## Data Reduction and Classification for Lumosity Data

## MSDS-6372

Yao Yao, Ian Kinskey, & Priyanka Tilak

## **Introduction**

In 2015 researchers at Lumos Labs, the Lumosity cognitive training games platform maker, sought to determine if cognitive training (via the Lumosity platform) would result in cognitive performance gains. The researchers conducted a study of 4,715 participants over the course of 10 weeks to make this determination. Participants were randomized to a treatment and active control group, and each group completed assorted cognitive performance assessments before and after the study period. The treatment group was assigned to perform cognitive training exercises (i.e. “brain training games”) on the Lumosity platform, and participants in the active control group were instructed to complete crossword puzzles. Each group was instructed to perform these activities at least five days per week for at least 15 minutes per day. The researchers found that participants in the treatment group achieved significantly better cognitive performance than those in the active control group.

## **Problem Statement**

Can the randomization grouping of participants in the original study be predicted? Utilizing cognitive ability measurements, participant activity measurements, and participants’ ages, we attempt to predict randomization grouping utilizing linear discriminant analysis and principal component analysis techniques.

## **Description of Data**

The Cognitive Ability Experiment data set consists of 5,013 observations of 80 variables where each observation corresponds to a unique study participant. 330 participants in the control group accessed the cognitive training exercises (i.e. the Lumosity brain training games) in violation of the experiment controls and are thus excluded from both the original paper and this analysis, resulting in 4,715 viable observations. The 80 variables are comprised of an identification variable (“participant\_id”), a variable indicating the exclusion status of the participant (“exclude”), a single response variable (“group”), and 77 potential explanatory variables. Appendix

The response variable, group, has two levels, “crosswords” and “Lumosity”, corresponding to the experimental group into which the participant was randomly allocated. The levels are relatively balanced, as seen in Display 1.1.

|  |
| --- |
| *Display 1.1 Frequency by Randomization Group Levels* |
|  |

The 77 candidate explanatory variables break down as shown in Display 1.

|  |  |  |  |
| --- | --- | --- | --- |
| *Display 1.2 Categorization of Explanatory Variables* | | | |
| Explanatory Variable Type | Categorical | Numeric | Total Variables |
| Cognitive Ability Measurements | 0 | 38 | 38 |
| Participant Activity Measurements | 0 | 4 | 4 |
| Participant Description | 1 | 1 | 2 |
| Participant Reported Outcomes | 0 | 33 | 33 |

The “Cognitive Ability Measurements” are the results of seven neuropsychological assessments:

|  |  |  |
| --- | --- | --- |
| *Display 1.3 Cognitive Ability Measurements* | | |
| Abbreviation | Assessment Name | Tests for |
| AR | Arithmetic Reasoning | Basic math skills (problem solving) |
| TTS | Two Target Search | Visual attention span (attention) |
| GNG | Go/No Go | Shape matching (flexibility) |
| GR | Grammatical Reasoning | Verbal syntax (language) |
| MS | Memory Span | Short term recollection (memory) |
| PM | Progressive Matrices | Deductive reasoning (problem solving) |
| RMS | Reverse Memory Span | Short term recollection (memory/attention) |

Detailed definitions of these assessments can be found in Appendix I. Participants in this data set completed each assessment before and after the study. These fourteen assessment scores were used to create an additional 31 variables. The 38 cognitive ability measurement variables are broken down in Display 1.3 below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Display 1.4 Cognitive Ability Measurement Variable Break Down* | | | | |
|  | Pre-Study | Post-Study | Grand Index | Variable Totals |
| Raw Assessment Score | 7 | 7 | 0 | 14 |
| Scaled Assessment Score | 7 | 7 | 2 | 16 |
| Variable Totals | 14 | 14 | 3 | 38 |

The pre and post study assessment scores were scaled, and a “Grand Index” variable, a composite of the seven assessment scores, was created for each set. Finally, an additional eight variables were created by calculating the difference between the pre and post study scaled assessment scores as well as the difference between the scaled pre and post study “Grand Indexes”.

The four variables categorized as “Participant Activity Measurements” are measures of individuals’ participation in the experimental environment. “crosswords\_active\_days” measures the number of unique days in which the participant was active in completing crossword puzzles. “Lumosity\_active\_days” measures the number of unique days in which the participant was active in completing cognitive training tasks. “active\_days” measures the number of unique days in which the participant was active in either completing crossword puzzles or cognitive training tasks. The “active\_days” value is either equal to the “crossword\_active\_days” value or the “Lumosity\_active\_days” value for all but those 330 participants who were excluded for reasons previously explained.

|  |  |
| --- | --- |
| *Display 1.4**Histogram of Participants, Grouped by Active Days and Age Bins* | |
|  |  |

Both “Participant Description” variables relate to the participant’s age. “age\_round” is the age of the participant at the time of the study rounded to the nearest year, and “age\_bin” is a categorical group of the participant’s age into one of 12 categories as shown in Display 1.3.

Finally, the “Participant Reported Outcomes” variables participants’ responses to a survey of ten questions related to cognitive failures and successes as well as emotional status, as well as the mean of questions one through nine. Participants completed the survey before and after the study, yielding 20 variables, and a further 11 variables were created by calculating the difference in survey responses (before less after study completion). Unlike other variables where measurements were taken under the control of the experimenters, the survey responses appear to have been optional, and thus there is a significant proportion of participants who did not respond to some or all of the survey questions. In consideration of these incomplete responses, this set of variables is excluded from this analysis.

## **Limitations and Concerns**

The participants in the study were recruited from the population of registered users of the Lumosity platform who were not paying customers. Because this sample was not randomly selected, no inferences can be made to this population. Instead, findings are limited to the study participants. Study participants who completed the study were compensated in the form of six months free use of the premium features of the Lumosity platform. In many studies such as this, some form of compensation is both common and necessary to ensure participants will adhere to and complete study requirements. However, compensation, and, in particular, this form of compensation, may have resulted in participation bias of the sample towards individuals on the Lumosity platform who are unable or unwilling to pay for its premium features.

This analysis essentially attempts to work in reverse of the Lumosity study, using features of the original data to predict a known variable—the randomized group assignment of participants. Although there may be scope limitations and biases as discussed above, the randomization should effectively eliminate confounding differences between the experimental groups.

