males took on average 114.41 s to fulfill all the grids ($SD = 49.96$) while females took 119.09 s ($SD = 41.14$); this difference was not statistically significant ($t(24.15) = -0.35$, $p = .73$). On average, participants significantly required more time to complete the triangles domain ($M = 63.03$ s, $SD = 23.49$) than the distance domain ($M = 54.97$ s, $SD = 25.02$) ($t(152) = 2.05$, $p = .04$).

Time-related analyses were also run on the time the participants required to fulfill the tests related to predictors measuring performance in Fluid intelligence and Objective numeracy; such time was based on the last click registered by Qualtrics on the test pages. Participants required 660.47 s ($SD = 130.47$) to complete the RPM where females took 650.26 s ($SD = 130.50$) and males 693.23 s ($SD = 125.10$); however, this difference was not statistically significant ($t(28.72) = -1.25$, $p = .22$). For what concerns the BNT, participants took 237.72 s ($SD = 53.13$) to complete the test; where males and females took respectively 251.05 s ($SD = 44.58$) and 233.66 s ($SD = 54.83$) to complete the task. Such a difference was not significant having a $t(33.49) = 1.34$ and a $p = .19$.

5 Discussion

The planning game we developed for the study reported here required players to plan and optimize their moves through a grid and avoid obstacles. Sequences of player moves displayed a wide range of variability, reflecting the planning and optimization processes at work. We found that higher fluid intelligence scores predict a better-planned and more optimized sequence of moves on the route through a grid to a goal. Fluid intelligence involves processes responsible for manipulating information in real-time, abstraction, and general reasoning, which suggests that these are important mechanisms for planning ahead and optimizing during gameplay. This is not surprising since fluid intelligence has previously been shown to be important to planning [3] and a wide range of complex behaviors and skills [20, 22].

Previous studies showed that numeracy is a predictor of superior decision-making in problems with uncertain and probabilistic outcomes [17]. Our game involved some uncertainty since the grids were complex enough to make it impossible for players to plan their entire route before making their first move. Unexpectedly, we did not find any significant relationship between numeracy and performance. While we found that males tend to have higher subjective numeracy scores compared to women, which means that they are more confident about their ability to solve numerical problems, they did not perform better in the game compared to women. Our results, also suggested that participants required longer to fulfill the triangles domain. Such results, together with the fact that only two participants found the optimal path for the triangles domain, suggest that the triangles domain may be harder than the distance domain.

We also found no effect of age and years of education on performance. This is surprising since previous research showed that age and years of education predict performance in solving mazes, but measures of intelligence do not [15]. Our study points in the other direction. That said, our sample was relatively young and homogenous with respect to age and years of education, as they were all students in the same University, so we remain cautious about just how significant this may be.

One limitation of our study is that it does not use a validated measure of planning performance. This means that we cannot draw strong conclusions about the extent the game we developed captures planning ability. We assumed that it does, because of its similarity to mazes widely used in psychology research to study spatial and social cognition [23], and especially those used to investigate planning, such as the Porteus Maze Test [15], or the more recently developed Electric Maze Test [3]. Our work gamifies these tests by adding extra mechanics at the expense of introducing potential confounds. We are confident about validity since previous work demonstrates that planning skills are important in video games that involve strategy [24].

Our work also makes a methodological contribution. We showcase how video games created using widely-accessible tools can be used to study human cognitive processes. Playing video games online requires little time and affords data to be sent automatically to the experimenter, which facilitates data collection significantly and has the added benefit of being outside of a laboratory setting. Furthermore, an online game like ours can be combined with non-invasive psychophysiological measures based on webcam recordings. This affords an opportunity to investigate the connection between behaviors and psychophysiological reactions to make inferences about psychological processes. Given all this, our paper demonstrates the potential of video games to tap into the same processes, constructs, and behaviors that are engaged in laboratory experiments, and real-life situations.

Acknowledgments. The research reported in this study is funded by the MasterMinds project, part of the RegionDeal Mid- and West-Brabant, and is co-funded by the Ministry of Economic Affairs and Municipality of Tilburg.

Appendix A

Planning game link: https://play.unity.com/mg/other/my-new-microgame-4526.
Identifier: Test2022, Password: logistics.

