Mirrors in Smog City – A serious game to assess collaboration potential

Rahul Kochar[1], Andrejs Kārkliņš[1], Tobias van den Hurk[1], Tomohiro Akutsu[1], Gijs Paardekooper[1], Robert Kooij[2], Rafael Bidarra[3]

Delft University of Technology, Netherlands
obsonebuttonstudio@gmail.com[1], {R.E.Kooij[2], R.Bidarra[3]}@tudelft.nl

Abstract. Collaboration is a critical factor for a team to be successful. Current practices to measure collaboration potential include many methods that are outdated, cumbersome or error prone, and therefore often inapplicable in many settings. Games have been increasingly applied to assess a variety of skills and the popularity of multi-player games makes them a strong alternative for measuring the collaboration potential of a team. We propose Mirrors in Smog City , a novel serious game in which the final score is explicitly conceived as a proxy for the collaboration potential of the players. This multi-player collaborative game is designed for simplicity and accessibility, so that it can be used by very diverse teams, with minimal on boarding in many settings. Although the current prototype still needs a thorough evaluation, our preliminary assessment has concluded that the current design seems to be well-suited for assessing the collaboration potential of its players. Apart from the small sample size, a thorough evaluation will also need to include a benchmark with an objective and, if available, standardized method for assessing collaboration potential.

Keywords: Serious game · Team collaboration · Communication · Teamwork · Simple game controller

1 Introduction

In the last VUCA (Vision, Understanding, Clarity and Agility) World[1], managers of organizations stressed the need to adapt and improve to defeat challenges like Covid-19 and climate change. One efficient way to prepare for a rapidly changing world is Human Resource Development [9]. According to Dhillon et al., [4], organizational culture is an increasingly important topic, and for that, collaboration is essential. Moreover, they describe collaboration beyond the position in an organization, such as collaboration between managers and employees, leading to flexibility of thought, so that an organization can respond to fierce changes in its external environment.

According to Bedwell et al., [2], collaboration is a higher-level process that encompasses many frequently studied constructs such as *cooperation*, *teamwork*

[1] https://www.vuca-world.org/

and *coordination*. Often collaboration is thought of as strongly correlated and/or connected with teamwork. Driskell et al. [6] say that 'At a broad level, teamwork is the process through which team members collaborate to achieve task goals'. Teamwork refers to the activities through which team inputs translate into team outputs such as team effectiveness and satisfaction. These notions imply that collaboration is a kind of process for several people to achieve a common goal. We assume that there is a definition of 'good collaboration' based on some subjective and context-dependent criteria.

Collaboration potential, or the potential to collaborate, is the degree to which a group of people are skilled to perform group tasks together. These tasks, by definition, cannot be performed by an individual alone or would take more time. Quantifying collaboration potential within a group of people can enable the formation of the team that is best equipped to tackle a problem, be it strategic, commercial, military or in any other domain.

Despite abundant research effort, a single definition of collaboration has not yet emerged. Even today, a lot of research often depends on questionnaires to assess traits around collaboration, teamwork, communication skills, etc. These methodologies may often be affected by individual biases in the survey formulations, by bias among the participants answering the survey, and by the subjectivity of their answers. We posit that a carefully designed collaborative game can provide a much better means for efficiently assessing the collaboration potential within a team, in an immersive and compelling setting.

2 Related work

Tong and Chignell [12] showed that a serious game can be used to perform a cognitive assessment. Among others, they measured the 'working memory' of the participant. Tong et al. [13] performed a validation study to check if using a serious game to conduct a cognitive assessment is valid and they found that it was indeed a valid use case. Their work showed that a serious game can be used to quantify traits that are hard to measure by other means. It has also been shown that it is possible to measure the degree of a certain bias in a group of people, e.g. using the serious game *Stranded* [7]. In general, there is an increasing body of evidence that serious games, if properly designed, are a powerful and versatile instrument for skills and behavior assessment.

