Exploring a choking under pressure effect in an escape room setting

Femmie Jansen (4082680)

Utrecht University

Author Note

Femmie Jansen, Master student Social, Health and Organisational Psychology, Social Influence track, Utrecht University; dr. Hans Marien, Professor at the Department of Social and Behavioural Sciences, Utrecht University

Correspondence concerning this article should be addressed to Hans Marien, Department of Social and Behavioural Sciences, Utrecht University

Contact: h.marien@uu.nl

First supervisor: dr. H. Marien

Second supervisor: dr. R. Custers

Date: 29-06-2018

Word count: 7989

This manuscript should be made publicly accessible after 01-08-2018

Abstract

The present study examined choking under pressure in a new research setting, namely an escape room. In this setting participants were affected by four types of pressure (outcome, time, monitoring and competition). Performance in three condition (pre-, post- and in the escape room) was measured on a Mental Rotation Task (MRT). It was expected that people would perform worse on the MRT in the escape room setting, due to a high amount of pressure. The main analysis did not show a significant effect. An exploratory analysis showed that participants performed particularly low on MRT items that should have been reported as different. They did not perform above chance on these items. After correcting these items, a significant choking under pressure effect was found. The findings of this study support the hypothesis that an escape room setting can be a sufficient setting to identify choking under pressure. The results show profound implications for subsequent research. Future research is necessary to investigate the right degree of difficulty and practice items for the MRT. Furthermore, it is advised to research which kind of pressure in the escape room causes the most choking.

Keywords: Choking under pressure, escape room, mental rotation task, MRT

Exploring a choking under pressure effect in an escape room setting

It is generally assumed that increasing motivation in people leads to a better performance. However, growing literature shows that increasing motivation can also cause people to perform worse (Beilock & Carr, 2001; Hill, Hanton, Matthews & Fleming, 2010). *Performance pressure*, an anxious desire to perform as well as possible in a personally important situation, is the reason for this decrease in performance (Hardy, Mullens & Jones, 1996). This pressure to perform well, can cause people to perform worse than they intended. This is called *choking under pressure*: the occurrence of inferior performance despite individual striving and situational demands for superior performance (Baumeister, 1984).

Two theories have been put forward to explain choking under pressure, the *distraction theory* and the *self-focus theory* (Mesagno & Beckmann, 2017). Distraction theories depict that high levels of pressure distract people from their task by shifting attention to task-irrelevant cues. Pressure serves to create a dual-task environment in which task performance and for example worries about the consequences compete for attention (Beilock, Kulp, Holt & Carr, 2004). Self-focus theories explain an opposite effect. In the distraction theory the attention shifts away from the main task, whereas the self-focus theory yields that under pressure attention is focused more on the task (Beilock & Carr, 2001). Pressure, for example the presence of others, raises self-awareness in people (Baumeister, 1984). Together with the anxiety to perform correctly, this contributes to the fact that people in pressure situations start to over-monitor their task. Too much focus on well-learned, high-level sensorimotor skills can cause people to perform worse. For example, over-monitoring can decrease performance on golf putting. Too much self-focus disturbs a golfer's automatic golf swing (Beilock & Carr, 2001).

There are different kinds of pressure that can induce choking under pressure. DeCaro, Thomas, Albert & Beilock (2011) researched two types of pressure and connected them to the two "choking under pressure systems", which are described above. The first pressure is *monitoring pressure*, the pressure of being watched by others. This pressure increases self-awareness and therefore people tend to over-monitor their tasks. Based on the self-focus theory this over-monitoring can evoke choking. Especially when the performance is evaluated in some manner people tend to perform worse. The other pressure DeCaro et al. (2011) described is *outcome pressure,* a pressure caused by offering an incentive if certain outcome is achieved. Outcome pressure places the focus on worries and consequences about not receiving the incentive. Due to the distraction theory this can cause people to perform worse on a task.

Besides reward and outcome pressure, the pressure of time can also affect performance. In the meta-analysis of Byron, Khazanch and Nazarian (2010) the relationship between time pressure and creative performance has been explored. They assume that there is a U-shaped relation between time pressure and performance, with very low and very high levels of pressure being detrimental for performance. Also participants in the lab study of Freedman and Edwards (1988) performed best on anagram tasks under moderate levels of time pressure. There appears to be a trade-off between rate of performance and quality of performance. A time limit leads to higher rates, but lower quality, of task performance. So even though people can finish more tasks elements, the performance on these elements is low (Kelly & McGrath, 1985). However, time pressure may sometimes result in positive outcomes, such as increased stimulation, particularly on tasks that are experienced as boring (Van Weerde, 2003). In general, people perform worse under high levels of time pressure than in situations with moderated time pressure.

Moreover, situations in which there is a competition element the chance of choking under pressure induces. Competition causes arousal within the participants. This arousal contributes to the fact that people start focusing more on themselves (Wegner & Giuliano, 1980). Based on the self-focusing theory, this disrupts performance of well-learned sensorimotor skills. An important moderator in this relationship is self-consciousness. People who scored low in self-consciousness were the ones who are more susceptible to choking under pressure than those high in it (Baumeister, 1984).

Measuring the effect of pressure is difficult. Naturally occurring situations are often complex and thus it is hard to ascribe a performance deterioration unambiguously to choking under pressure (Lindner, 2017). Therefore, a lot choking under pressure studies use sport contests as a setting (Hill, Hanton, Matthews & Fleming, 2010). Sport events are chosen, because failure and success are easy to measure (score a point or not), the task is one-dimensional (e.g., kicking a ball) and they can be described as authentic working conditions with high stake incentives (Lindner, 2017). However, not many choking under pressure studies research outside the sport setting, what causes a one-sided view of the choking under pressure research (Beilock & Carr, 2001)

In the current research choking under pressure will be examined in a new setting, namely in an escape room. Nicholson (2015) describes the escape room as “a live action team based game where players discover clues, solve puzzles and accomplish tasks in one or more rooms in order to accomplish a specific goal (usually escaping from the room) in a limited amount of time” (p. 1). Recently, this activity has become very popular[1] around the world (Wiemker, Elumir, Clare, 2015). This specific setting has been chosen as a research setting for choking under pressure, because all kinds of pressures described above are present during an escape room game. Research of Nicholson (2015), in which results from a survey answered by 175 escape room facilities around the world were compared, shows that in every escape room there is limited time to escape the room. This time limit induces time pressure. Moreover, there is also monitoring pressure. By 95 percent of the facilities the room is monitored by someone, mostly via cameras (76% of the time) or by being present in the room (16% of the time). The most common structure in escape rooms is a structure, in which puzzles must be solved to unlock the next puzzle that eventually leads to the code for the end door. People in the escape room face outcome pressure, because they have to perform well on every single task to receive the big incentive (their freedom) in the end (Nicholson, 2015). Moreover, some escape room literally reward the ones who were able to escape, for example taking a group photo is a small common used price for winners. Lastly, in most escape rooms there is a competition element. Many escape rooms keep a ranking of the record times and put these teams or individuals on a leader board (Wiemker et al., 2015).

