MIE1402 Group Project Report

Cynthia Li 1005934599
Kristin Feng 1004878867

1. Research Questions and Hypotheses

A cognitive experiment was conducted with undergraduate students from the Industrial Engineering Department at the University of Toronto. The experiment comprised five cognitive assessment games—Again, Only, Quick, Bigger, and Switch—each designed to assess different cognitive abilities: Working Memory, Response Inhibition, Cognitive Speed, Stroop Test, and Cognitive Flexibility, respectively. These games were performed using three distinct skins: TagMe, PlantMe, and ShuffleMe, with some games containing multiple levels of difficulty.

This paper primarily focuses on the impact of the different skins on each game. It discusses how different skins influence the task purity of each game. Task purity is defined as the extent to which a game measures a single psychological ability. Additionally, the paper examines how the correlations between games vary based on the skin types used.

Another particularly interesting research question addressed in the paper is the effect of skin variations on the grouping of games. This analysis is crucial as it reveals how different skins contribute to or alter the similarities of performances between distinct games.

Lastly, the paper proposes an objective criterion for evaluating skins, based on their observed impacts on the games. These findings offer valuable insights for the design and setup of future cognitive assessments.

In an ideal scenario, a robust cognitive assessment experiment should primarily concentrate on the tasks it measures, while minimizing the influence of extraneous design elements, such as the skins. Accordingly, this paper hypothesizes that the influence of different skins on task purity and game correlations should be consistent, suggesting that the impact of skin design on any specific game would not be significant. Additionally, different skins shouldn't influence the game clusterings too much, and games should be grouped primarily on their game types.

2. Data preparation and Preliminary analysis

In the preliminary data cleaning stage, we performed several steps. Firstly, all rows with the interaction type "Bad Hit" were removed. This interaction, representing the player making a response but hitting somewhere other than a target, was outside the research scope. Secondly, we removed rows with inconsistent reaction times.

Specifically, a row showing a reaction time of -1533 was viewed as an obvious outlier and excluded. Furthermore, instances with a reaction time of -1, indicating no response from the participant, were adjusted. For the interaction type “Miss”, we replaced the -1 reaction time with the longest mole duration times to better represent the data for further analysis.

After completing the data cleaning, we defined performance measurements for each of the five game types and calculated the scores for all 21 games for each participant. As different games had distinct potential interaction types, the score computation for each game was also different. For the “Again” and “Only” games, we excluded reaction time from the score calculation due to the lack of reaction time for the “Correct Rejection” interaction. Instead, we used D-prime (Z(Correct Hit Rate) - Z(False Alarm Rate)) to assess their performance. For the “Quick” game, which only included “Correct Hit” and “Miss” interactions, performance was calculated using a standardized correct hit rate minus standardized reaction time, as the reaction time was a crucial part of the performance measurement for the cognitive assessment games. The “Bigger” game, which only contained “Correct Hit” and “False Alarm” interactions, had its performance measures calculated by subtracting the standardized reaction time from a standardized accuracy rate. The same method was applied for the “Switch” game. The table in Appendix 1 summarizes these performance measures. All scores were standardized again after calculation for finalization.

After calculating 21 game scores for each participant, we dropped all rows of participants with NA scores in any game for consistency in data. This step left us with the final scores of 107 participants across the 21 games.

Before conducting the correlation tests, it was crucial to address the assumptions regarding outliers and normality. For each game, we identified the outliers as the score data points that deviated by more than 3 IQR from the game median and removed the entire row with outliers. This outlier cleaning process resulted in a refined dataset of 78 participants for further analysis.

The Shapiro-Wilk test was performed to test the normality of the final scores for each game. Unfortunately, the result showed that 20 out of the 21 games did not normally distribute ($p<0.05$). Despite this, the paper continued to explore the correlation between the games. However, the results of the correlation test need to be carefully interpreted due to the non-normality of the data.

3. Analysis and Results

To understand the task purity, we conducted three correlation tests between different games under each skin type. The results were composed into three correlation matrices for skin types ‘TagMe’, ‘PlantMe’, and ‘ShuffleMe’ as in Appendix 2.

Furthermore, to compare the effect of skin type on game purity directly, the mean shared variance was computed for each game. As a first step, the correlation was squared to get the shared variance. Then, we calculated the mean shared variance for a specific game by averaging the shared variance between the game versus all other same-skin games. For example, the mean shared variance for TagMeOnly was the average squared correlation of all other TagMe games except the TagMeOnly itself. The mean shared variances for each game are displayed in Appendix 3.

The mean shared variance was viewed as a measurement of task purity, as it indicated the extent to which a game measures the targeted cognitive abilities without being confounded by other abilities. A lower mean shared variance shows more pure of the task.