In 2005, Dougherty and Larson [5] measured collaboration between a nurse and a physician using questionnaires. The disadvantage of paper questionnaires and interviews is the risk of 'favoring' socially desirable answers. Accommodating assessment in a game can provide a richer, less repetitive and more immersive context in which players might act more naturally and intuitively. In order to design games for structured assessment existing methods do not sufficiently accommodate interactions between game, the player's mental state and decisions in the game.

Schneider et al. [10] discuss measuring collaboration between two individuals through techniques like computer vision, to identify facial expressions, but these can likely be fooled with a poker face or high-quality acting. The authors claim

to measure collaboration, but in fact, what they measure is rather whether two people fit each other and would be comfortable working with each other.

Rousseau et al. [14] established a framework for a model of collaborative working. Their framework contains numerous factors like team building, goals, experience and communication. For example, they state that 'communication in collaborative work underpins how people understand each other and how knowledge is transferred. (...) The more that the right information is conveyed to the right people at the right time in the most appropriate way, the more effective collaboration will be.'. Collaboration, they conclude, means cooperating with each other to achieve a common task/objective.

Mayer et al. [8] proposed a 3D multiplayer puzzle game *TeamUP* to assess team performance. Players collaborate to solve puzzles, and teamwork is measured by quantitative indicators like 'time needed to complete the task', 'speak time' and 'avoidable mistakes'. They observed that the teams in which team members had high cohesion were able to complete the game in less time. They suggest the time needed to complete the game, or the number of mistakes made, can be used to measure collaboration in a team.

Some collaborative serious games have gone a step further, attempting to improve rather than simply assess the conditions for collaboration in a team. The serious game Grapplenauts [1], for example, was designed to accelerate ideation conditions for newly formed teams, by progressively challenging its members, first individually, and then requiring from them increasing collaboration. Similarly, the collaborative game *Maze Maestro* [11] aimed at helping break the ice within new teams, by spreading team members in a maze, with limited vision, and enforcing their collaboration to find the exit together.

Wang et al. [15] identified four game mechanic categories to promote collaboration and indicators to assess the effectiveness of collaboration. The first category is *Space*: by isolating players in their own space, they can be given individual tasks, while combining them within a shared space will foster collaboration. The second category relates to *Rules and Goals*, and it can include common goals, joint rewards and group victories. The third category is *Skills*, regarding the roles and abilities of individual players. Lastly, the category *Chance* refers to surprise elements that add fun to the game and create unexpected situations, which give the players additional opportunities to collaborate and thus, extra events where collaboration potential can be measured.

They also described two methods of measuring collaboration: *self-report*, where participants report how they felt collaborating, and *data mining* which uses collected gameplay data, like chatting logs, behavior logs, video footage or audio recordings. In the latter, researchers have to analyze the data after the game, which implies that conclusions or feedback are not derived in real time.

For so-called competitive-collaborative serious games, which include both competitive and collaborative gameplay elements, Buchinger et al. [3] identified various design features. They claim that games with the following elements tend to be more effective to promote knowledge acquisition:

- **Synchronization:** multiple players playing at the same time is better than turn-based
- **Roles:** players are often uniquely skilled
- **Score:** motivates players to do well and push themselves to do even better
- **Challenge:** gives an achievement feeling when the level/game was solved
- **Reward:** motivates players to finish the level/game and try to perform better for better rewards
- **Operationalization:** sufficient players/controllers/resources must be available

Finally, the authors also point out additional guidelines, two of which are most relevant to our research: (i) they suggest that gameplay should strongly induce *communication*, especially in a novel team, because otherwise individual characteristics might dominate over a united team's characteristics; (ii) they recommend that *assessments* should be neatly and cleanly integrated into gameplay; otherwise, if assessments are superimposed to the game, some players might get distracted from the game itself.

3 Game design

From the research work surveyed above, we have formulated the following Research Questions (RQ):

1. How could a game measure a team's collaboration potential with minimum dependence on player feelings, biases and skills?
2. How can we make that game into an expedite and very accessible measurement method?

In this section, we identify our set of design requirements for such a game, and introduce the main game loop and mechanics of Mirrors in Smog City , the game that we came up with as an answer to these questions.