The fact that all these types of pressure are already present in an escape room setting, is a great advantage and increases the likelihood of finding a choking under pressure effect in an escape room setting. Moreover, it is expected that the external validity of the experiment will be better than in other experimental settings in which pressure is manipulated, because in this setting pressure is part of the game. Therefore, the manipulation will be less obvious. Furthermore, it is expected that the motivation to perform well, which is a criterion for choking under pressure (Baumeister, 1984), will be high. Escape rooms are currently very popular and visitors are in general motivated to escape on time (Nicholson, 2015; Wiemker et al., 2015). Hence, the research question for this experiment was: *Can choking under pressure be identified in an escape room setting?*

[1] The first escape room was established in the Netherlands in 2013. Within three years this grew to 133 escape room locations in the end of 2015 and it is expected that this will result in at least 400 locations in 2020 (Rietveld & Oote, 2016)

To increase the probability of finding a choking under pressure effect, a specific task was chosen. In the choking under pressure literature two kinds of tasks on which choking often occurs are mentioned. Choking under pressure caused by over-monitoring occurs mainly at tasks that require high-level sensorimotor skills (DeCaro et al. 2011), for example in sports like golf (Beilock & Car, 2001) and basketball (Gómez, Lorenzo, Jiménez, Navarro, & Sampaio, 2015). According to distraction theories, tasks that rely heavily on working memory are most negatively impacted under pressure (Gimmig, Huguet, Caverni & Cury, 2006). Working memory tasks, for example mathematical problems (Beilock et al., 2004; Beilock & Carr, 2005), require attention. However, in high pressure situations this attention shifts to task irrelevant cues, for example worries about consequences. Cognitive performance decreased, because it reduces working memory available for processing the main task (Markman, Maddox, & Worthy, 2006). Despite the fact that some escape rooms include puzzles that require sensorimotor skills, for example games that require hand eye coordination, most escape room puzzles are cognitive tasks that require working memory (Nicholson, 2015).

A task that requires high levels of attention and information from the capacity-limited working memory is the Shepard and Metzler’s (1971) Mental Rotation Task (MRT) (Pannebakker et al., 2011). This test was chosen as the performance measure in the current research, because according to the distraction theory pressure may result in an underperformance on a task like this. Moreover, performance can be easily measured by calculating the amount of correct answers given. By comparing the scores on the MRT in three conditions (pre-, post- and in the escape room) it was attempted to identify choking under pressure in an escape room setting.

Because in an escape room setting people got exposed to outcome-, monitor-, time- and competition pressure it is expected that choking under pressure will occur. Therefore, a significant quadratic effect between the three measurements is predicted. More specifically, participants will presumably perform worse on the MRT in the escape room condition in comparison to the pre- and post escape room conditions.

Method

Participants

In total 25 people participated in the experiment. To ensure that individuals were performing above chance on the MRT a minimum accuracy criterion was implemented (Beilock et al., 2004). This criterion was an accuracy on the MRT in the pre escape room condition higher than 55 percent. Due to this criterion, twelve participants were excluded from the experiment. In total the responses of thirteen participants, seven men and six women, were analysed. The age range of the participants was between 17 and 63 with a mean of 32.69 year (*SD* = 17.92). From most people the highest achieved or current educational level was HBO with 61.5 percent. Following on this were the HAVO and MBO educational level with both 15.4 percent. One participant had a WO educational level (Need & De Jong, 2001). Most individuals had some experience with playing escape rooms (*M*= 3.38, *SD* = 2.50). For two participants this escape room was their first. Five participants were not able to escape the room within the given time, the other eight people were. The record time was 9 minutes and 5 seconds and from the participants who escaped the room the mean escape time was 935,88 seconds (*SD* = 218,02). This is approximately 16 minutes. Moreover, the amount of hints given to the participants varied between zero to five. Two participants were able to escape the room without any hints, four people needed one hint, two used two hints, one person needed three hints, three other participants used four hints and one of them needed the maximum of five hints.

Materials

Questionnaire. Participants completed a small demographic questionnaire. Because choking under pressure only occurs when individuals feel motivated to perform their best (Baumeister, 1984), participants were asked to rate their motivation. Therefore, the following statement was included on the questionnaire; *Beforehand I was motivated to escape the room within the given time.* The participants could score on a 7-point Likert scale variating from 1 (*Strongly disagree)* to 7 *(Strongly agree).* Moderate task importance has been a criterion in choking under pressure studies before, so only participants who scored at the midpoint of higher were included in the analysis (Beilock & Gray, 2007). However, all participants scored a 5 or higher, so no participants were excluded because of this criterion (*M* = 6.54, *SD* = .66)

Mental rotation task (MRT). Participants were seated on a chair in front of a 15.6-inch laptop screen. On the screen a computerized version[2] of a MRT designed for the recent experiment was shown. The task consists of two-dimensional drawings of three-dimensional objects produced by Shepard and Metzler (1971). These items were used by Vandenberg and Kuse (1978) to create a paper and pencil MRT with 24 items. Every item consists of a criterion figure and a series of four alternatives. For each criterion figure, two correct

[2] This computerized experiment was administered using Inquisit 5, an experimental program developed by Sean C. Draine (1998).

alternatives and two incorrect alternatives, or distractors, were presented. This version of the MRT was used to develop a new MRT for the recent study. First of all, the 24 items were randomly assigned to the three conditions. After this, every item on the Vandenberg and Kuse (1978) test was split into four new items. Instead of showing the criterion figure and all the alternatives together, in the new test a criterion figure was shown together with one of the four alternatives. Participants were instructed to identify if the two figures, presented on the screen, were the same or different from each other. Figure 1 illustrates an example of how items were presented to the participants. Matching figures are identical to the criterion figure but are shown in a rotated position. The distractors are rotated mirror-images of the criterion figure or rotated figures of another criterion figure. In total, participants had to accomplish 32 items in every condition. Participants could choose their answer by pressing the F (figures are the same) or J (figures are different) button on the keyboard (Beilock et al., 2004). Following on each categorization selection, immediate feedback was displayed, with the words “correct” or “incorrect” appearing on the screen. At the end of the test, the amount of correct answers was presented on the screen.