From Appendix 3 we could see that for the Again games, the easy level tended to be purer under TagMe skin while the medium and hard level games were purer with PlantMe. The Only games were less pure under TagMe while showing similar performance under PlantMe and ShuffleMe skins. We also noticed that the OnlyColour game performed better than OnlySuit under the ShuffleMe skin. For the Quick games, ShuffleMe had a better performance than PlantMe while PlantMe skin had better purity than Tagme. The TagMe skin had a very low mean shared variance for the Bigger game, which indicated a high task purity. Meanwhile, ShuffleMe skin had higher mean shared variances on the Bigger games, indicating it probably affected the test outcome. Last but not least, for the Switch games, ShuffleMe had higher purity than TagMe, while TagMe was better than PlantMe.

To sum up, there was no obvious conclusion that one skin affected the game's purity more than the others across all games. ShuffleMe skin performed well with the Quick and Switch games but had a higher skin influence for Bigger games. On the other hand, TagMe skin showed relatively low purity with Only and Quick games but performed well with Bigger. We could not conclude that one skin was objectively better solely based on the task purity measurements.

A loophole was identified in the previous computation: the correlation between two games that were the same type and only differed in difficulty levels could potentially be higher since they were intentionally designed to measure the same ability. To reduce the impact of the same type of games, another matrix was constructed with the mean shared variance now calculated as the average squared correlation of a game versus other same-skin different types of games. For instance, the mean shared variance of the ShuffleMeBiggerSuit game now was the averaged squared correlation against all ShuffleMe games except ShuffleMeBiggerColour and ShuffleMeBiggerSuit. A modified matrix is shown in Appendix 4.

Appendix 4 showed a significant decrease in the mean shared variances for games with the same type of different difficulties, which aligned with our speculation, that a game would have higher correlations with the same type of games. The Again games' mean shared variance dropped most significantly, as it had more same type different difficulties games than others. But overall, the sequence of skins' performances on task purity for each game remained almost the same, which showed robustness in the test results.

To investigate how the correlations between games vary depending on the skin types, we created the correlation matrices between different skins for each game (as shown in Appendix 5). Almost all correlations were positive, showing that the same type of games were in positive relations with each other. Overall, the Quick games showed the most inconsistent performances as skin type changed, while the Switch games were the most consistent ones as their correlations were relatively higher.

Focusing on the Only games, ShuffleMeOnlySuit, TagMeOnly, and PlantMeOnly had a small to medium correlation (ranging from about 0.1 to 0.2) with each other. However, ShuffleMeOnlyColour demonstrated a distinct pattern. It only shared a small to medium correlation (0.187) with ShuffleMeOnlySuit, while its correlations with TagMeOnly and PlantMeOnly were negligible, nearly zero. It suggested a lack of a significant relationship in performance with the other two skins. This might indicate that ShuffleMeOnlyColour's design was affecting the outcome.

For the Quick games, the highest correlation (0.587) existed between TagMeQuick and PlantMeQuick, indicating a strong correlation. However, the correlation coefficient between ShuffleMeQuick and TagMeQuick was only 0.04, showing no relationship between them. Overall, for the Quick games, the ShuffleMe skin type had the lowest correlation with other skin, which may imply that the skin may affect the performance of the game Quick.

Focusing on the Bigger games, most of the correlation coefficients consistently ranged from 0.2 to 0.35, showing a consistent correlation relationship between small to medium sizes. However, it was worth noticing that PlantMeBigger had a notably small correlation (lower than 0.1) with ShuffleMeBiggerSuit and ShuffleMeBiggerColour.

Finally, the Switch games had a consistency (from 0.25 to 0.4) correlation between each of them. All pairs of skins have a medium to large positive correlation with the others, demonstrating that the skin types didn't affect the performance outcome of the Switch games.

Clustering analysis was conducted to identify the effect of skin variations on the grouping of games. The hierarchical clustering results of average linkage are shown in the below dendrogram.