References

1. Garling, T.: Processing of time constraints on sequence decisions in a planning task. Eur. J. Cogn. Psychol. **6**(4), 399–416 (1994)
2. Fernandez, F., Liu, H.: Examining relationships between soft skills and occupational outcomes among US adults with—and without—university degrees. J. Educ. Work. **32**(8), 650–664 (2019)
3. Sheppard, K.W., Cheatham, C.L.: Validating the electric maze task as a measure of planning. J. Cogn. Dev. **18**(2), 309–322 (2017)
4. Peters, E., Shoots-Reinhard, B.: Numeracy and the motivational mind: the power of numeric self-efficacy. Med. Decis. Making, 729–740 (2022)
5. Zikmund-Fisher, B.J., Smith, D.M., Ubel, P.A., Fagerlin, A.: Validation of the subjective numeracy scale: effects of low numeracy on comprehension of risk communications and utility elicitations. Med. Decis. Making **27**(5), 663–671 (2007)
6. Kokkinakis, A.V., Cowling, P.I., Drachen, A., Wade, A.R.: Exploring the relationship between video game expertise and fluid intelligence. PLoS ONE **12**(11), e0186621 (2017)
7. NorouzzadehRavari, Y., Strijbos, L., Spronck, P.: Investigating the relation between playing style and national culture. IEEE Trans. Games (2020)
8. Buiel, E., et al.: Synchro mania-design and evaluation of a serious game creating a mind shift in transport planning. In: 46th International Simulation and Gaming Association Conference, ISAGA, pp. 1–12 (2015)
9. Allal-Chérif, O., Bidan, M.: Collaborative open training with serious games: relations, culture, knowledge, innovation, and desire. J. Innov. Knowl. **2**(1), 31–38 (2017)
10. Phillips, L.H.: The role of memory in the Tower of London task. Memory **7**(2), 209–231 (1999)
11. Kotovsky, K., Hayes, J.R., Simon, H.A.: Why are some problems hard? Evidence from the Tower of Hanoi. Cogn. Psychol. **17**(2), 248–294 (1985)
12. Delorme, R., et al.: Shared executive dysfunctions in unaffected relatives of patients with autism and obsessive-compulsive disorder. Eur. Psychiatry **22**(1), 32–38 (2007)
13. Riccio, C.A., Wolfe, M.E., Romine, C., Davis, B., Sullivan, J.R.: The Tower of London and neuropsychological assessment of ADHD in adults. Arch. Clin. Neuropsychol. **19**(5), 661–671 (2004)
14. Robinson, S., Goddard, L., Dritschel, B., Wisley, M., Howlin, P.: Executive functions in children with autism spectrum disorders. Brain Cogn. **71**(3), 362–368 (2009)
15. Krikorian, R., Bartok, J.A.: Developmental data for the Porteus maze test. Clin. Neuropsychol. **12**(3), 305–310 (1998)
16. Zook, N.A., Davalos, D.B., DeLosh, E.L., Davis, H.P.: Working memory, inhibition, and fluid intelligence as predictors of performance on Tower of Hanoi and London tasks. Brain Cogn. **56**(3), 286–292 (2004)

17. Sobkow, A., Olszewska, A., Traczyk, J.: Multiple numeric competencies predict decision outcomes beyond fluid intelligence and cognitive reflection. Intelligence **80**, 101452 (2020)
18. Faul, F., Erdfelder, E., Buchner, A., Lang, A.G.: Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav. Res. Methods **41**, 1149–1160 (2009)
19. Hamel, R., Schmittmann, V.D.: The 20-minute version as a predictor of the Raven Advanced Progressive Matrices Test. Educ. Psychol. Measur. **66**(6), 1039–1046 (2006)
20. Al-Bokaia, H., Al-Subaihib, A.A.: Standard Progressive Matrices (SPM): validity and reliability. Int. J. Innov. Creativity Change **15**(4), 276–293 (2021)
21. Akinwande, M.O., Dikko, H.G., Samson, A.: Variance inflation factor: as a condition for the inclusion of suppressor variable(s) in regression analysis. Open J. Stat. **5**(07), 754 (2015)
22. Kulikowski, K., Orzechowski, J.: All employees need job resources: testing the" Job Demands-Resources Theory" among employees with either high or low working memory and fluid intelligence. Medycyna Pracy **69**(5) (2018)
23. Dudchenko, P.A.: 'A history of 'maze' psychology', Why People Get Lost: The Psychology and Neuroscience of Spatial Cognition (2010)
24. Gabbiadini, A., Greitemeyer, T.: Uncovering the association between strategy video games and self-regulation: a correlational study. Pers. Individ. Differ. **104**, 129–136 (2017)