Vandenberg and Kuse (1978) described that participants are in general able to complete their MRT within ten minutes. Therefore, Mental Rotation Tasks are typically administered with a time limit between 6 and 10 minutes (Voyer, 2011). A time limit of 8 minutes (i.e., five seconds per item) was chosen for the recent experiment. The time limit was visualized on the screen with a timer counting down from five to zero. If the participant answered too late, the text “Too late, please answer faster” appeared.

The pre escape room condition included an explanation part about the MRT, followed by eight practice items and 32 test items. In the escape room condition a letter was given for every fourth correct answer. With these letters the word “ZEVENTIEN” (seventeen) could be composed, which was the code for the end door. These letters where given at the end of the task, so there was no extra difficultly of remembering the letters during this condition. The better the participants performed on the MRT in the escape room, the more letters they could gather and therefore the better chance they had to escape the room. The post escape room condition included again 32 test items.

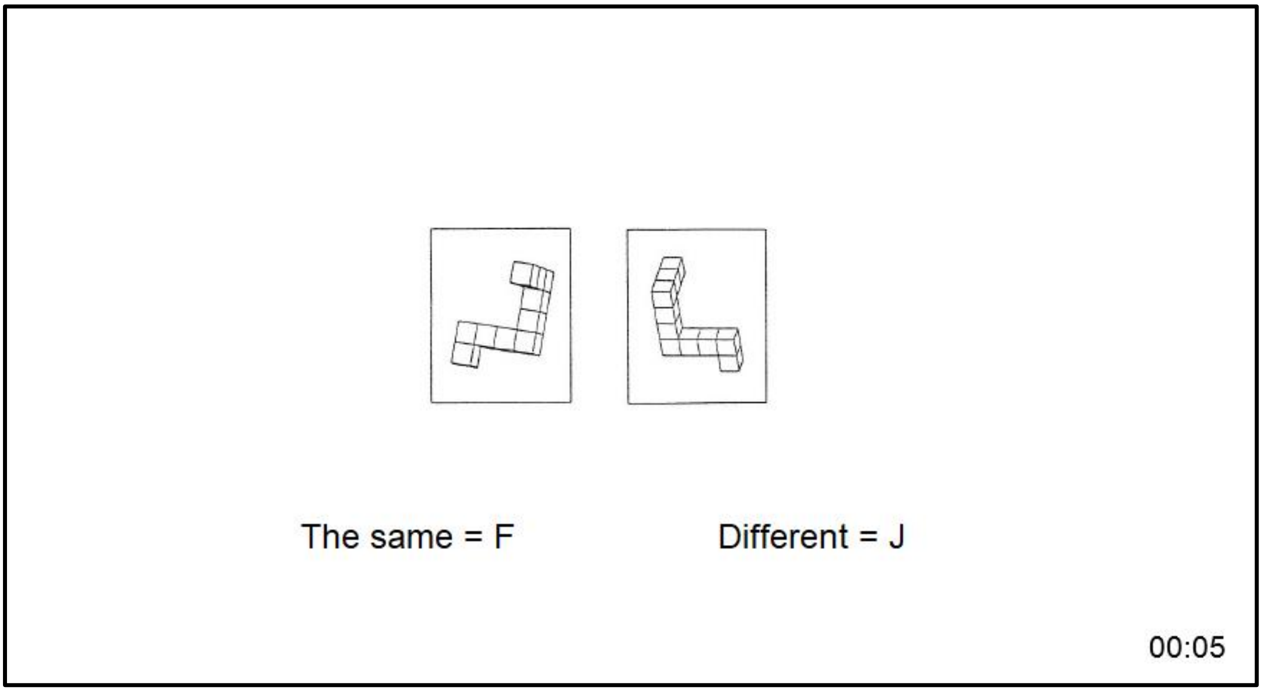

Figure 1. Screen shot of items that were presented to the participants during the MRT.

Procedure

Participants were recruited by posting a request on social media pages of the experimenter. Via snowball sampling[3] enough participants were recruited. They were welcomed by the experimenter in the house next to the escape room. In some cases participants arrived together. Then one of the participants waited in a waiting room and the other joined the researcher to a quiet room, where the first part of the experiment started. In this room the experimenter explained to the participant that to heighten the chance of escaping the room, it is important to practise his or her mental rotation skills. Therefore, a pre escape room game has been developed and the participant was asked to accomplish this before entering the escape room. The experimenter told the participant that the escape room will be prepared and the participant accomplished the first MRT. After five minutes the experimenter returned and guided the participant to the escape room. Before entering, the experimenter shortly explained the rules of the escape room. These rules were: 1. The participant has twenty minutes to escape the room. 2. He or she can do this by playing several games that will bring them steps wise to the end solution. 3. The experimenter will follow his or her performance via cameras and when necessary hints will be given. Moreover, the experimenter told the participant what the record time was at that moment. After this the participant played

[3] Snowball sampling is a social research method in which participants who are already participating in the experiment recruit other participants (Bryman, 2004).

the 20 minute version of the escape room including the second part of the MRT. After twenty minutes or if the participant escaped before, the game ended. The escape time and the amount of hints was kept track by the experimenter, who was watching the participants via cameras in the control room of the escape room. After completing the escape room, the solutions of the other games in the escape room were shortly discussed with the participant. After this, the participant and the experimenter went back to the other room and the participant was asked to complete a third MRT. Again the experimenter left the room and return after five minutes. After completing the last series of the MRT, participants filled out the demographic questionnaire and the participants were briefed about the purpose of the experiment. In the debriefing the experimenter also shortly asked the participant how they experienced the escape room and the MRT.