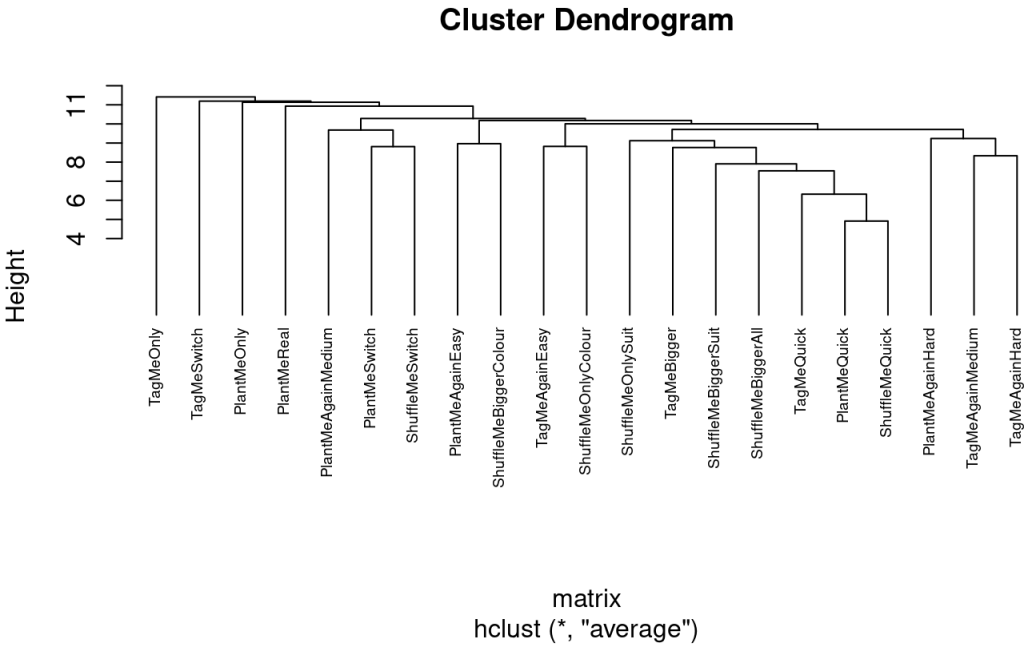


Figure 1: Cluster Dendrogram of The 21 Games

As shown in Figure 1, the 21 games were primarily grouped by their game types. For example, on the right side of the dendrogram, the Again games were close to each other. Next to them were all the Quick games. This indicated that the distinct skin types didn't influence the clustering of games, which aligned with our hypothesis, that one game's performance measures should be more similar to the same type of games instead of the same type of skins.

In addition, the team also performed a K-means clustering with K-value = 2. The results are shown in Figure 2.

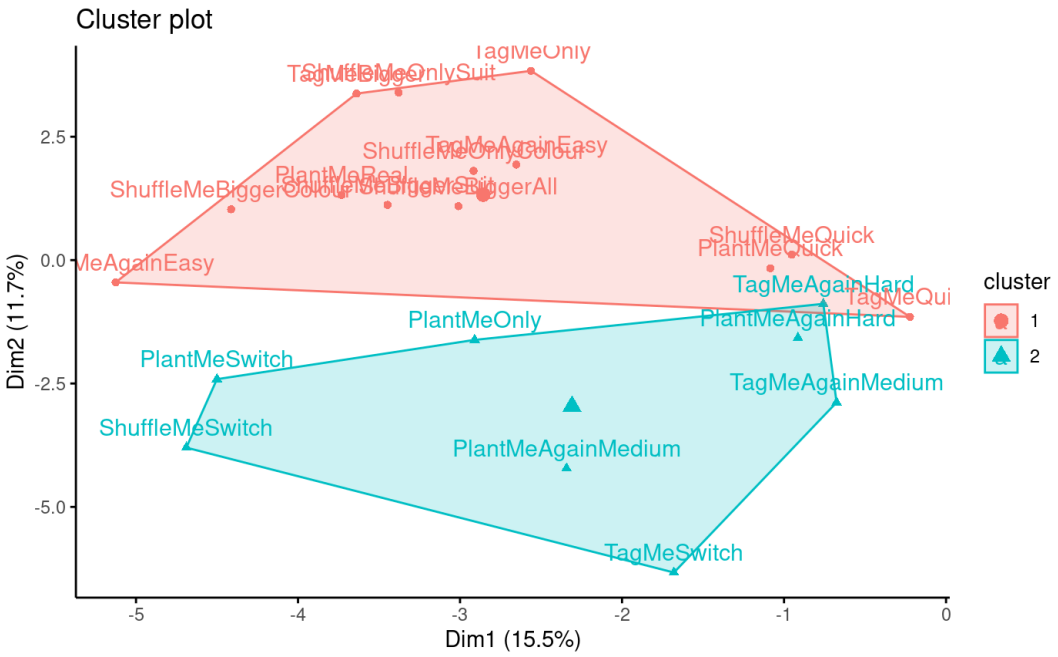


Figure 2: K-means Clustering of The 21 Games

The K-means clustering results were surprisingly interesting. One cluster of games was mostly filled up with medium to hard levels Again games and Switch games. The second cluster contained all other games. An underlying feature of the first cluster was that the harder Again games and Switch games were more cognitively challenging and required thinking rather than intuition from the participants. As opposed to them, the other games, Only, Quick, and Bigger, relied more on the instant reactions of participants. To sum up, the K-means clustering results were based on the nature of the tasks instead of the skin types of games.