3.1 Design requirements

Taking into account all recommendations from previous work, we identified a number of guidelines that we pose in the form of Game Requirements (GR):

1. To truly measure a team's collaboration potential, each player should have an equal and indispensable contribution, so that fewer than all players cannot succeed nor get a good score. (RQ 1)
2. This score should be automatically derived from player data and performance, not from their answers to any survey. (RQ 2)
3. The score for collaboration potential should be objectively and quickly given to the team. (RQ 1, 2)
4. The game mechanics should be very intuitive, so that most time is spent on collaborating, not on learning to play the game. (RQ 2)
5. For the same mechanics, the game should offer easy and hard difficulty modes, to tailor for various target groups. (RQ 1, 2)
6. The simplicity of the gameplay should make it accessible to a wide range of people, yet keep it challenging and fun for all teams. (RQ 2)

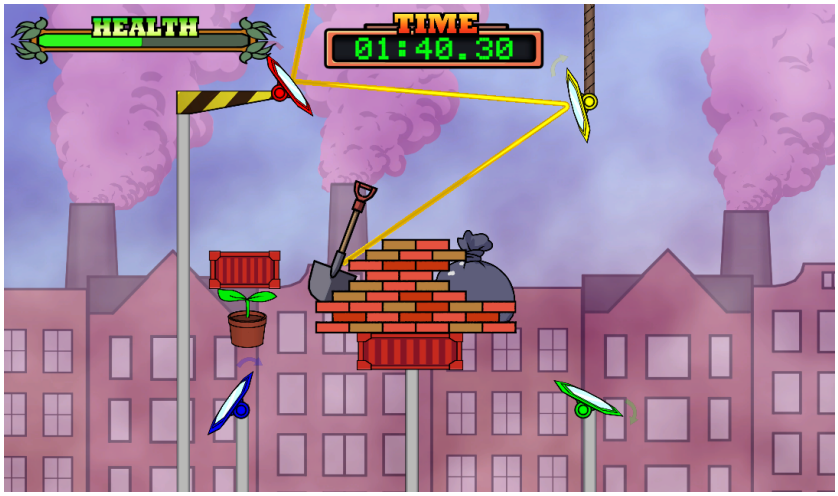


Fig. 1: The polluted urban environment of Mirrors in Smog City

3.2 Game story and synopsis

A city is polluted by heavy, dense and filthy smog, which is pierced by just one ray of sunlight (see Figure 1). Players have to work together, using mirrors to direct that light ray onto a plant. Each player controls one mirror and, at each level, the team has to solve increasingly challenging puzzles. Upon each success, the plant grows and, as a result, the city air gets cleaner.

3.3 Main game loop

Rousseau et al. [14] propose that there are three stages when a team tackles a common task (see Figure 2): (i) during the *preparation* stage, the team communicates to sketch their approach to solve the challenges at hand; (ii) at *execution*