Design and statistical analysis

Escape room design. The escape room included a series of puzzles (see Appendix A). Every puzzle led to a new key or code to gain access to the next puzzle, with eventually the code for the end door. One of the puzzles included in this sequence was the MRT, from which the performance was measured. All participants were locked in the room on their own and they had 20 minutes to escape. The experimenter could give the participants hints if they got stuck in the game. The hint system was included in the experiment first of all because it is usually part of the escape room experience. In the research of Nicholson (2015), 82% of the researched escape rooms used a hint system. Moreover, hints were given to ensure that people started the MRT at last 5 minutes before the end of the game, so that they will be in the escape room setting when completing the MRT. The first hint was given to all participants after five minutes, after this the experimenter decided when a next hint was necessary. During the escape room, the music and light changed as well. In the beginning there was quite medieval music, because this matched the theme of the escape room that had been used in this experiment. By the time the participant started the MRT the music and lights changed. The lights were dimmed, only one spotlight directed to the end door stayed on, and the music became more thrilling. Research of Rickard (2004) has shown that exciting music causes more arousal. As described above, high levels of arousal can induce choking under pressure. Moreover, to further induce time pressure from this moment on the time left was emphasized via a microphone every minute. The time stopped when the participant escaped through the end door.

Experimental design. This experiment used a within-subject design. Participants completed the MRT in three different conditions and the results of these three tests were compared. The first condition (T1) is the baseline measurement before the participant enters the escape room. In this condition there was no competition element and no time-, monitor- or outcome pressure. Moreover, this condition was used to bring the performance of the participant on the MRT to a certain level. To ensure that participants scored above chance a criterion accuracy of 55 percent was implemented in the study (Beilock & Gray, 2007). The second condition (T2) is the escape room condition. In this condition time pressure was further emphasized by the experimenter by naming the time the participant had left every five minutes and in the end every minute. Monitoring pressure was emphasized by telling the participant that there were cameras to observe him or her while playing the escape room. Research of Belletier et al. (2015) has shown that being watched by an experimenter also induces monitoring pressure. Furthermore, outcome pressure was present because the better the participant perform on MRT the more letters he or she could collect. These letters constitute the code for the end door, so performing well on the MRT increases their chance on freedom. The competition element of the escape room was emphasized by announcing the record time to the participant before they entered the room. The last condition (T3) was the post escape room condition. This condition was the same as the pre escape room condition so no pressure- and competition elements were presented to the participants.

Data preparation and statistical analyses. The responses and reaction times (in ms) of the participants on the MRT in each condition were recorded with Inquisit 5 software. This information was exported to SPSS statistics 24. Post hoc comparisons using the repeated measures ANOVA with Bonferroni correction indicated that the amount of missing values in the pre escape room condition (M = 2.92, SD = 1.50) was significantly higher than the amount of missing values in the escape room condition (M=1.31, SD = 1.25, p = .001) and post escape room condition (M = 1.31, SD = 1.377, p = .022). Thus, in the pre escape room conditions participants answered more often too late in comparison to the other two conditions. Therefore, accuracy on the MRT has been corrected for the amount of missing values. Accuracy on the MRT was computed by dividing the amount correct answers by the amount of answers that the participant actual responded within the given 5 seconds multiplied by hundred.

The hypothesis of the experiment was tested with a repeated measures ANOVA with accuracy on the MRT as the dependent variable and condition as independent variable. Before analysing the data the assumption regarding repeated measurement ANOVA's were tested. Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated, χ^2 (2) = 1.07, *p* = .586. Histograms and Q-Q Plots of the dependent variable response accuracy in every condition have shown that the distribution was approximately normal. Moreover, boxplots did not show any warning outliers.

Results

Main analysis

A repeated measures ANOVA with accuracy on the MRT as dependent variable and condition as independent variable, did not show the expected quadratic effect, *F*(1, 12) = 2.39, *p* = 0.148. As can been seen in Figure 2 accuracy decreased in the escape room condition (*M* = 63.31, *SD* = 3.85) in comparison to the pre escape room condition (*M* = 68.06, *SD* = 2.94, *p* = .453) and the post escape room condition (*M* = 67.78, *SD* = 3.46, *p* = .637), but this effect was not significant. Moreover, there was no linear effect, *F*(1, 12) = 0.11, *p* = .918. With a mean difference of 0.27 percent (*SE* = 2.583, *p* = 1.000) the accuracies in the pre and post escape room were almost the same.

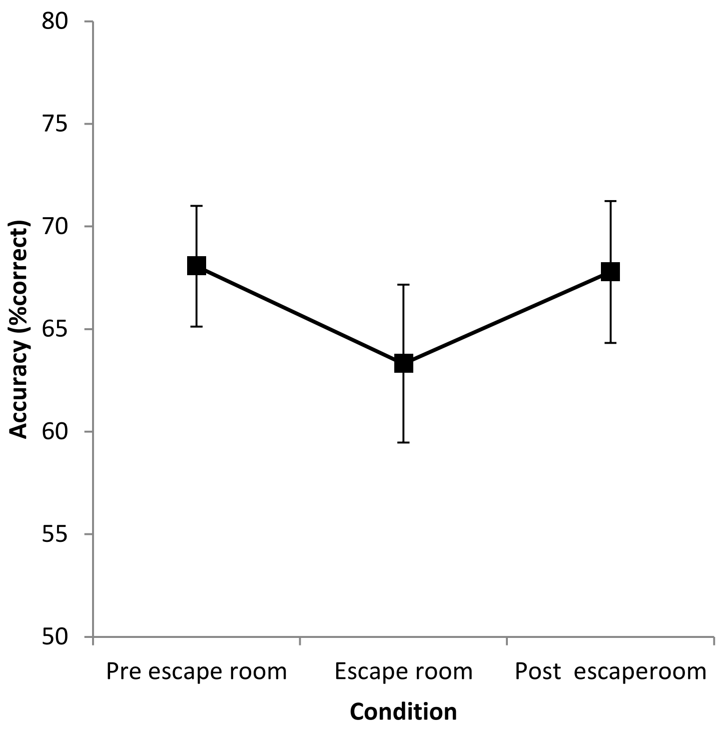

Figure 2. Mean accuracy (% correct) on the MRT (*n* = 13) in the pre escape room, escape room and post escape room condition. Error bars represent standard errors.

Analysis of the reaction time (RT) for trials on which responses were correct in the different conditions showed a significant quadratic effect, $F(2, 12) = 5.30$, $p = .040$, $\eta^2_p = .306$. The average RT in the escape room condition (M = 2850.06 ms, SD = 728.88) was significantly lower than the average RT in the pre- (M = 3213.93 ms, SD = 459.48), and post escape room condition (M = 3141.86 ms, SD = 524.77). Moreover, no linear effect was found, $F(2, 12) = .27$, $p = .615$. So the participants responded significantly faster in the escape room condition. To ensure that the differences in accuracy reported above were not caused by a trade-off with RT, the correlation between accuracy and RT in every condition was researched. As can be seen in table 1, there was no significant correlation of accuracy and RT in either the pre escape room condition ($r = -.11$, $p = .723$), the escape room condition ($r = .30$, $p = .316$) or the post escape room condition ($r = .04$, $p = .896$). The negative correlation in the pre escape room condition is in the opposite direction of what one would expect if speed-accuracy trade-off were playing a role. The fact that there is no significant correlation between accuracy and RT in all conditions, provides insights that the speed-accuracy trade-off is probably not the reason for the decrease in performance in the escape room condition.