## **Exploratory Data Analysis**

For initial analysis, we remove the “Participant Reported Outcomes” variables from consideration for the study. We then create derived variables for Cognitive Ability Measurements for each assessment as follows:

The scatterplot for the derived variables and the other explanatory variables indicate that age (age\_round) and active hours (est\_hours) may require logarithmic transformation. It is evident from the scatterplot that there is a possible correlation between active days variable and active hours.

Linear discriminant analysis and principal component analysis models assume multivariate normality of the dependent variables. As a proxy for multivariate normality, we evaluate pairwise scatterplots for bivariate normality (as characterized by an ellipsis formed point cloud) in Display 2.1 and 2.2. Not all pairwise scatterplots appear to be bivariate normal, but the very high of each group obviates this concern.

Display 2.1 and 2.2illustrate the reduction in the skew of the active days, age, and active hours variables after logarithmic transforms.

|  |  |
| --- | --- |
| *Display 2.1: Scatterplot*  *before log transformation* | *Display 2.2 Scatterplot*  *after log transformation* |
|  |  |

While skew is reduced for age, a distinct bimodal shape emerges after the transformation. Display 2.3 confirms that this bimodal shape is not a function of the randomization group.

|  |
| --- |
| *Display 2.3 Histogram of Age By Randomization Group (After Logarithmic Transform)* |
|  |

**Multicollinearity**

Multicollinearity amongst the cognitive ability measurement variables may arise due to both the nature of the assessments as well as the aptitudes of the participants. It is reasonable to distinguish an individual’s cognitive ability based on six attributes, but the seven cognitive ability assessments overlap in their measurement of these abilities. The Reverse Memory Span assessment measures both memory and attention—cognitive abilities which are also measured by the Memory Span and Two Target Search assessments. Furthermore, multicollinearity may be introduced by test techniques and their relationship to an individuals’ aptitudes. For example, several assessments use shapes as means of assessing certain abilities outside the constraints of verbal and mathematical methods. Some individuals may perform better across tests using shapes, regardless of the cognitive abilities being evaluated, due to their greater aptitude for working with shapes.

## **QDA Stepwise Elimination**

Due to multicolinearity of test scores, we must first eliminate variables that are redundant from each other to create a model.

For the QDA automatic variable selection, we chose to use the scaled pre and post results as well as active days, rounded age, and estimated hours as candidates for stepwise selection. The “delta” cognitive ability assessment variables were excluded as they are essentially redundant representations of the pre and post scaled scores.

Percent change from pre and post is also a derived result from pre and post but because it is not easily obtained through simple subtraction, we also decided to feed in percent change into another analysis in addition to the ones stated in the first QDA automatic selection. The reason that we kept in percent change in one QDA automatic selection and left it out for another was to compare whether or not the percent change would be more or less useful than pre and post scores alone. Subtraction from pre and post would result in 10 if the pre score is 80 and the post score is 90, where it would be a 12.5% increase. However, subtraction from pre and post would still result in 10 if the pre score is 10 and the post score is 20, where it would be a 100% increase. Simply put, because percent change preserves more data integrity towards the relative change in pre and post scores, we decided to run both options simultaneously for the rest of the analysis for cross comparison on which route had a better prediction estimate.

For QDA stepwise selection of scaled pre and post scores, log hours spent, active days, and log rounded age, the order in which the variables are kept to increase QDA prediction levels are as follows: Log of hours spent, active days, grand index pre score, grand index post score, grammatical reasoning post score, two target score post score, arithmetic reasoning pre score, arithmetic reasoning post score, log of age, and progressive matrices pre score (Display 3.1). Also note that memory span post score was added in step 3 of the stepwise selection and then removed on step 8. This is because the model could be predicted better with arithmetic reasoning pre score, arithmetic reasoning post score, log of age, and progressive matrices pre score than memory span post score because they are collinear of each other and the latter does a better job with the predictive modeling.

|  |
| --- |
| *Display 3.1: QDA automatic selection of pre and post results, active days, log rounded age, and log estimated hours* |
|  |

For QDA stepwise elimination of scaled pre and post scores, percent change between pre and post scores, log hours spent, active days, and log rounded age, the order in which the variables are kept to increase QDA prediction levels are as follows: grand index percent change, Log of hours spent, active days, grammatical reasoning post score, memory span post score, go/no go post score, log of age, and two target score post score (Display 3.2).

|  |
| --- |
| *Display 3.2: QDA automatic selection of pre and post results, percent change of pre and post scores, active days, log rounded age, and log estimated hours* |
|  |

In comparison with the first stepwise QDA auto selection, second QDA auto selection preferred percent change if both pre and post scores were used and also only kept post scores when necessary (Display 3.2). When percent change is introduced into the QDA auto selection, arithmetic reasoning and grand index pre and post scores are merged into their percent changes, respectively. Two target search percent change was preferred over its post score. Active days, log of age, grammatical reasoning, log of age, and log of hours are kept the same. Progressive matrices pre score was left out while Go/No-Go post score and memory span post score are newly entered into the model for the QDA auto selection with percent change.

## **QDA vs LDA**

All discriminant analysis models are quadratic rather than linear due to violations of the equality of covariance matrices assumption. Display 4 shows the results of the Bartlett’s test for each model evaluated.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Display 4: P-value significant to the chi squared F test for unequal variance* | | | | | |
| **Run** | **PCA?** | **Priors** | **Chi-Square** | **Degrees of Freedom** | **p-Value** |
| Pre and Post | No | 50-50 | 435.623072 | 55 | <.0001 |
| Pre and Post | No | 43-57 | 435.623072 | 55 | <.0001 |
| Pre, Post, and Percentage | No | 50-50 | 468.882611 | 45 | <.0001 |
| Pre, Post, and Percentage | No | 43-57 | 468.882611 | 45 | <.0001 |
| Pre and Post | Yes | 50-50 | 946.778256 | 91 | <.0001 |
| Pre and Post | Yes | 43-57 | 946.778256 | 91 | <.0001 |
| Pre, Post, and Percentage | Yes | 50-50 | 1120.607107 | 91 | <.0001 |
| Pre, Post, and Percentage | Yes | 43-57 | 1120.607107 | 91 | <.0001 |

## **QDA**

|  |  |  |
| --- | --- | --- |
| *Display 5: Cross validation summary reports and error rates of QDA categorical prediction based on auto selection with and without percent change as well as priors set at default and actual ratio* | | |
| QDA | **Auto selection without percent change** | **Auto selection with percent change** |
| **Priors set at 0.5 and 0.5** |  |  |
| **Priors set at actual group ratio** |  |  |

For the QDA model without percent change in the auto selection, the error rate was 0.3877 when the priors were set at 0.5 and 0.5 and was 0.3678 when the priors were set at the actual group ratio (Display 5). When the priors were adjusted to actual ratio, the error rate decreased and Lumosity users were predicted better than that of crossword.