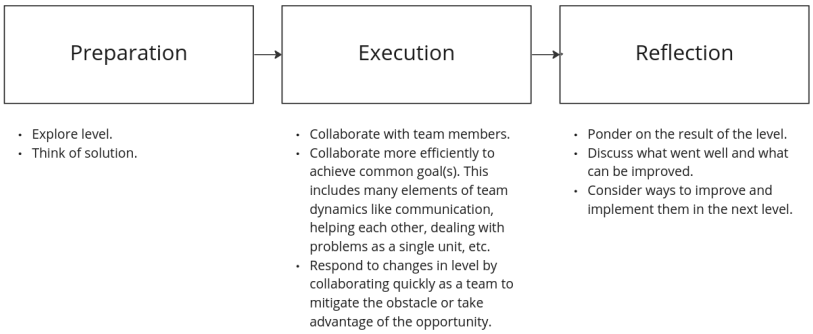


Fig. 2: The three stages of tackling a common task

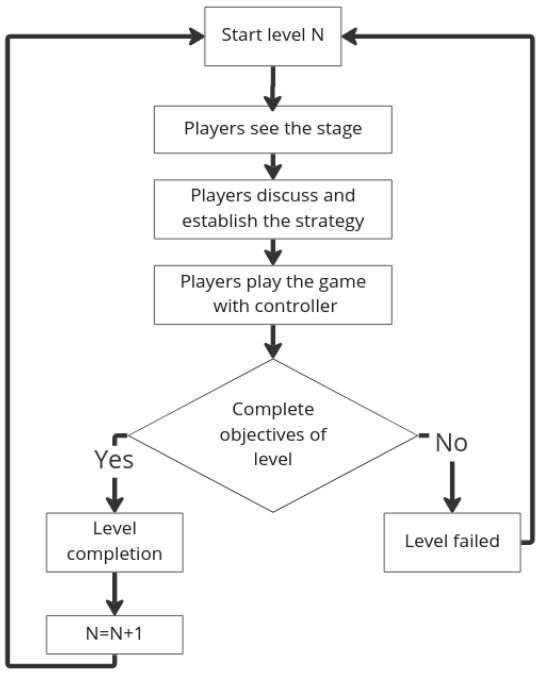


Fig. 3: Detailed main game loop

stage, the team implements what they planned and improvises along the way to deal with unplanned events; and (iii) at *reflection* stage, after completing the task, the team evaluates what went well and how to improve what did not go well for the next level.

We structured the core game loop of Mirrors in Smog City along these same stages, as indicated in Figure 3.

3.4 Core mechanics

To fulfill the game requirements in Subsection 3.1, we came up with the following core game mechanics:

1. **Light, mirror reflection and rotation**
   First and foremost, the basic mechanics consists of leading a ray of sunlight by reflecting it with mirrors to a target. Each player is assigned a mirror they can rotate. Pointing the light at the back of another player's mirror will damage it. Such simple mechanics makes the game intuitive and accessible to almost everybody, so that the team can quickly learn to play the game. This levels the field for all players and reduces bias in estimating collaboration potential. (GR 1, 4, 6)
2. **Simple one-button controller**
   We chose to use one-button controllers (see Figure 4) for each player to con-

Fig. 4: Custom one-button controllers

trol the rotation of their mirror: pressing and holding the button will rotate the mirror. This control is just enough to play the game and it eliminates distractions caused by complex controllers with many buttons. Another advantage is that by having just a button, each player is actively limited in their capabilities, which gently pushes players to collaborate to achieve their joint objectives in the game. (GR 1, 4, 6)

3. **Plant growth, healthy environment**
   The clear objective of each game level is a positive and attractive goal: to direct the sunlight ray to the plant and make it grow. This will purify the air and reduce the smog in the city. (GR 4, 6)
4. **Scoring system**
   The scoring system estimates the collaboration potential among the team members. Feedback contains several parameters on the various aspects of collaboration described in Subsection 3.6. (GR 1, 3, 2)
5. **Game level progression**
   The complexity of levels increases by the inclusions of a variety of objects, so that players face more challenging obstacles as they progress in the game. Level 1 (Figure 1) shows simple obstacles to help players familiarize themselves with the basic mechanics, controllers and objectives. Level 2 (Figure 5) features a moving obstacle that blocks the ray of light and forces a timeout on the game, while in Level 3 (Figure 6), the green mirror is placed 'inconveniently', blocking access to the plant with the risk of getting damaged. A switch workaround allows the team to change the level configuration. Level 4 (Figure 7) introduces static mirrors, which increase confusion so that the team must work together to steer the sunlight ray out of the maze of mirrors. Levels 1 to 4 address Game Requirements 1, 4, 5 and 6. Lastly, Level 5 is

a dark level: to further increase the players' dependency on each other, the scene is hidden behind thick dark smoke (Figure 8). The players must work together with their mirrors to first explore the level, learn the obstacles they face and then find a solution. Level 5 addresses Game Requirements 5 and 6.