4. Interpretation and Discussion

Based on the previous results discussed in Section 3, we were able to draw conclusions and interpretations. Appendixes 3 and 4 for the task purity test showed almost consistent mean shared variances between various skins across all games. All mean shared variances were between 0.01 to 0.05, which were relatively small and showed overall good performances in task purity. For Quick and Switch games, ShuffleMe performed the best. TagMe had the highest purity with Bigger and Again Easy games, while PlantMe showed higher purity with harder Again games and Only games. There was no significant pattern of which skin performed the best or worst. It could be interpreted that the design of the skins was indeed successful, as they allowed each task to focus on the cognitive ability that it was intended to measure. The result for task purity was consistent with our initial hypothesis.

The correlation between the four games varied between different skins to some extent. According to Appendix 5, the ShuffleMe skin performed not as well as the TagMe and PlantMe for games Only, Quick, and Bigger, especially when using the

ShuffleMeColour version. We interpreted the reason why ShuffleMe had lower correlations with the others was that it had distinct game designs within one task. For the Only and Bigger games, it compared both the color and suit of the card, which inevitably added to the game's complexity and differed from the design of the other two skins. It may be concluded that the ShuffleMe skin brought a distinct element to the cognitive assessment and potentially affected the consistency of results.

The clustering results showed that the skin types didn't affect the groupings of games. In the hierarchical dendrogram, the same tasks tend to be located closely to each other. In the K-means clustering graph, tasks with a similar nature (cognitively challenging vs. intuitive) were grouped. The result of the clustering tests was consistent with our initial hypothesis.

According to the results of all the tests conducted in this paper, we could not strongly conclude which skin was better than the other, although it may be suggested that ShuffleMe might not be consistent with the other two skins in terms of the measure of cognitive ability**.** Additionally, according to the result from the linear mixed model (Appendix 6), game-skin interaction didn't significantly influence participant's performances. A detailed discussion of the model is presented in Appendix 6.

In general, we believe that for any cognitive assessment, the objective metrics to asses skins are the mean shared variance for the same skin in different games and the correlation for the same game in different skins. Better skin should have less effect on the task purity (low mean shared variance with other same-skin games), and its games' performances were strongly correlated (high correlation) with other skins' same type of game. The logic behind this metric is simple: the skin types should just be a blank canvas of the cognitive assessment test, while the cognitive abilities should be the painter.

5. Appendix

1. Performance Measurement for Games Table

Game Name	Again	Only	Quick	Bigger	Switch
Performance Measurement	Z(d-prime)	Z(d-prime)	-Z(reaction time) + Z(Correct Hit Rate)	-Z(reaction time) + Z(Accuracy rate)	-Z(reaction time) + Z(Accuracy rate)

Table 1: Performance Measurement for Games Table

2. Correlation Matrices for Skins

	TagMeAgainEasy	TagMeAgainMedium	TagMeAgainHard	TagMeOnly	TagMeQuick	TagMeBigger	TagMeSwitch
TagMeAgainEasy	1.000000000	0.09659586	0.24489551	0.13178760	-0.05452456	0.12051706	-0.004638214
TagMeAgainMedium	0.096595863	1.00000000	0.28305653	-0.22429175	0.09469670	-0.03981540	0.223240406
TagMeAgainHard	0.244895515	0.28305653	1.00000000	0.01405194	-0.08369227	0.02867901	0.114473622
TagMeOnly	0.131787598	-0.22429175	0.01405194	1.00000000	-0.37187221	0.03717758	-0.243955083
TagMeQuick	-0.054524558	0.09469670	-0.08369227	-0.37187221	1.00000000	0.10093134	0.214485582
TagMeBigger	0.120517060	-0.03981540	0.02867901	0.03717758	0.10093134	1.00000000	-0.108138532
TagMeSwitch	-0.004638214	0.22324041	0.11447362	-0.24395508	0.21448558	-0.10813853	1.000000000

Figure 3: TagMe Skin Correlation Matrix

	ShuffleMeOnlyColour	ShuffleMeOnlySuit	ShuffleMeQuick	ShuffleMeBiggerSuit	ShuffleMeBiggerColour	ShuffleMeBiggerAll	ShuffleMeSwitch
ShuffleMeOnlyColour	1.000000000	0.186513347	0.078232228	0.07986482	0.16228028	0.1724415	-0.006085171
ShuffleMeOnlySuit	0.186513347	1.000000000	-0.009324602	0.28217540	0.09831312	0.2950449	-0.037969687
ShuffleMeQuick	0.078232228	-0.009324602	1.000000000	0.07367774	-0.08093724	0.1748309	0.136372859
ShuffleMeBiggerSuit	0.079864823	0.282175402	0.073677737	1.00000000	0.18841934	0.2640946	0.237395901
ShuffleMeBiggerColour	0.162280282	0.098313122	-0.080937242	0.18841934	1.00000000	0.1656501	0.105281665
ShuffleMeBiggerAll	0.172441452	0.295044899	0.174830901	0.26409458	0.16565014	1.0000000	0.224587077
ShuffleMeSwitch	-0.006085171	-0.037969687	0.136372859	0.23739590	0.10528166	0.2245871	1.000000000