For the QDA model with percent change in the auto selection, the error rate was 0.3957 when the priors were set at 0.5 and 0.5 and was 0.3779 when the priors were set at the actual group ratio (Display 5. When the priors were left unadjusted (50-50), the QDA model predicted crossword participants at a lower error rate than Lumosity participants, but after adjusting priors (57-43), the error rate of Lumosity classifications fell below that of crossword classifications.

Although the percent change variables encode more information than pre and post scores, they may risk overfitting due to also pre and post scores being candidates for stepwise variable selection.

## **PCA**

For PCA, data reduction was implemented on scaled pre and post scores, log hours spent, active days, and log rounded age with and without percent change in score.

For PCA data reduction without percent change in score, 10 principal components account for 82.9% of the variance (Display 6.1). The Eigen value scree plot starts at 5.56 and reduces to 0.69 after those 10 principal components.

|  |
| --- |
| *Display 6.1: Eigen values of the correlation matrix, scree plot, and variance explained for PCA data reduction without percent change in score* |
|  |

For principal component 1 to account for 29.1% of the variance, grand index post score and grand index pre score account for 0.390 and 0.389 of the correlation, respectively while log of age and log of hours account for 0.013 and 0.082 of the correlation, respectively (Display 6.2). There are 19 principal components created from 19 initial variables.

|  |
| --- |
| *Display 6.2: Eigen vectors of principal component correlation with initial variables* |
|  |

For PCA data reduction with percent change in score, 11 principal components account for 82.8% of the variance (Display 7.1). The eigen value scree plot starts at 5.57 and reduces to 1.20 after those 11 principal components.

|  |
| --- |
| *Display 7.1: Eigen values of the correlation matrix, scree plot, and variance explained for PCA data reduction with percent change in score* |
|  |

For principal component 1 to account for 20.5% of the variance, grand index pre score and grand index post score account for 0.392 and 0.386 of the correlation, respectively, while reverse memory span percent change and go no go percent change account for -0.024 and -0.030 of the correlation, respectively (Display 7.2). There are 27 principal components created from 27 initial variables.

|  |
| --- |
| *Display 7.2: Eigen vectors of principal component correlation with initial variables* |
|  |

|  |  |  |  |
| --- | --- | --- | --- |
| *Display 8: Comparison of principal component 1 correlation with initial variables from largest to smallest influence* | | | |
| **Auto selection without percent change** | | **Auto selection with percent change** | |
|  | **Prin1** |  | **Prin1** |
| GI\_2 | 0.390087 | GI\_1 | 0.391976 |
| GI\_1 | 0.389261 | GI\_2 | 0.38655 |
| AR\_2 | 0.30071 | AR\_1 | 0.300705 |
| AR\_1 | 0.299549 | AR\_2 | 0.299398 |
| GR\_2 | 0.257981 | GR\_2 | 0.257897 |
| GR\_1 | 0.253276 | GR\_1 | 0.253397 |
| RMS\_1 | 0.225913 | RMS\_1 | 0.229732 |
| MS\_2 | 0.220251 | MS\_2 | 0.21759 |
| MS\_1 | 0.215588 | MS\_1 | 0.217547 |
| PM\_2 | 0.212342 | PM\_2 | 0.211529 |
| RMS\_2 | 0.211194 | RMS\_2 | 0.207005 |
| PM\_1 | 0.199338 | PM\_1 | 0.19861 |
| TTS\_2 | 0.173887 | TTS\_2 | 0.173121 |
| TTS\_1 | 0.164312 | TTS\_1 | 0.164698 |
| GNG\_1 | 0.140878 | GNG\_1 | 0.144797 |
| GNG\_2 | 0.129273 | GNG\_2 | 0.12514 |
| active\_days | 0.095933 | active\_days | 0.094931 |
| log\_est\_hours | 0.082578 | log\_est\_hours | 0.081525 |
| log\_age\_round | 0.013397 | log\_age\_round | 0.012845 |
|  |  | PM\_p | 0.008614 |
|  |  | TTS\_p | 0.007067 |
|  |  | GR\_p | 0.0063 |
|  |  | AR\_p | -0.00875 |
|  |  | MS\_p | -0.00912 |
|  |  | GI\_p | -0.01774 |
|  |  | RMS\_p | -0.02446 |
|  |  | GNG\_p | -0.02963 |

Principal component 1 has the largest determining factor when explaining for variance to the initial data set. The overall ordering of the correlation of variables by influence stayed the same after percent change was added into PCA data reduction (Display 8). Percent change of progressive Matrices and two target search had slightly positive correlation while the rest of the percent change had negative correlation.

No clustering analysis is shown on the principal components because the random clouds were on top of each other similar to the initial matrix plots of the variables. Due to space and lack of results, we did not include these side by side graph comparisons.

## **QDA Stepwise Elimination after PCA**

Due to multicollinearity of test scores, we must first eliminate variables that are redundant from each other to do the model.

For the QDA automatic selection after PCA, 13 of the 19 principal components were kept in the model for categorical prediction (Display 9.1).

|  |
| --- |
| *Display 9.1: QDA automatic selection after PCA of 19 principal components without percent change* |
|  |