Fig. 5: Level 2

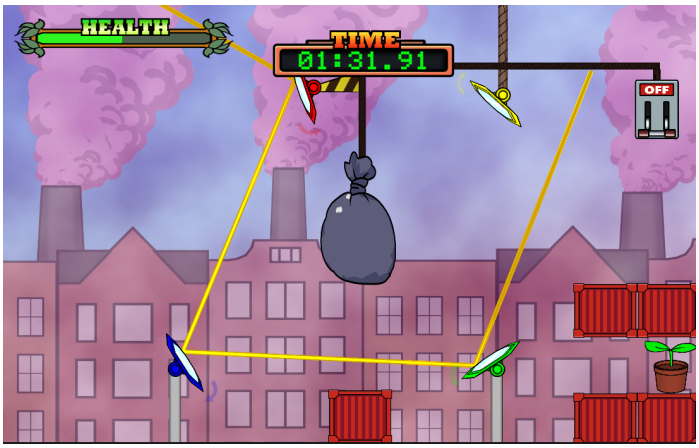


Fig. 6: Level 3

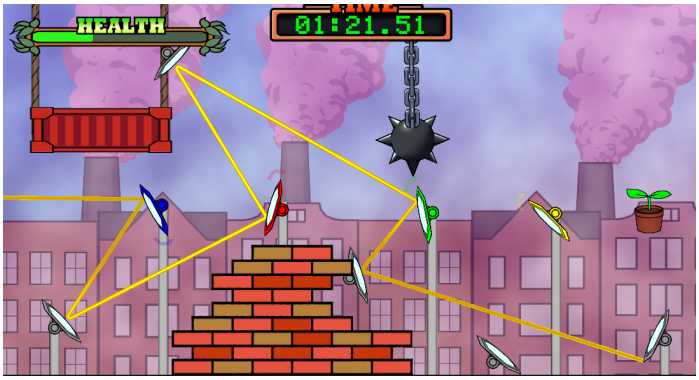


Fig. 7: Level 4

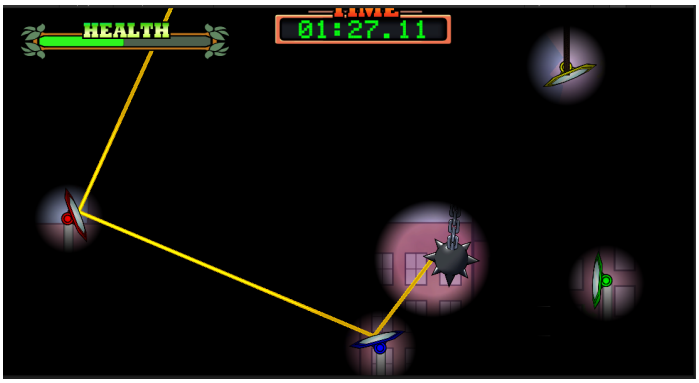


Fig. 8: Level 5

3.5 Easy and hard difficulty modes

The easy mode offers mainly simple and straightforward gameplay, ideal for players who are less familiar with video games. The hard mode is better suited for players who are more experienced with video games and are insufficiently challenged in the easy mode. The hard mode gameplay requires more collaboration, as well as quicker and more precise response by each player. Firstly, the mirrors initially assigned to the players, get re-mapped over the controllers after some random period. The players then have to find a strategy to quickly discover the new mapping and rapidly resume solving the level before more random remaps are applied. Secondly, the time available for completing each level is shorter in the hard mode. Finally, in hard mode, the mirrors' health deteriorates faster, upon an incident ray, leaving the team less room for mistakes.

3.6 Scoring system

The score of a level is the weighted sum of Time Score (S_t), Friendly Damage Score (S_f) and Retry penalty for failing a level (S_r) between 0 and 100 (best

score). Different components of the score can be given different weights and tailored to specific needs. Currently, the game uses equal weights, because it is simple and gives a broad picture of the various metrics measured. The final score of the game is the average of the scores obtained in all levels. A game is played in the same difficulty mode even though the scoring system is uniform across all levels and difficulty modes, because changing difficulty mode would introduce new variables and skew any comparisons among teams.