Figure 4: Plant Me Skin Correlation Matrix

	PlantMeAgainEasy	PlantMeAgainMedium	PlantMeAgainHard	PlantMeOnly	PlantMeQuick	PlantMeReal	PlantMeSwitch
PlantMeAgainEasy	1.0000000000	0.18053954	0.07746541	0.0003006538	0.12964477	0.12870201	0.25758909
PlantMeAgainMedium	0.1805395375	1.00000000	0.12830767	0.0747184331	0.11436316	0.04393525	0.21564494
PlantMeAgainHard	0.0774654057	0.12830767	1.00000000	0.2036199565	-0.02149666	0.06364033	0.06173195
PlantMeOnly	0.0003006538	0.07471843	0.20361996	1.0000000000	-0.15048920	0.21494478	0.30273052
PlantMeQuick	0.1296447691	0.11436316	-0.02149666	-0.1504892040	1.00000000	0.13072134	0.24571551
PlantMeReal	0.1287020108	0.04393525	0.06364033	0.2149447849	0.13072134	1.00000000	0.19848051
PlantMeSwitch	0.2575890876	0.21564494	0.06173195	0.3027305221	0.24571551	0.19848051	1.00000000

Figure 5: ShuffleMe Skin Correlation Matrix

3. Mean Shared Variances of One Game vs. All Other Games

	TagMe	PlantMe	ShuffleMe
AgainEasy	0.017365225	0.02305327	*NA*
AgainMedium	0.033357926	0.01935870	*NA*
AgainHard	0.026870560	0.01204131	*NA*
Only	0.044509569	0.03458968	*NA*
OnlyColour	*NA*	*NA*	0.01723231
OnlySuit	*NA*	*NA*	0.03544263
Quick	0.035570822	0.02174333	0.01122498
Bigger	0.006699227	0.02087141	*NA*
BiggerSuit	*NA*	*NA*	0.04217239
BiggerColour	*NA*	*NA*	0.01942954
BiggerAll	*NA*	*NA*	0.04916311
Switch	0.030029015	0.05134702	0.02299278

Figure 6: Mean Shared Variances of One Game vs. All Other Games

4. Mean Shared Variances of One Game vs. Different Types of Games

	TagMe	PlantMe	ShuffleMe
AgainEasy	0.008721693	0.02493105	*NA*
AgainHard	0.005282137	0.01244603	*NA*
AgainMedium	0.027673950	0.01677371	*NA*
Bigger	0.006699227	0.02087141	*NA*
BiggerAll	*NA*	*NA*	0.04944819
BiggerColour	*NA*	*NA*	0.01340886
BiggerSuit	*NA*	*NA*	0.03694664
Only	0.044509569	0.03458968	*NA*
OnlyColour	*NA*	*NA*	0.01372133
OnlySuit	*NA*	*NA*	0.03557371
Quick	0.035570822	0.02174333	0.01122498
Switch	0.030029015	0.05134702	0.02299278

Figure 7: Mean Shared Variances of One Game vs. Different Types of Games

5. Correlation Matrices for games

	TagMeOnly	PlantMeOnly	ShuffleMeOnlyColour	ShuffleMeOnlySuit
TagMeOnly	1.00000000	0.1185163876	-0.0105389412	0.18521703
PlantMeOnly	0.11851639	1.0000000000	0.0008891254	0.08541153
ShuffleMeOnlyColour	-0.01053894	0.0008891254	1.0000000000	0.18651335
ShuffleMeOnlySuit	0.18521703	0.0854115289	0.1865133474	1.00000000

Figure 8: Only Games Correlation Matrix

	TagMeQuick	PlantMeQuick	ShuffleMeQuick
TagMeQuick	1.00000000	0.5870907	0.04146431
PlantMeQuick	0.58709074	1.0000000	0.15759800
ShuffleMeQuick	0.04146431	0.1575980	1.00000000

Figure 9: Quick Games Correlation Matrix

	TagMeBigger	PlantMeReal	ShuffleMeBiggerSuit	ShuffleMeBiggerColour	ShuffleMeBiggerAll
TagMeBigger	1.0000000	0.11366787	0.34734010	0.25933549	0.2986719
PlantMeReal	0.1136679	1.00000000	0.09841896	0.05741565	0.2626849
ShuffleMeBiggerSuit	0.3473401	0.09841896	1.00000000	0.18841934	0.2640946
ShuffleMeBiggerColour	0.2593355	0.05741565	0.18841934	1.00000000	0.1656501
ShuffleMeBiggerAll	0.2986719	0.26268486	0.26409458	0.16565014	1.0000000