In layman's terms, the variables that are highly correlated to those auto-selected principal components, without percent change, from most to least common are Go/No-Go pre and post score, Two Target Search pre score, Grammatical Reasoning pre and post score, Progressive Matrices pre and post score, arithmetic reasoning pre and post score, grand index pre score, log of age, log of hours, memory span post score, active days, and reverse memory span pre and post score (Display 9.2).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Display 9.2: Top 3 correlated variables from PCA for the autoselected principal components without percent change* | | | | | | |
| **Autoselected Variable** | **Top 3 correlated variables from PCA and their correlation (without percent change)** | | | | | |
| **Prin17** | log\_est\_hours | 0.709 | AR\_1 | 0.036 | GR\_2 | 0.02 |
| **Prin16** | AR\_1 | 0.683 | GNG\_2 | 0.111 | GI\_1 | 0.086 |
| **Prin2** | log\_est\_hours | 0.638 | active\_days | 0.627 | log\_age\_round | 0.365 |
| **Prin7** | TTS\_1 | 0.549 | GI\_1 | 0.349 | PM\_1 | 0.28 |
| **Prin4** | GNG\_2 | 0.614 | GNG\_1 | 0.562 | MS\_2 | 0.122 |
| **Prin8** | log\_age\_round | 0.859 | GR\_1 | 0.215 | PM\_1 | 0.168 |
| **Prin10** | MS\_2 | 0.523 | TTS\_1 | 0.459 | GNG\_2 | 0.112 |
| **Prin15** | GR\_2 | 0.653 | GNG\_1 | 0.158 | PM\_1 | 0.157 |
| **Prin6** | TTS\_2 | 0.622 | TTS\_1 | 0.392 | PM\_2 | 0.319 |
| **Prin13** | PM\_2 | 0.508 | RMS\_1 | 0.345 | GNG\_2 | 0.315 |
| **Prin9** | GR\_2 | 0.456 | GR\_1 | 0.416 | RMS\_2 | 0.166 |
| **Prin18** | TTS\_1 | 0.16 | GNG\_1 | 0.16 | GR\_1 | 0.158 |
| **Prin3** | GNG\_1 | 0.291 | GNG\_2 | 0.275 | AR\_2 | 0.258 |

For the QDA automatic selection after PCA with percent change, 13 of the 27 principal components were kept in the model for categorical prediction (Display 10.1).

|  |
| --- |
| *Display 10.1: QDA automatic selection after PCA of 27 principal components with percent change* |
|  |

In layman's terms, the variables that are highly correlated to those auto-selected principal components, with percent change, from most to least common are grammatical reasoning pre and post score with percent change, log of age, active days, arithmetic reasoning pre and post score with percentage change, Go-No Go pre and post score with percent change, reverse memory span pre and post score with percent change, two target search pre and post score with percent change, grand index percent change, log of hours, memory span pre score and percent change, and progressive matrices pre and post score (Display 10.2).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Display 10.2: Top 3 correlated variables from PCA for the autoselected principal components with percent change* | | | | | | |
| **Autoselected Variable** | **Top 3 correlated variables from PCA and their correlation (with percent change)** | | | | | |
| **Prin17** | active\_days | 0.70364 | log\_age\_round | 0.024565 | PM\_1 | 0.010552 |
| **Prin2** | GI\_p | 0.559435 | TTS\_p | 0.289769 | RMS\_p | 0.265774 |
| **Prin10** | GNG\_2 | 0.588794 | GNG\_1 | 0.486139 | AR\_p | 0.201075 |
| **Prin11** | GNG\_p | 0.252305 | AR\_1 | 0.247253 | GNG\_2 | 0.237525 |
| **Prin3** | log\_est\_hours | 0.624369 | active\_days | 0.614062 | log\_age\_round | 0.358084 |
| **Prin25** | AR\_2 | 0.48234 | GR\_1 | 0.374701 | GR\_p | 0.343941 |
| **Prin14** | log\_age\_round | 0.85477 | GR\_2 | 0.19513 | GR\_1 | 0.142629 |
| **Prin5** | MS\_1 | 0.414791 | TTS\_p | 0.275251 | GR\_p | 0.263237 |
| **Prin24** | GR\_1 | 0.43182 | GR\_p | 0.422022 | AR\_1 | 0.357402 |
| **Prin4** | RMS\_p | 0.38556 | TTS\_1 | 0.352421 | RMS\_2 | 0.293566 |
| **Prin8** | GR\_p | 0.689824 | GR\_2 | 0.267585 | MS\_p | 0.199228 |
| **Prin15** | GR\_2 | 0.434679 | GR\_1 | 0.431531 | RMS\_1 | 0.146562 |
| **Prin26** | PM\_2 | 0.162182 | GNG\_1 | 0.160086 | TTS\_2 | 0.159347 |

In comparison with the first stepwise QDA auto selection after PCA, second QDA auto selection after PCA preferred percent change if both pre and post scores were used and sometimes preferred pre score and percent change over post scores (Display 9 and 10). When percent change is introduced into the QDA auto selection, arithmetic reasoning, Go/Not Go, grammatical reasoning, reverse memory span, and two target search expanded from pre and post score into pre and post score with percent change. Grand index percent change was preferred over its pre score and memory span pre score and percent change was preferred over its post score. Active days, log of age, log of age, log of hours, and Progressive Matrices pre and post score are kept the same.

## **QDA after PCA**

|  |  |  |
| --- | --- | --- |
| *Display 11: Cross validation summary reports and error rates of QDA after PCA categorical prediction based on auto selection with and without percent change as well as priors set at default and actual ratio* | | |
| QDA | **Auto selection without percent change** | **Auto selection with percent change** |
| **Priors set at 0.5 and 0.5** |  |  |
| **Priors set at actual group ratio** |  |  |

For the QDA model without percent change in the auto selection, the error rate was 0.4131 when the priors were set at 0.5 and 0.5 and was 0.4131 when the priors were set at the actual group ratio (Display 11). Even though error rate did not change, the distribution of correctly predicting crosswords and Lumosity is more balanced for when the priors were adjusted to actual ratio.

For the QDA model with percent change in the auto selection, the error rate was 0.4134 when the priors were set at 0.5 and 0.5 and was 0.4271 when the priors were set at the actual group ratio (Display 11). When the priors were adjusted to actual ratio, the error rate increased and the prediction went from overwhelmingly predicting crossword users better to still predicting crossword users better.

Even though percent change more data integrity towards the relative change in pre and post scores, it might be over fitting because the pre and post scores are also in the model for auto selection to eliminate. It was surprising that adjusting the priors to the actual group ratio for PCA with percent change had a relatively higher error rate.

## **Conclusion**

For auto selection with percent change without PCA, the stepwise QDA elimination preferred percent change over pre and post scores. For auto selection with percent change with PCA, the stepwise QDA elimination preferred percent change in addition to pre and post scores.

For both LDA selection with and without PCA, grammatical reasoning, arithmetic reasoning, grand index, two target search, active days, log of hours, and log of age were all kept as variables indicating the QDA prediction of crossword and Lumosity groups.