Table 1.

Correlation between accuracy and reaction time (RT) on the MRT in the pre escape room (T1), escape room (T2) and post escape room (T3) conditions (n = 13).

	1	**2**	**3**	**4**	**5**	**6**	***M***	***SD***
1. T1Accuracy	-						68.06	10.61
2. T2Accuracy	0.62*	-					63.31	13.89
3. T3Accuracy	0.69**	0.57*	-				67.78	14.46
4. T1RT	-0.11	0.05	-0.01	-			3213.93	459.48
5. T2RT	-0.12	0.30	-0.02	0.69**	-		2850.06	728.88
6. T3RT	-0.07	0.12	0.04	0.48	0.56*	-	3141.86	524.77

*Note.** $p < 0.05$, ** $p < 0.01$

Exploratory analysis

Results from the main analysis did show a decrease in performance in the escape room condition even though this was not significant. This trend is presumably not the result of a speed-accuracy trade-off. Therefore, it was interesting to further explore this decrease in performance. Moreover, twelve out of the twenty-five participants were excluded from the main analysis because their score in the pre escape room did not meet the criterion score of 55 percent correct. So half of the participants could not perform above chance on the first MRT. In the next part the decrease in performance in the escape room condition found in the main analysis is further investigated by exploring differences between the types of MRT items.

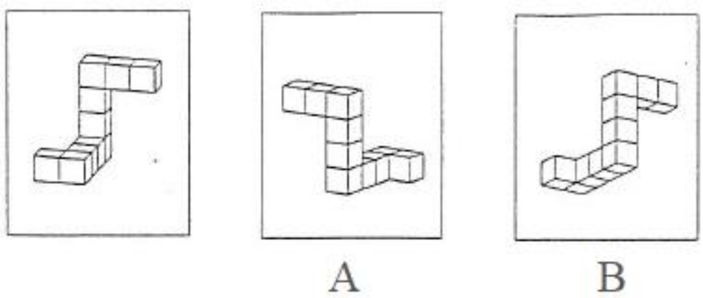

Figure 3. Example of a criterion figure with alternatives that should be reported as the same (A) or as different (B) from each other.

When looking at the MRT there are two kinds of correct answers, namely reporting items that are the same correctly and reporting items that are different from each other correctly (see Figure 3). T-tests in the different conditions showed that the performance on those types was significantly different. In the pre escape room condition there was a significant difference in accuracy to report items that are the same (*M* = 65.57, *SD* = 15.45) in comparison to reporting items as different from each other (*M* = 47.3, *SD* = 21.08); $t(24) = 4.23$, $p < .001$. Also in the escape room condition this difference between items that are the same (*M* = 64.13, *SD* = 16.72) and items that are different from each other (*M* = 52.3, *SD* = 21.78) was found; $t(24) = 2.27$, $p = .033$. Lastly, this effect was also found in the post escape room condition with a mean score of 73.45 percent (*SD* = 15.43) for the items that are the same and 55.65 percent for different items (*SD* = 18.00); $t(24) = 4.03$, $p < .001$. As can be seen in Figure 4 correctly reporting items that are different from each other showed a lower accuracy than reporting items that are the same. Moreover, in all conditions, pre escape room: $t(24) = -.64$, $p = .528$, escape room condition; $t(24) = .53$, $p = .600$ and post escape room condition; $t(24) = 1.57$, $p = .130$, the accuracy on reporting items that are different from each other did not significantly differ from chance (*M*accuracy = 50 percent). In other words, the participants did not show a higher accuracy on items that should be reported as different, than the mean accuracy they would get if they would guess all the answers. Therefore, an extra analysis was conducted in which items that should have been reported as different from each other were excluded.

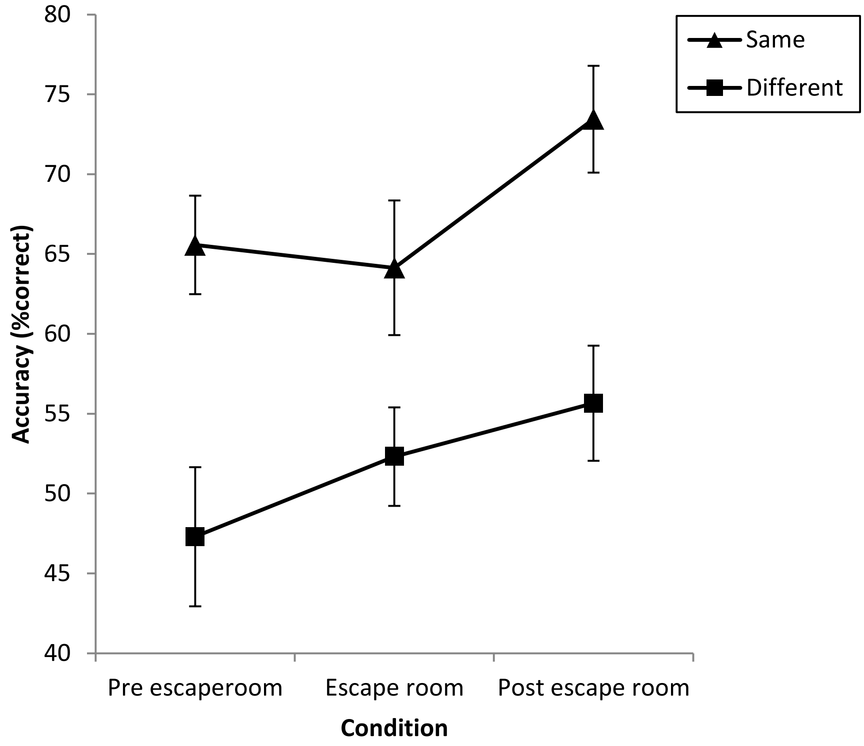

Figure 4. Mean accuracy (% correct) scored by participants ($N = 25$) on items that should be reported as the same and items that should be reported as different in the pre escape room, escape room and post escape room condition. Error bars represent standard errors.