Figure 10: Bigger Games Correlation Matrix

	TagMeSwitch	PlantMeSwitch	ShuffleMeSwitch
TagMeSwitch	1.0000000	0.2547648	0.3373532
PlantMeSwitch	0.2547648	1.0000000	0.3949742
ShuffleMeSwitch	0.3373532	0.3949742	1.0000000

Figure 11: Switch Games Correlation Matrix

6. Linear Mixed Model

To further explore the effects of different games and skins, the team also constructed a Linear Mixed Model. The PlantMe skin and the AgainEasy task were selected to be the baseline. The results are shown in Figure 12.

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: Score ~ Game * Skin + (1 | participantNumber)
   Data: df_long

REML criterion at convergence: 4143

Scaled residuals:
   Min     1Q Median     3Q    Max
-5.264 -0.554  0.042  0.670  4.517

Random effects:
 Groups            Name        Variance Std.Dev.
 participantNumber (Intercept) 0.0595   0.244
 Residual                      0.6825   0.826
Number of obs: 1638, groups:  participantNumber, 78

Fixed effects:
                            Estimate Std. Error        df t value Pr(>|t|)
(Intercept)                 8.51e-02   9.75e-02  1.43e+03    0.87    0.383
GameAgainHard              -1.33e-01   1.32e-01  1.54e+03   -1.00    0.316
GameAgainMedium            -1.04e-01   1.32e-01  1.54e+03   -0.79    0.430
GameBigger                  3.96e-02   1.32e-01  1.54e+03    0.30    0.765
GameBiggerAll               6.69e-02   1.87e-01  1.54e+03    0.36    0.721
GameBiggerColour            1.39e-01   1.87e-01  1.54e+03    0.74    0.457
GameBiggerSuit              1.48e-01   1.87e-01  1.54e+03    0.79    0.429
GameOnly                   -3.71e-02   1.32e-01  1.54e+03   -0.28    0.779
GameOnlyColour              9.30e-03   1.87e-01  1.54e+03    0.05    0.960
GameOnlySuit                7.64e-02   1.87e-01  1.54e+03    0.41    0.683
GameQuick                   2.23e-01   1.32e-01  1.54e+03    1.69    0.092 .
GameSwitch                  4.59e-02   1.32e-01  1.54e+03    0.35    0.729
SkinShuffleMe              -5.54e-02   1.32e-01  1.54e+03   -0.42    0.676
SkinTagMe                  -2.31e-02   1.32e-01  1.54e+03   -0.17    0.862
GameQuick:SkinShuffleMe     5.41e-03   1.87e-01  1.54e+03    0.03    0.977
GameAgainHard:SkinTagMe     2.99e-04   1.87e-01  1.54e+03    0.00    0.999
GameAgainMedium:SkinTagMe -1.56e-02   1.87e-01  1.54e+03   -0.08    0.934
GameBigger:SkinTagMe       -6.75e-02   1.87e-01  1.54e+03   -0.36    0.718
GameOnly:SkinTagMe          4.96e-02   1.87e-01  1.54e+03    0.26    0.791
GameQuick:SkinTagMe        -1.37e-01   1.87e-01  1.54e+03   -0.73    0.463
GameSwitch:SkinTagMe       -1.66e-01   1.87e-01  1.54e+03   -0.89    0.375
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 12: Result of Linear Mixed Model

Surprisingly, the linear mixed model results revealed that the effects of game, skin, and game-skin interactions were all not statistically significant. This result further proved that the skins did not affect the participants’ game performance. In addition, it might imply that there were other confounding factors influencing game performances, such as the individual differences of participants. However, because the data was not normally distributed and the normality assumption was not satisfied, the robustness of the test results might not be strong enough.

7. Group Presentation Slide

MIE1402 Group Project

- A Research on Cognitive Ability Test Games

Cynthia Li 1005934599
Kristin Feng 1004878867

Research Question

How do the different skins affect the issue of **task purity**?

Do the **correlations** between the four games vary between the **different skins**?

Is there any **objective way** to say that one of the **skins is better** than the others in terms of cognitive assessment?

Task Purity & Method

Task purity reflects the extent to which a task/game represents a **single psychological construct**.

E.g., to what extent is TAG-ME Only measuring other constructs **in addition to response inhibition**?

If the result correlation between game Only and the others is high, it may represent that the game Only also related to other cognitive ability. A lower correlation means more task purity.