We found out that including percent change in addition to pre and post scores is considered over fitting because the error rate increased for both QDA with and without PCA. This is also seen for principal component 1's correlation where percent change had the lowest correlation of all the variables when determining the largest variance for data reduction.

Our highest QDA categorical prediction accuracy rate without PCA is 0.6322 while our highest QDA categorical prediction error rate with PCA is 0.5869. Adjusting for priors based on actual ratio of groups either lowers error rate or balances the prediction accuracy for both groups. We cannot adequately predict the different groups from the dataset. The best performing model yielded a 63% accuracy rate, little better than the 50% that would be expected by randomly predicting the randomization group.

## **Appendix I – Data Dictionary**

|  |  |  |
| --- | --- | --- |
| **Column name** | **Variable Type** | **Description** |
| participant\_id | Identification | Unique identifier for each participant |
| group | Response | Randomized condition - either "Lumosity" (cognitive training) or "crosswords" (crossword puzzle control) |
| AR\_pre | Cognitive Ability Measurements | Neurocognitive battery pre-test - raw Arithmetic Reasoning score (number of correct responses) |
| TTS\_pre | Cognitive Ability Measurements | Neurocognitive battery pre-test - raw Two-Target Search score (negative threshold presentation time) |
| GNG\_pre | Cognitive Ability Measurements | Neurocognitive battery pre-test - raw Go/No-Go score (negative time) |
| GR\_pre | Cognitive Ability Measurements | Neurocognitive battery pre-test - raw Grammatical Reasoning score (net number of correct responses) |
| MS\_pre | Cognitive Ability Measurements | Neurocognitive battery pre-test - raw Forward Memory Span score (span) |
| PM\_pre | Cognitive Ability Measurements | Neurocognitive battery pre-test - raw Progressive Matrices score (number correct) |
| RMS\_pre | Cognitive Ability Measurements | Neurocognitive battery pre-test - raw Reverse Memory Span score (span) |
| age\_round | Participant Description | Integer age at time of pre-test (floor of absolute age) |
| age\_bin | Participant Description | 5-year age bin |
| AR\_post | Cognitive Ability Measurements | Neurocognitive battery post-test - raw Arithmetic Reasoning score (number of correct responses) |
| TTS\_post | Cognitive Ability Measurements | Neurocognitive battery post-test - raw Two-Target Search score (negative threshold presentation time) |
| GNG\_post | Cognitive Ability Measurements | Neurocognitive battery post-test - raw Go/No-Go score (negative time) |
| GR\_post | Cognitive Ability Measurements | Neurocognitive battery post-test - raw Grammatical Reasoning score (net number of correct responses) |
| MS\_post | Cognitive Ability Measurements | Neurocognitive battery post-test - raw Forward Memory Span score (span) |
| PM\_post | Cognitive Ability Measurements | Neurocognitive battery post-test - raw Progressive Matrices score (number correct) |
| RMS\_post | Cognitive Ability Measurements | Neurocognitive battery post-test - raw Reverse Memory Span score (span) |
| AR\_1 | Cognitive Ability Measurements | Neurocognitive battery pre-test - scaled Arithmetic Reasoning score |
| TTS\_1 | Cognitive Ability Measurements | Neurocognitive battery pre-test - scaled Two-Target Search score |
| GNG\_1 | Cognitive Ability Measurements | Neurocognitive battery pre-test - scaled Go/No-Go score |
| GR\_1 | Cognitive Ability Measurements | Neurocognitive battery pre-test - scaled Grammatical Reasoning score |
| MS\_1 | Cognitive Ability Measurements | Neurocognitive battery pre-test - scaled Forward Memory Span score |
| PM\_1 | Cognitive Ability Measurements | Neurocognitive battery pre-test - scaled Progressive Matrices score |
| RMS\_1 | Cognitive Ability Measurements | Neurocognitive battery pre-test - scaled Reverse Memory Span score |
| GI\_1 | Cognitive Ability Measurements | Neurocognitive battery pre-test - scaled Grand Index score |
| AR\_2 | Cognitive Ability Measurements | Neurocognitive battery post-test - scaled Arithmetic Reasoning score |
| TTS\_2 | Cognitive Ability Measurements | Neurocognitive battery post-test - scaled Two-Target Search score |
| GNG\_2 | Cognitive Ability Measurements | Neurocognitive battery post-test - scaled Go/No-Go score |
| GR\_2 | Cognitive Ability Measurements | Neurocognitive battery post-test - scaled Grammatical Reasoning score |
| MS\_2 | Cognitive Ability Measurements | Neurocognitive battery post-test - scaled Forward Memory Span score |
| PM\_2 | Cognitive Ability Measurements | Neurocognitive battery post-test - scaled Progressive Matrices score |
| RMS\_2 | Cognitive Ability Measurements | Neurocognitive battery post-test - scaled Reverse Memory Span score |
| GI\_2 | Cognitive Ability Measurements | Neurocognitive battery post-test - scaled Grand Index score |
| AR\_d | Cognitive Ability Measurements | Neurocognitive battery - change in scaled Arithmetic Reasoning score |
| TTS\_d | Cognitive Ability Measurements | Neurocognitive battery - change in scaled Two-Target Search score |
| GNG\_d | Cognitive Ability Measurements | Neurocognitive battery - change in scaled Go/No-Go score |
| GR\_d | Cognitive Ability Measurements | Neurocognitive battery - change in scaled Grammatical Reasoning score |
| MS\_d | Cognitive Ability Measurements | Neurocognitive battery - change in scaled Forward Memory Span score |
| PM\_d | Cognitive Ability Measurements | Neurocognitive battery - change in scaled Progressive Matrices score |
| RMS\_d | Cognitive Ability Measurements | Neurocognitive battery - change in scaled Reverse Memory Span score |
| GI\_d | Cognitive Ability Measurements | Neurocognitive battery - change in Grand Index score |
| lost\_track\_details\_pre | Participant Reported Outcomes | Participant reported outcomes pre-test - Q1 |
| misplaced\_items\_pre | Participant Reported Outcomes | Participant reported outcomes pre-test - Q2 |
| lost\_concentration\_pre | Participant Reported Outcomes | Participant reported outcomes pre-test - Q3 |
| remembered\_names\_pre | Participant Reported Outcomes | Participant reported outcomes pre-test - Q4 |
| felt\_creative\_pre | Participant Reported Outcomes | Participant reported outcomes pre-test - Q5 |
| good\_concentration\_pre | Participant Reported Outcomes | Participant reported outcomes pre-test - Q6 |
| felt\_anxious\_pre | Participant Reported Outcomes | Participant reported outcomes pre-test - Q7 |
| in\_bad\_mood\_pre | Participant Reported Outcomes | Participant reported outcomes pre-test - Q8 |
| felt\_sad\_pre | Participant Reported Outcomes | Participant reported outcomes pre-test - Q9 |
| felt\_training\_benefits\_pre | Participant Reported Outcomes | Participant reported outcomes pre-test - Q10 (excluded from average, see manuscript) |
| rwc\_ave\_pre | Participant Reported Outcomes | Participant reported outcomes pre-test - average of Q1-9 |
| lost\_track\_details\_post | Participant Reported Outcomes | Participant reported outcomes post-test - Q1 |
| misplaced\_items\_post | Participant Reported Outcomes | Participant reported outcomes post-test - Q2 |
| lost\_concentration\_post | Participant Reported Outcomes | Participant reported outcomes post-test - Q3 |
| remembered\_names\_post | Participant Reported Outcomes | Participant reported outcomes post-test - Q4 |
| felt\_creative\_post | Participant Reported Outcomes | Participant reported outcomes post-test - Q5 |
| good\_concentration\_post | Participant Reported Outcomes | Participant reported outcomes post-test - Q6 |
| felt\_anxious\_post | Participant Reported Outcomes | Participant reported outcomes post-test - Q7 |
| in\_bad\_mood\_post | Participant Reported Outcomes | Participant reported outcomes post-test - Q8 |
| felt\_sad\_post | Participant Reported Outcomes | Participant reported outcomes post-test - Q9 |
| felt\_training\_benefits\_post | Participant Reported Outcomes | Participant reported outcomes post-test - Q10 (excluded from average, see manuscript) |
| rwc\_ave\_post | Participant Reported Outcomes | Participant reported outcomes post-test - average of Q1-9 |
| lost\_track\_details\_d | Participant Reported Outcomes | Participant reported outcomes - change in Q1 |
| misplaced\_items\_d | Participant Reported Outcomes | Participant reported outcomes - change in Q2 |
| lost\_concentration\_d | Participant Reported Outcomes | Participant reported outcomes - change in Q3 |
| remembered\_names\_d | Participant Reported Outcomes | Participant reported outcomes - change in Q4 |
| felt\_creative\_d | Participant Reported Outcomes | Participant reported outcomes - change in Q5 |
| good\_concentration\_d | Participant Reported Outcomes | Participant reported outcomes - change in Q6 |
| felt\_anxious\_d | Participant Reported Outcomes | Participant reported outcomes - change in Q7 |
| in\_bad\_mood\_d | Participant Reported Outcomes | Participant reported outcomes - change in Q8 |
| felt\_sad\_d | Participant Reported Outcomes | Participant reported outcomes - change in Q9 |
| felt\_training\_benefits\_d | Participant Reported Outcomes | Participant reported outcomes - change in Q10 (excluded from average, see manuscript) |
| rwc\_ave\_d | Participant Reported Outcomes | Participant reported outcomes - change in average of Q1-9 |
| crosswords\_active\_days | Participant Activity Measurements | Number of unique days participant started a crossword puzzle during study period |
| Lumosity\_active\_days | Participant Activity Measurements | Number of unique days participant completed at least one Lumosity game during study period |
| active\_days | Participant Activity Measurements | Total number of unique days participant engaged with either program |
| exclude | Experiment Meta Variable | Was control participant excluded from final analysis for completing cognitive training during study period? (1 = excluded, 0 = included) |
| est\_hours | Participant Activity Measurements | Estimate of total hours trained during the study period (see Supplementary Information for more detail on how this was calculated) |