In the extra analysis only items that should have been reported as the same were included, resulting in 16 items per condition. To control for chance the criterion accuracy of 55 percent had been used again (Beilock & Gray, 2007). This time only three participants were excluded from the experiment due to this criterion. With a mean score of 6.45 (*SD* = 1.06) on the 7-point Likert scale the participants who were included in the experiment were motivated to perform well. There was again a difference in the amount of missing values on the MRT in the different conditions. In the pre escape room condition the amount of missing values was significantly higher (*M* = 1.18, *SD* = 1.10) than in the escape room condition (*M* = .32, *SD* = .48, *p* = .002). Also in comparison to the post escape room condition, there were more missing values in the pre escape room condition, however this effect was not significant (*M* = .64, *SD* = .73 *p* = .147). Again it was chosen to correct the missing values. Accuracy was therefore computed by the amount of correct answers on items that should be reported as the same divided by amount of items of this type that have been answered within the given time multiplied by hundred. Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated, $\chi^2(2) = 2.67$, $p = .263$. Histograms and Q-Q Plots of the dependent variable response accuracy in every condition depicts that the distribution was approximately normal. Moreover, there were no warning outliers[4].

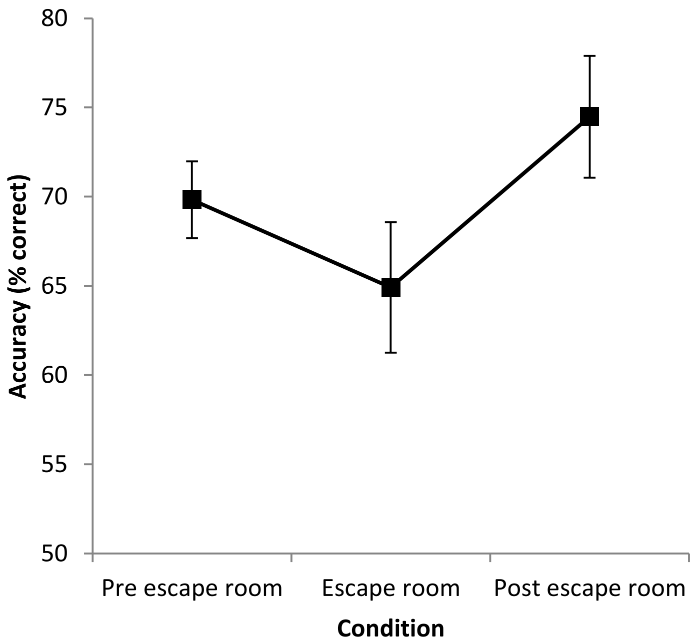

Figure 5. Mean accuracy (% correct) on items on the MRT (*n* = 22) that should be reported as the same in the pre escape room, escape room and post escape room condition. Error bars represent standard errors.

[4] The boxplots did show two outliers. These were participant 13 in T3 and participant 25 in T2. Howell (1998) depicts that outliers that differ from mean with plus or minus three standard deviations should be excluded. However, these outliers did not meet this criterion and therefore they could stay in the analysis.

A repeated measures ANOVA with the dependent variable accuracy on items that should have been reported as the same and condition as independent variable showed a significant quadratic effect, $F(1, 21) = 5.21$, $p = .033$, $\eta^2_p = .199$. It implies that after taking out the items that should have been reported as different, the participants performed significantly worse in the escape room condition as in other conditions. Moreover, there was no linear effect; $F(1, 21) = 2.89$, $p = .104$ (see Figure 5). Post- hoc comparison with Bonferroni correction showed that participants performed significantly worse in the escape room condition ($M = 64.95$, $SD = 17.15$) in comparison to the post escape room ($M = 74.49$, $SD = 16.02$, $p = .019$). However, this analysis did not show a significant difference between the escape room condition and the post escape room condition ($M = 69.83$, $SD = 10.12$, $p = .605$). So the main difference was between the accuracy in the escape room and the post escape room condition.

Analysis of the average RT to give a correct answer on items that should be reported as the same showed the same results as in the main analysis. In the escape room condition ($M = 2745.69$ ms, $SD = 771.60$) participants responded significantly faster than in the pre escape room ($M = 3146.13$ ms, $SD = 514.45$) and post escape room condition ($M = 2899.82$ ms, $SD = 622.49$); $F(1, 21) = 6.846$, $p = .016$, $\eta^2_p = .246$. Furthermore, this analysis also showed a linear trend; $F(1, 21) = 4.418$, $p = .048$, $\eta^2_p = .174$. This means that besides that the reaction time was significantly lower in the escape room condition in comparison to the other conditions, that the participants also show a linear decrease in RT. So over time the participants answered significantly faster. However, the quadratic effect was a lot stronger than this linear decrease.

Again the relationship between RT and accuracy was researched by looking at the correlation between those two variables in every condition. As can be seen in table 2, there was no significant correlation between accuracy and RT in either the pre escape room condition ($r = -.03$, $p = .913$), the escape room condition ($r = .17$, $p = .448$) or the post escape room condition ($r = -.04$, $p = .860$). The fact that there is no correlation between the variables reduces the chance that the decrease in performance during the escape room is due to a speed-accuracy trade-off. Due to the fact that the participants were in general motivated to perform well on the MRT, no clear speed-accuracy effect was found and the pressure in the escape room setting was high, the significant decrease found in this experiment can presumably be qualified as choking under pressure.

Table 2.

Correlation between accuracy and reaction time (RT) on items from the MRT that should be reported as the same in the pre escape room (T1), escape room (T2) and post escape room (T3) conditions (n = 22).

	1	**2**	**3**	**4**	**5**	**6**	***M***	***SD***
1.T1Accuracy	-						69.83	10.12
2.T2Accuracy	0.27	-					64.92	17.15
3.T3Accuracy	0.60**	0.60**	-				74.49	16.02
4.T1RT	-0.03	0.03	-0.10	-			3146.13	514.45
5.T2RT	-0.14	0.17	-0.11	0.69**	-		2745.69	771.60
6.T3RT	-0.02	-0.10	-0.04	0.55**	0.67**	-	2899.82	662.49

*Note.** $p < 0.05$, ** $p < 0.01$

Discussion

The present work was designed to explore the impact of pressure in an escape room setting that might make performance on a MRT susceptible to choking. Choking under pressure had been hard to detect outside sport settings over the recent years. Therefore, a new phenomenon in which almost all kinds of pressure are present, the escape room, was chosen for this study. In the first place, no significant choking under pressure effect was found. However, after taking out MRT items that should be reported as different the exploratory analysis did show a significant effect. This supports the hypothesis that an escape room setting can be a sufficient setting to identify choking under pressure.