Game Name (`name`)

	TagMe (Whacamole)	PlantMe (Gardening)	ShuffleMe (Card)	**Cognitive Abilities**
Again (n-back)	TagMeEasy (1-back) TagMeMedium (2-back) TagMeHard (3-back)	PlantMeAgainEasy (1-back) PlantMeAgainMedium (2-back) PlantMeAgainHard (3-back)	--	Working Memory
Only (Inhibition)	TagMeOnly	PlantMeOnly	ShuffleMeOnlyColour ShuffleMeOnlySuit	Response Inhibition
Quick	TagMeQuick	PlantMeQuick	ShuffleMeQuick	Cognitive Speed
Bigger	TagMeBigger	PlantMeReal	ShuffleMeBiggerSuit ShuffleMeBiggerColour ShuffleMeBiggerAll	Inhibition (Stroop Task)
Switch	TagMeSwitch	PlantMeSwitch	ShuffleMeSwitch	Cognitive Flexibility

Data Preparation - Score Calculation

Preliminary data cleaning:

- remove bad-hits
- remove all switch game with switch = 1 (doesn't reflect cognitive flexibility, just pure guesses)
- replace all reactiontime = -1 (no reaction) with the mole duration time (longest duration time)

Data Preparation - Score Calculation

Calculate the performance measurement score for all 21 types of games for each participant

Game Name	Again	Only	Quick	Bigger	Switch
Performance Measurement	-Z(reaction time) + Z(d-prime)	-Z(reaction time) + Z(d-prime)	-Z(reaction time) + Z(Correct Hit Rate)	-Z(reaction time) + Z(Accuracy rate)	-Z(reaction time) + Z(Accuracy rate)

Correct Hit Rate = total # of correct hits/ (total # of correct hits + total # of misses)

False Alarm Rate = total # of false alarms/ (total # of false alarms + total # of correct rejections)

Accuracy = (total # of correct hits + total # of correct rejections)/ total # of trials (excluding bad hits)

D-prime = Z(Correct Hit Rate) - Z(False Alarm Rate)

Game Name	Correct Hit	Miss	False Alarm	Correct Rejection	Bad Hit
Quick	X	X			X
Only	X	X	X	X	X
Again	X	X	X	X	X
Switch	X		X		
Bigger/Real	X		X		

Data Preparation - Score Calculation

Calculate the performance measurement score for all 21 types of games for each participant

Game Name	Again	Only	Quick	Bigger	Switch
Performance Measurement	-Z(reaction time) + Z(d-prime)	-Z(reaction time) + Z(d-prime)	-Z(reaction time) + Z(Correct Hit Rate)	-Z(reaction time) + Z(Accuracy rate)	-Z(reaction time) + Z(Accuracy rate)

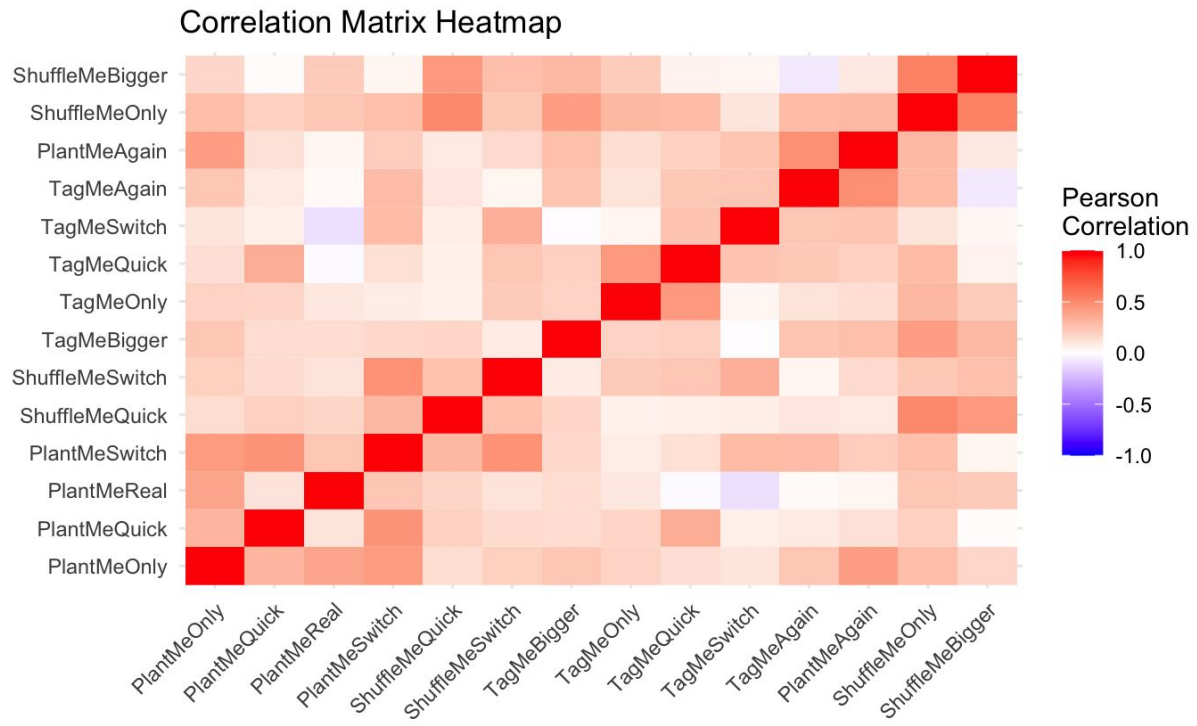
Average the scores among different difficulties for same game same skin