## **Appendix II – SAS Code**

/\* IMPORT DATA \*/

FILENAME REFFILE '/home/yaoy890/proj 2/S1\_Dataset.csv';

PROC IMPORT DATAFILE=REFFILE

DBMS=CSV

OUT=Lumositydata;

GETNAMES=YES;

RUN;

/\*

DATA CLEAN UP

- remove all survey variables

- remove redundant variables (crosswords\_active\_days, Lumosity\_active\_days, age\_bin)

- remove extraneous "exclude" variable

- natural log transform estimated hours

- exclude observations where "exclude" = 1

- calculate "P" variables (percentage change of scaled pre/post assessment scores)

\*/

DATA LumosityCleaned;

SET Lumositydata;

WHERE exclude=0; /\*exclude participants which are marked as exclude=1\*/

DROP age\_bin lost\_track\_details\_pre misplaced\_items\_pre lost\_concentration\_pre remembered\_names\_pre felt\_creative\_pre good\_concentration\_pre felt\_anxious\_pre in\_bad\_mood\_pre felt\_sad\_pre felt\_training\_benefits\_pre rwc\_ave\_pre lost\_track\_details\_post misplaced\_items\_post lost\_concentration\_post remembered\_names\_post felt\_creative\_post good\_concentration\_post felt\_anxious\_post in\_bad\_mood\_post felt\_sad\_post felt\_training\_benefits\_post rwc\_ave\_post lost\_track\_details\_d misplaced\_items\_d lost\_concentration\_d remembered\_names\_d felt\_creative\_d good\_concentration\_d felt\_anxious\_d in\_bad\_mood\_d felt\_sad\_d felt\_training\_benefits\_d rwc\_ave\_d crosswords\_active\_days Lumosity\_active\_days exclude AR\_pre TTS\_pre GNG\_pre GR\_pre MS\_pre PM\_pre RMS\_pre AR\_pre AR\_post TTS\_post GNG\_post GR\_post MS\_post PM\_post RMS\_post AR\_d TTS\_d GNG\_d GR\_d MS\_d PM\_d RMS\_d GI\_d; \*remove clutter and columns not used;

log\_est\_hours=log(est\_hours+1); \*due to skew in histogram;

log\_age\_round=log(age\_round+1); \*due to skew in histogram;

AR\_p = (AR\_2 - AR\_1) / AR\_1 \* 100;

TTS\_p = (TTS\_2 - TTS\_1) / TTS\_1 \* 100;

GNG\_p = (GNG\_2 - GNG\_1) / GNG\_1 \* 100;