1. **Time Score**
   Time Score expresses how fast players can solve the levels. Levels have time limits to create a sense of urgency, and evoke in players emotions that are similar to real-life situations. Team performance in adverse conditions is measured through time spent; naturally, teams with higher collaboration potential require less time.

$$S_t = \lfloor \frac{T_l - T_u}{T_l} \times 100 \rfloor \tag{1}$$

   S_t : Calculated Score.
   T_l : Time Limit of a level.
   T_u : Time taken to complete a level.

2. **Friendly Damage Score**
   This score represents the damage a player does to team mates by pointing the light ray at the back of his colleague's mirror. A higher score can be earned by collaborating efficiently. Mayer et al. [8] mention that the number of errors inversely correlates with the efficiency of information exchange and hence collaboration potential.

$$S_f = \max(0, \lfloor \frac{H_t - H_l}{H_t} \times 100 \rfloor) \tag{2}$$

   S_f : Calculated Score.
   H_l : Damage.
   H_t : Health.

   H_t starts at 1000 in all levels and it is the collective health of the mirrors i.e. doing 250 points of damage to 4 mirrors will cause the team to fail the level. Pointing the light ray at the back of another mirror will start doing damage after 3 and 2 consecutive seconds in easy and hard modes respectively. Damage per second in easy mode is 25 and 50 for hard mode.

3. **Retry penalty for failing a Level**
   The team can fail a level by losing all health to friendly damage (Eq. 2 or by exhausting all their time; in either case, they have to retry the level). Based on the above argument, stronger collaboration potential correlates with fewer retries. When a level is retried, a penalty of 20 is applied (Eq. 3 and the score of a successful attempt is taken. Re-trying successfully solved levels to get a better score is not allowed because then the solution of the puzzle is known and finding the solution together is also a part of collaboration. The

penalty value of 20 was empirically selected and should likely be adjusted after gaining more insights into the game, in particular by benchmarking it with other objective assessments.

$$S_r = -20 \times N_r \tag{3}$$

N_r : Number of failures

4 Evaluation

In order to evaluate Mirrors in Smog City , two experiments were conducted. In the first experiment, of subjective nature, a group of participants was asked whether they felt the game had a collaboration element in it. In the second, objective experiment, the game score of two teams were compared in order to assess which of the two teams has a better collaboration potential.

4.1 Subjective experiment

20 participants (5 teams of 4 players) were introduced to the controls, objectives and backstory of the game. They were given a maximum of 10 minutes to play the game without any help or intervention from the research group and filled out a questionnaire immediately afterwards. All questions were answered on a scale of 1 (strongly disagree) to 5 (strongly agree). The survey had 3 questions (Table 1) on collaboration potential.

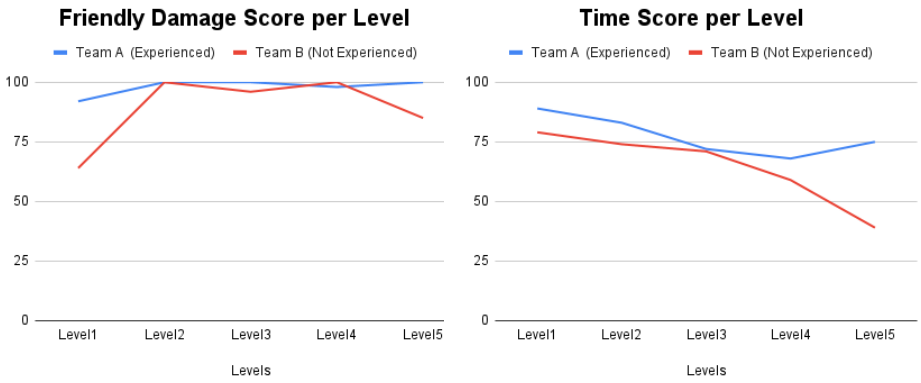
Question	**Average Response**
How much did you depend on your colleagues?	3.81
Was communication required in your team?	4.06
Was collaboration required in your team?	4.43

Table 1: Result of the survey on perception of collaboration element.