A limitation of the recent study was the small amount of useful participants in the main analysis. Due to the criterion accuracy twelve out of twenty-five participants were excluded, because they did not performance above chance. This raises questions about the difficulty of the MRT, perhaps the task was too hard. A possible adjustment could be to prolong the maximum respond time. During the debriefing, eleven participants explicitly mentioned that they found the five seconds that the recent experiment uses too short. This is in line with research of Goldstein, Haldane and Mitchell (1990) which depicted that item difficultly on a MRT is primarily a product of strengthen time limits. Moreover, a stricter time limit would result in more response omissions and unanswered test items (Voyer, Rodgers & McCornick, 2004). Even though that the time limit used in the experiment was based on previous research (Voyer, 2011), it did not have the desirable result. Maybe this was due to the fact that the Vandenberg and Kuse (1978) test was a pencil and paper test and the time criterion could not be generalized to the computerised version created for the recent study.

Another possible adjustment to heighten accuracy is to include more practice items. Before the baseline measurement participants got eight practice items, however with a mean score of 4.62 (*SD* = 1.609) they did not perform that well on these relatively "easy" items. One participant even scored only one point. By offering the participants the chance to practise mental rotation more, it is expected that the amount of participants who score above the criterion will increase. Some studies do this by including the criterion value into their test. For example, DeCaro et al. (2011) included a learning criterion of eight correct trials in a row or a 200-trial maximum in their test. The computer kept presenting new practice items until the participants scored above the criterion. However, Beilock et al. (2004) explained that too much practice can decrease choking under pressure. Once problems were repeatedly practised choking under pressure was no longer observed. This is because answers were retrieved directly from long-term memory into working memory. Likewise, tasks should also not become too easy. Previous research showed that choking under pressure can only be identified on tasks that demand a high amount of working memory (Beilock et al., 2004; Beilock & Car; 2005). Future research should focus on finding the right amount of practice items and a task that has the right difficulty.

The recent study adapted to these problems by adding a second analysis. Exploring the mental rotation items showed that the participants scored especially poorly on items that should be reported as different from each other. This accuracy is maybe a result of the small time limit that, especially on this type of items, causes people to perform worse. This is in compliance with research of Vandenberg and Kuse (1978) who showed that overall mean reaction time for different pairs was nearly a second longer than the mean reaction time on items that are the same. So probably the time limit was too short for the participants to identify that two items were different from each other. Excluding this kind of items led eventually to a significant choking under pressure effect. Thus, even though that the MRT designed for this experiment was probably not the most effective task to identify choking under pressure in the first place, after correcting for different items it was a useful tool to identify choking pressure in an escape room setting.

Another point that should be noted about this study is that there was no control condition. The experiment used a within subject design including a baseline measurement and post escape room measurement. This showed a significant decrease within participant's performance in the escape room setting. For future research it is advisable to include a between-subject design with a control group for the second measurement. Participants in this group will do the exact same games as the ones that should be accomplished in the escape room. However, they will not actually enter the escape room. Instead, they stay in the room were the pre- and post escape room measurements were conducted. The game sequence, including the MRT, will be the same. However, the escape room setting, with the competition element and monitoring-, outcome- and time pressure, will be absent. By including this control group and comparing their results to the results of the group who does enter the escape room, more certainty can be given that the decrease in performance is caused by the pressure in the escape room.

In conclusion, the findings of the present study lend support to the notion that pressure in an escape room situation causes people to perform worse than they beforehand attended. Therefore, the results of this study have profound implications for subsequent choking under pressure research. The fact that this effect is found in a small sample size highlights the strength of researching choking under pressure in an escape room. Future research can build upon on this and use the escape room as a research tool. It has the same benefits as a lab study, for example high internal validity, but the manipulations will be less obvious, which improves external validity. Therefore, insights gathered in the escape room setting can bring more understanding about choking under pressure in real life settings. This can help people understand why they choke in certain situation and learn to overcome it. Moreover, it would be interesting to research what kind of pressure in an escape room causes the most choking. Research of DeCaro et al. (2011) tells us that performance of working memory tasks, like the puzzles in the escape room, got mostly disrupted by outcome pressure. So probably the concerns about the outcome, getting your freedom or not, distracts people from the puzzles which causes them to perform worse. However, this is just a speculation. During the escape room almost all kind of pressures are present. Research about the effect of all pressure types would give interesting insights about what pressure affects people's performance the most in an escape room setting.

References

Baumeister, R. F. (1984). Choking under pressure: self-consciousness and paradoxical effects of incentives on skilful performance. *Journal of personality and social psychology, 46*(3), 610-620. https://doi.org/10.1037/0022-3514.46.3.610

Beilock, S.L., & Carr, T.H. (2001). On the fragility of skilled performance: What governs choking under pressure. *Journal of Experimental Psychology*, *130*(4), 701-725. https://doi.org/10.1037/0096-3445.130.4.701

Beilock, S.L., & Carr, T.H. (2005). When high-powered people fail: Working memory and choking under pressure in math. *Psychological Science, 16*(2), 101-105. https://doi.org/10.1111/j.0956-7976.2005.00789.x

Beilock, S. L., & Gray, R. (2007). Why do athletes "choke" under pressure? In G. Tenenbaum and R.C. Ecklund (Eds.), *Handbook of sport psychology,* 3rd Ed. (pp. 425-444) . Hoboken, NH: John Wiley & Sons.