GR\_p = (GR\_2 - GR\_1) / GR\_1 \* 100;

MS\_p = (MS\_2 - MS\_1) / MS\_1 \* 100;

PM\_p = (PM\_2 - PM\_1) / PM\_1 \* 100;

RMS\_p = (RMS\_2 - RMS\_1) / RMS\_1 \* 100;

GI\_p = (GI\_2 - GI\_1) / GI\_1 \* 100;

RUN;

/\* PRINT SAMPLE \*/

proc print data=LumosityCleaned(obs=50);

run;

quit;

proc sgscatter data=LumosityCleaned;

matrix age\_round AR\_1 TTS\_1 GNG\_1 GR\_1 MS\_1 PM\_1 RMS\_1 GI\_1 active\_days est\_hours / diagonal=(histogram) group = group;

run;

proc sgscatter data=LumosityCleaned;

matrix age\_round AR\_2 TTS\_2 GNG\_2 GR\_2 MS\_2 PM\_2 RMS\_2 GI\_2 active\_days est\_hours / diagonal=(histogram) group = group;

run;

proc sgscatter data=LumosityCleaned;

matrix age\_round AR\_p TTS\_p GNG\_p GR\_p MS\_p PM\_p RMS\_p GI\_p active\_days est\_hours / diagonal=(histogram) group = group;

run;

\*log transform age and hours while others have normal distribution;

proc sgscatter data=LumosityCleaned;

matrix log\_age\_round AR\_1 TTS\_1 GNG\_1 GR\_1 MS\_1 PM\_1 RMS\_1 GI\_1 active\_days log\_est\_hours / diagonal=(histogram) group = group;

run;

proc sgscatter data=LumosityCleaned;

matrix log\_age\_round AR\_2 TTS\_2 GNG\_2 GR\_2 MS\_2 PM\_2 RMS\_2 GI\_2 active\_days log\_est\_hours / diagonal=(histogram) group = group;

run;

proc sgscatter data=LumosityCleaned;

matrix log\_age\_round AR\_p TTS\_p GNG\_p GR\_p MS\_p PM\_p RMS\_p GI\_p active\_days log\_est\_hours / diagonal=(histogram) group = group;

run;

\*prove that double peaking is prevalent in both groups and not just 1 group;

\*prove why we need to log age and est hours;

\*prove active days is left skewed but CLT applies when dealing w/ large data sets;

proc univariate data=LumosityCleaned ;

class group;

var log\_age\_round active\_days log\_est\_hours;

histogram;

run;

\*see the correlation among variables;

proc corr data=LumosityCleaned;

var log\_age\_round AR\_p TTS\_p GNG\_p GR\_p MS\_p PM\_p RMS\_p GI\_p active\_days log\_est\_hours;

run;

/\*findings: GI\_d is corelated to other variables, since it's derived variable\*/

\*automatic variable selection using stepdisc for LDA variable selection;

\*pre post without d because redundant;

proc stepdisc data=LumosityCleaned bsscp tsscp;

class group;

var log\_age\_round active\_days log\_est\_hours AR\_1 TTS\_1 GNG\_1 GR\_1 MS\_1 PM\_1 RMS\_1 GI\_1 AR\_2 TTS\_2 GNG\_2 GR\_2 MS\_2 PM\_2 RMS\_2 GI\_2;

run;

\*log\_est\_hours active\_days GI\_1 GI\_2 GR\_2 TTS\_2 AR\_1 AR\_2 log\_age\_round PM\_1;

\*pre post and precentage change;

proc stepdisc data=LumosityCleaned bsscp tsscp;

class group;

var log\_age\_round active\_days log\_est\_hours AR\_1 TTS\_1 GNG\_1 GR\_1 MS\_1 PM\_1 RMS\_1 GI\_1 AR\_2 TTS\_2 GNG\_2 GR\_2 MS\_2 PM\_2 RMS\_2 GI\_2 AR\_p TTS\_p GNG\_p GR\_p MS\_p PM\_p RMS\_p GI\_p;

run;

\*GI\_p log\_est\_hours active\_days AR\_p GR\_2 MS\_2 GNG\_2 log\_age\_round TTS\_p;

/\*LDA without dimensionality reduction\*/

\*pre and post;

proc discrim data=LumosityCleaned pool=test crossvalidate;

class group;

var log\_est\_hours active\_days MS\_2 GI\_1 GI\_2 GR\_2 TTS\_2 AR\_1 AR\_2 log\_age\_round PM\_1;

priors "crosswords"=0.5 "Lumosity"=0.5;

run;

\*error rate 0.3906 with MS\_2;

\*pre and post adjusted priors;

proc discrim data=LumosityCleaned pool=test crossvalidate;

class group;

var log\_est\_hours active\_days MS\_2 GI\_1 GI\_2 GR\_2 TTS\_2 AR\_1 AR\_2 log\_age\_round PM\_1;

priors "crosswords"=0.434358 "Lumosity"=0.565642;

run;

\*error rate 0.3743 with MS\_2;

proc discrim data=LumosityCleaned pool=test crossvalidate;

class group;

var log\_est\_hours active\_days GI\_1 GI\_2 GR\_2 TTS\_2 AR\_1 AR\_2 log\_age\_round PM\_1;

priors "crosswords"=0.5 "Lumosity"=0.5;

run;

\*error rate 0.3877 w/o MS\_2;

\*adjusted;

proc discrim data=LumosityCleaned pool=test crossvalidate;

class group;

var log\_est\_hours active\_days GI\_1 GI\_2 GR\_2 TTS\_2 AR\_1 AR\_2 log\_age\_round PM\_1;

priors "crosswords"=0.434358 "Lumosity"=0.565642;

run;

\*error rate 0.3678 w/o MS\_2;

\*pre post and percentage;

proc discrim data=LumosityCleaned pool=test crossvalidate;

class group;

var GI\_p log\_est\_hours active\_days AR\_p GR\_2 MS\_2 GNG\_2 log\_age\_round TTS\_p;

priors "crosswords"=0.5 "Lumosity"=0.5;

run;

\*error rate 0.3957;

\*adjusted;

proc discrim data=LumosityCleaned pool=test crossvalidate;

class group;

var GI\_p log\_est\_hours active\_days AR\_p GR\_2 MS\_2 GNG\_2 log\_age\_round TTS\_p;

priors "crosswords"=0.434358 "Lumosity"=0.565642;

run;