Final scores are **scaled** for further comparison

Remove participants with NA scores

	TagMe (Whacamole)
Again (n-back)	TagMeEasy (1-back) TagMeMedium (2-back) TagMeHard (3-back)

Correlation Between Each 14 games



Correlation between games

After scores are prepared:

1. Calculate the correlation between different games under each skin type
2. Compose correlation matrixs
3. Calculate shared variance by squaring the correlation
4. Calculate mean shared variance for each game under each skin type

	TagMeAgain_avg	TagMeOnly	TagMeQuick	TagMeBigger	TagMeSwitch
TagMeAgain_avg	1.0000000	0.11804013	0.2325204	0.243549685	0.241545780
TagMeOnly	0.1180401	1.00000000	0.4353982	0.186325756	0.034919071
TagMeQuick	0.2325204	0.43539825	1.0000000	0.196852751	0.252696558
TagMeBigger	0.2435497	0.18632576	0.1968528	1.000000000	-0.008832308
TagMeSwitch	0.2415458	0.03491907	0.2526966	-0.008832308	1.000000000

Correlation matrix for Tag-Me skin games

	tagme_mean_shared_variance
Bigger	0.03321569
Only	0.05986043
Quick	0.08656098
Switch	0.03087432
Again	0.04641500

Mean shared variance for Tag-Me skin games

Results and Interpretation on Task Purity

Higher correlations with other games mean low in purity

	TagMe	PlantMe	ShuffleMe
Again	0.04641500	0.05815934	*NA*
Only	0.05986043	0.14995499	0.1995430
Quick	0.08656098	0.08420489	0.1697910
Bigger	0.03321569	0.05567990	0.1843568
Switch	0.03087432	0.12323931	0.0635181

Mean shared variance table for same skin different games

Again	Pure across skin types
Only	Not very pure in PlantMe and ShuffleMe
Quick	Not very pure in ShuffleMe
Bigger	Not very pure in ShuffleMe
Switch	Not very pure in PlantMe

Tag-Me skin doesn't affect task purity as much as the other 2 skins. Overall, skins' affects on task purity not very significant.

Correlation between skins

Revert what was previously done to see the correlations between skins under each game

1. Calculate the correlation between different skins for each game
2. Compose correlation matrixs
3. Calculate shared variance by squaring the correlation
4. Calculate mean shared variance for each skin under each game

	TagMeOnly	PlantMeOnly	ShuffleMeOnly_avg
TagMeOnly	1.0000000	0.1813524	0.3028361
PlantMeOnly	0.1813524	1.0000000	0.2765994
ShuffleMeOnly_avg	0.3028361	0.2765994	1.0000000

Correlation matrix for Only type of games

	Game_Type	Mean_Shared_Variance
1	TagMeOnly	0.06229919
2	PlantMeOnly	0.05469797
3	ShuffleMeOnly_avg	0.08410846

Mean shared variance for Only type of games

Results and Interpretation on Correlation

	TagMe_vs_PlantMe	TagMe_vs_ShuffleMe	PlantMe_vs_ShuffleMe
Only	0.1813524	0.30283608	0.2765994
Quick	0.3474426	0.06471694	0.1989823
Bigger	0.1453943	0.29561839	0.2204752
Switch	0.2828427	0.33744044	0.4668226

Correlation Matrix for same game different skins

Game	Correlation with other skin types
Only	PlantMe < TagMe < ShuffleMe
Quick	ShuffleMe < TagMe < PlantMe
Bigger	PlantMe < TagMe < ShuffleMe
Switch	TagMe < PlantMe < ShuffleMe

Overall, no obvious patterns in the correlation rankings of skins.

Skin performance measurements

Skin performance measures:

1. Mean shared variance for same skin different games
2. Correlation for same game different skins

Better Skin:

1. Higher task purity in each game (one game only measure one specific cognitive ability)

   Low mean shared variance for same skin different games

2. Higher consistency with other skin measurement (same game measurement should not different from other skins)

   High correlation for same game different skins

Question: what else skin measurements are related to this research question?