Beilock, S.L., Kulp, C.A., Holt, L.E., & Carr, T.H. (2004). More on the fragility of choking under pressure in mathematical problem solving. *Journal of Experimental Psychology*, *133*(4), 584-600. https://doi.org/10.1037/0096-3445.133.4.584

Belletier, C., Davranche, K., Tellier, I. S., Dumas, F., Vidal, F., Hasbroucq, T., & Huguet, P. (2015). Choking under monitoring pressure: being watched by the experimenter reduces executive attention. *Psychonomic bulletin & review*, *22*(5), 1410-1416. https://doi.org/10.3758/s13423-015-0804-9

Byron, K., Khazanchi, S., & Nazarian, D. (2010). The relationship between stressors and creativity: A meta-analysis examining competing theoretical models. *Journal of Applied Psychology*, *95*(1), 201-212. https://doi.org/10.1037/a0017868

Bryman, A. (2004) *Social Research Methods* (2nd edition). Oxford: Oxford University Press

DeCaro, M. S., Thomas, R. D., Albert, N. B., & Beilock, S. L. (2011). Choking under pressure: Multiple routes to skill failure. *Journal of Experimental Psychology: General*, *140*(3), 390-406. https://.doi.org/10.1037/a0023466

Draine, S. C. (1998). Inquisit (Version 5) [Computer software], Seattle,WA: Millisecond Software

Markman, A. B., Maddox, W. T., & Worthy, D. A. (2006). Choking and excelling under pressure. *Psychological Science, 17*(11), 944-948. https://doi.org/10.1111/j.1467-9280.2006.01809.x

Mesagno, C., & Beckmann, J. (2017). Choking under pressure: theoretical models and interventions. *Current opinion in psychology*, *16*, 170-175. https://doi.org/10.1016/j.copsyc.2017.05.015

Need, A., & De Jong, U. (2001). Educational differentials in the Netherlands: Testing rational action theory. *Rationality and Society, 13*(1), 71-98. https://doi.org/10.1177/104346301013001003

Nicholson, S. (2015). *Peeking behind the locked door: A survey of escape room facilities.* White Paper available at http://scottnicholson.com/pubs/erfacwhite.pdf

Freedman, J. L., & Edwards, D. R. (1988). Time pressure, task performance, and enjoyment. In J. McGrath (Ed.), *The social psychology of time: New perspectives* (Vol. 91, pp. 113–133). Newbury Park, CA: Sage.

Gimmig, D., Huguet, P., Caverni, J. P., & Cury, F. (2006). Choking under pressure and working memory capacity: When performance pressure reduces fluid intelligence. *Psychonomic Bulletin & Review, 13*(6), 1005-1010. https://doi.org/10.3758/BF03213916

Goldstein, D., Haldane, D., & Mitchell, C. (1990). Sex differences in visual-spatial ability: The role of performance factors. *Memory & Cognition, 18*(5), 546-550. https://doi.org/10.3758/BF03198487

Gómez, M. Á., Lorenzo, A., Jiménez, S., Navarro, R. M., & Sampaio, J. (2015). Examining choking in basketball: effects of game outcome and situational variables during last 5 minutes and overtimes. *Perceptual and motor skills, 120*(1), 111-124. https://doi.org/10.2466/25.29.PMS.120v11x0

Hardy, L., Mullen, R., & Jones, G. (1996). Knowledge and conscious control of motor actions under stress. *British Journal of psychology, 87*(4), 621-636. https://doi.org/10.1111/j.2044-8295.1996.tb02612.x

Hill, D.M., Hanton, S., Fleming, S., & Matthews, N. (2009). A re-examination of choking in sport. *European Journal of Sport Science*, *9*(4), 203-212. https://doi.org/10.1080/17461390902818278

Hill, D. M., Hanton, S., Matthews, N., & Fleming, S. (2010). Choking in sport: A review. *International Review of Sport and Exercise Psychology*, *3*(1), 24-39. https://doi.org/10.1080/17509840903301199

Howell, D. C. (1998). *Statistical methods in human sciences.* New York: Wadsworth.

Kelly, J. R., & McGrath, J. E. (1985). Effects of time limits and task types on task performance and interaction of four-person groups. *Journal of personality and social psychology*, *49*(2), 395- 407. https://doi.org/10.1037/0022-3514.49.2.395

Lindner, F. (2017). Choking under pressure of top performers: Evidence from Biathlon competitions. *Working Papers in Economics and Statistics, 1,* 1-27

Pannebakker, M. M., Jolicœur, P., van Dam, W. O., Band, G. P., Ridderinkhof, K. R., & Hommel, B. (2011). Mental rotation impairs attention shifting and short-term memory encoding: Neurophysiological evidence against the response-selection bottleneck model of dual-task performance. *Neuropsychologia*, *49*(11), 2985-2993. https://doi.org/10.1016/j.neuropsychologia.2011.06.021

Rickard, N. S. (2004). Intense emotional responses to music: a test of the physiological arousal hypothesis. *Psychology of music, 32*(4), 371-388. https://doi.org/10.1177/0305735604046096

Rietveld, L & Ootes, M. (2016) *De Escape Room in beeld; Can you escape in time?* Van Spronsen & Partners horeca-advies. Retrieved from https://www.spronsen.com/wordpress/wp-content/uploads/De-Escape-Room-in-beeld-2016.pdf

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*(3972), 701-703. https://doi.org/10.1126/science.171.3972.701

Vandenberg, S. .G. & Kuse, A. R. (1978) Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills,47*(2), 559-601. https://doi.org/10.2466/pms.1978.47.2.599

Van Eerde, W. (2003). A meta-analytically derived nomological network of procrastination. *Personality and individual differences, 35*(6), 1401-1418. https://doi.org/10.1016/S0191-8869(02)00358-6

Voyer, D. (2011). Time limits and gender differences on paper-and-pencil tests of mental rotation: a meta-analysis. *Psychonomic bulletin & review*, *18*(2), 267-277. https://doi.org/10.3758/s13423-010-0042-0

Voyer, D., Rodgers, M. A., & McCormick, P. A. (2004). Timing conditions and the magnitude of gender differences on the Mental Rotations Test. *Memory & Cognition*, *32*(1), 72-82. https://doi.org/10.3758/BF03195821

Wegner, D. M. & Giuliano, T. (1980) Arousal-induced attention to self. *Journal of personality and Social Psychology, 38*(5), 719-726. https://doi.org/10.1037/0022-3514.38.5.719

Wiemker, M., Elumir, E., & Clare, A. (2015, november). *Escape room games: "Can you transform an unpleasant situation into a pleasant one?"* Retrieved from https://thecodex.ca/wp-content/uploads/2016/08/00511Wiemker-et-al-Paper-Escape-Room-Games.pdf

Appendix A

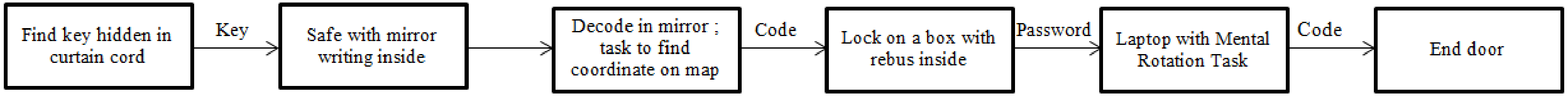

Figure A1. Sequence of all the games participants had to solve in the escape room including the MRT.