The "Dependence on colleagues"-question likely received a relatively low score because hard difficulty mode was not yet developed at the time the experiment was conducted. Also, for some participants, the first levels turned out to be easier than expected.

4.2 Objective experiment

Two teams of four participants, with no prior knowledge of Mirrors in Smog City , volunteered for this experiment. Team A has worked together on a project for 3 months and claims to have good team dynamics. On the other hand, Team B are four random participants who met each other for the first time when they started playing the game. Team A has prior experience collaborating and Team B does not. This is expected to be reflected in the objective outcome of the experiment, namely in the game scores of both teams, as defined in the previous section.



(a) Friendly Damage Score (health) (b) Time Score

Fig. 9: Scores of Team A vs Team B per level

Figure 9a shows the Friendly Damage Score for each team. Clearly Team A (average score 98) does much better than Team B (average score 89). Team B also had to retry a level because they destroyed one of their mirrors due to bad collaboration (deciding who and how to deal with the situation). This failed attempt resulted in a deduction in the final score. Note that Figure 9a shows the damage done in the following successful try. Interestingly, Team B quickly learns how to collaborate successfully, although that does not go without struggle, as seen in the drop of the score at Level 5.

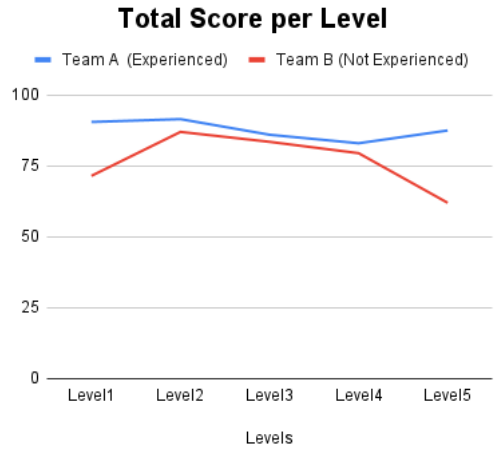


Fig. 10: Total Score per level

Team B has an average time score of 64.4 and shows a consistent drop in time taken to complete a level (Figure9b), because level complexity increases requiring stronger and more efficient collaboration among team members. Team A has an average time score of 77.4 and shows smaller drops in Time Score as the complexity of the levels increases, possibly indicating good collaboration potential. The final score is calculated as the sum of the two previous scores (Fig. 10) and as expected Team A does better on all levels. In the easiest and hardest level, Team A strongly outperforms Team B. Teams A and B got a final score of 87.7 and 76.7, respectively, which is not surprising. Of course, this experiment only gives anecdotal evidence that Mirrors in Smog City can be used to assess collaboration potential. Apart from a larger sample size, a thorough evaluation will also need to include a benchmark with an objective and, if possible, standardized method for assessing collaboration potential.

5 Conclusion and future work

Mirrors in Smog City estimates collaboration potential in a team of 4 players and also answers the Research Questions in Section 3. The process is independent of surveyor bias and individual feelings of players, because surveys are not used. Instead, results are given immediately after playing the game. The game is also accessible to most age groups by reducing dependence on complex controllers and using simple intuitive game elements. The current version of the game, however, is still just a proof of concept, and as such does not yet guarantee an accurate assessment of collaboration potential.

In Section 4.2 we described an experiment in which two teams play the game: one that has prior experience working together and another team that has never met before. The results were that the scores obtained by the first team were all higher than those of the second team. This outcome indeed hints at evidence that the game can be used to assess collaboration potential. However, we conducted the experiment for just two teams. More experiments with larger sample sizes need to be conducted to guarantee the effectiveness of the game. A drawback is that the game does not measure what happens outside of the game, e.g. the quality/quantity of verbal/non-verbal communication, nor does it track how individual players evolve through the game, because the idea is to observe team dynamics. Precision in the estimation of the degree of collaboration potential can likely be increased by analyzing communication, facial expressions, and other means to evaluate the atmosphere and interactions of the players. Naturally, a thorough evaluation will also need to include a benchmark with an objective existing method for assessing collaboration potential.

Lastly, the serious game can further be extended by adding several more levels and interactive elements. Many building blocks already exist in the game. Their use is very flexible and with a few more elements, many more unique levels can be created. Complex elements like a chaos moment where the scene is randomly rearranged (objects are moved, mirrors are rotated, switches are reset, etc) or dependent mirrors (moving a mirror will also move other mirrors) can be added. This will likely add more surprise, and allow better estimation of the collaboration potential of the team.

References

1. Alaka, S., Lopes Cunha, M., Vermeer, J., Salamon, N.Z., Balint, J.T., Bidarra, R.: Stimulating ideation in new teams with the mobile game grapplenauts. International Journal of Serious Games **6**(4), 87–101 (dec 2019), http://dx.doi.org/10.17083/ijsg.v6i4.325
2. Bedwell, W.L., Wildman, J.L., DiazGranados, D., Salazar, M., Kramer, W.S., Salas, E.: Collaboration at work: An integrative multilevel conceptualization. In: Human Resource Management Review. vol. 22, pp. 128–145 (2012)
3. Buchinger, D., da Silva Hounsell, M.: Guidelines for designing and using collaborative-competitive serious games. Computers & Education **118**, 133–149 (2018)
4. Dhillon, Raj and Nguyen, Quynh Chi: Strategies to respond to a vuca world (2021), last accessed 14 January 2023
5. Dougherty, M.B., Larson, E.: A review of instruments measuring nurse-physician collaboration. JONA: The Journal of Nursing Administration **35**(5), 244–253 (2005)
6. Driskell James E, Salas Edurado, D.T.: Foundations of teamwork and collaboration. American Psychologist **73**(4), 334–348 (2018)
7. Huang, A., Fu, Y., Leeuwen, D.v., Wu, L., Sebus, S., Song, Y., Bidarra, R.: Stranded: a classroom game for implicit bias elicitation and recognition. In: Proceedings of ISAGA 2022 - 53rd conference of the International Simulation and Gaming Association (2022)
8. Mayer, I., van Dierendonck, D., Van Ruijven, T., Wenzler, I.: Stealth assessment of teams in a digital game environment. In: Games and Learning Alliance: Second International Conference, GALA 2013, Paris, France, October 23-25, 2013, Revised Selected Papers 2. pp. 224–235. Springer (2014)
9. Okoye, P., Ezejiofor, R.A.: The effect of human resources development on organizational productivity. International Journal of Academic Research in Business and Social Sciences **3**(10), 250 (2013)
10. Schneider, B., Dowell, N., Thompson, K.: Collaboration analytics—current state and potential futures. Journal of Learning Analytics **8**(1), 1–12 (2021)
11. Sjolund, A., Straatman, M., Osch, M.v., Findra, O., Patil, P., Bueno, M., Salamon, N.Z., Bidarra, R.: Misusing mobile phones to break the ice: the tabletop game maze maestro. In: Proceedings of FDG 2020 - International Conference on the Foundations of Digital Games. ACM Press (sep 2020)
12. Tong, T., Chignell, M.: Developing a serious game for cognitive assessment: Choosing settings and measuring performance. In: Proceedings of the Second International Symposium of Chinese CHI. p. 70–79. Chinese CHI '14, Association for Computing Machinery, New York, NY, USA (2014)
13. Tong, T., Chignell, M., Tierney, M.C., Lee, J., et al.: A serious game for clinical assessment of cognitive status: validation study. JMIR serious games **4**(1), e5006 (2016)
14. Vincent Rousseau, Caroline Aubé, A.S.: Teamwork behaviors: A review and an integration of frameworks. Sage Journals **37**, 540–570 (2006)
15. Wang, C., Huang, L.: A systematic review of serious games for collaborative learning: Theoretical framework, game mechanic and efficiency assessment. International Journal of Emerging Technologies in Learning **16**(6) (2021)