\*error rate 0.3779;

/\*PCA to reduce the dimensions\*/

\*pre and post;

ods graphics on;

proc princomp data=LumosityCleaned plots=all out=LumosityPCA;

var log\_age\_round active\_days log\_est\_hours AR\_1 TTS\_1 GNG\_1 GR\_1 MS\_1 PM\_1 RMS\_1 GI\_1 AR\_2 TTS\_2 GNG\_2 GR\_2 MS\_2 PM\_2 RMS\_2 GI\_2;

run;

ods graphics off;

\*pre post and percentage;

ods graphics on;

proc princomp data=LumosityCleaned plots=all out=LumosityPCAp;

var log\_age\_round active\_days log\_est\_hours AR\_1 TTS\_1 GNG\_1 GR\_1 MS\_1 PM\_1 RMS\_1 GI\_1 AR\_2 TTS\_2 GNG\_2 GR\_2 MS\_2 PM\_2 RMS\_2 GI\_2 AR\_p TTS\_p GNG\_p GR\_p MS\_p PM\_p RMS\_p GI\_p;

run;

ods graphics off;

\*automatic variable selection using stepdisc for LDA variable selection after PCA;

\*pre post without d because redundant;

proc stepdisc data=LumosityPCA bsscp tsscp;

class group;

var Prin1 Prin2 Prin3 Prin4 Prin5 Prin6 Prin7 Prin8 Prin9 Prin10 Prin11 Prin12 Prin13 Prin14 Prin15 Prin16 Prin17 Prin18 Prin19;

run;

\*Prin17 Prin16 Prin2 Prin7 Prin4 Prin8 Prin10 Prin15 Prin6 Prin13 Prin9 Prin18 Prin3;

\*pre post and precentage change;

proc stepdisc data=LumosityPCAp bsscp tsscp;

class group;

var Prin1 Prin2 Prin3 Prin4 Prin5 Prin6 Prin7 Prin8 Prin9 Prin10 Prin11 Prin12 Prin13 Prin14 Prin15 Prin16 Prin17 Prin18 Prin19 Prin20 Prin21 Prin22 Prin23 Prin24 Prin25 Prin26 Prin27;

run;

\*Prin17 Prin2 Prin10 Prin11 Prin3 Prin25 Prin14 Prin5 Prin24 Prin4 Prin8 Prin15 Prin26;

/\*LDA with PCA\*/

\*pre and post;

proc discrim data=LumosityPCA pool=test crossvalidate;

class group;

var Prin17 Prin16 Prin2 Prin7 Prin4 Prin8 Prin10 Prin15 Prin6 Prin13 Prin9 Prin18 Prin3;

priors "crosswords"=0.5 "Lumosity"=0.5;

run;

\*error rate 0.4131;

\*adjusted;

proc discrim data=LumosityPCA pool=test crossvalidate;

class group;

var Prin17 Prin16 Prin2 Prin7 Prin4 Prin8 Prin10 Prin15 Prin6 Prin13 Prin9 Prin18 Prin3;

priors "crosswords"=0.434358 "Lumosity"=0.565642;

run;

\*error rate 0.4131;

\*pre post and percentage;

proc discrim data=LumosityPCAp pool=test crossvalidate;

class group;

var Prin17 Prin2 Prin10 Prin11 Prin3 Prin25 Prin14 Prin5 Prin24 Prin4 Prin8 Prin15 Prin26;

priors "crosswords"=0.5 "Lumosity"=0.5;

run;

\*error rate 0.4134;

\*adjusted;

proc discrim data=LumosityPCAp pool=test crossvalidate;

class group;

var Prin17 Prin2 Prin10 Prin11 Prin3 Prin25 Prin14 Prin5 Prin24 Prin4 Prin8 Prin15 Prin26;

priors "crosswords"=0.434358 "Lumosity"=0.565642;

run;

\*error rate 0.4271;

## **Appendix III – Cognitive Assessment Details**

**Source:** Joseph L Hardy, Rolf A Nelson, Moriah E Thomason, Daniel A Sternberg, Kiefer Katovich, Faraz Farzin, and Michael Scanlon. "Enhancing Cognitive Abilities with Comprehensive Training: A Large, Online, Randomized, Active-Controlled Trial." PLoS ONE 10.9: E0134467. Web.

A more thorough description of the seven neuropsychological assessments used to measure cognitive performance at pre-test and post-test.

**Descriptions of Assessments Used in Primary Outcome Battery**

These assessments feature a white background with black text, and fill a window on the computer screen 640 pixels in width by 480 pixels in height, unless otherwise noted. Each assessment is introduced through text directions as well as two trials of interactive practice at the lowest difficulty level. Practice trials are repeated until the participant correctly completes the trials, ensuring that the participant understands the task. Feedback regarding correctness is given during practice trials, but is not given during regular trials of the assessment, unless otherwise noted.

**Forward Memory Span**

This assessment is based on the Corsi Blocks tasks [1]. Blue circles with radii equal to 1/20 of the window height are placed at randomized, non-overlapping spatial locations and individually highlighted in orange following a particular sequence. Circles are highlighted for 500 msec with a 100 msec inter-stimulus interval. The participant is asked to recall the sequence by clicking on each circle in same order as originally presented. The length of the sequence increases by one every two trials. The session ends when the participant gives two incorrect answers at the same span level. The total number of correct responses is the dependent measure. This task is considered a measure of visual short-term memory.

**Reverse Memory Span**

This assessment is identical to the forward visual memory span assessment, except that the participant is asked to recall the sequence of circles in the reverse order. It is considered a measure of visual working memory.

**Grammatical Reasoning**

The Grammatical Reasoning assessment [2] measures the participant’s facility with rapidly and accurately evaluating a potentially confusing grammatical statement. A blue square and a blue equilateral triangle, each with height equal to 1/5 of the window height are shown side by side, with a logical statement written below. For example, the square may be positioned to the left of the triangle on a particular trial. The participant could be prompted with a statement of the form, “The square is not to the left of the triangle.” In this case, the answer would be “false.” The participant responds whether the statement is true or false by pressing a key on the keyboard that is indicated as corresponding to true or false. The probability that the statement includes a negative (“not”) is 50%. The net number of correct responses (number correct – number incorrect) in 45 seconds is the dependent measure, with a floor of zero. This test is a measure of cognitive flexibility and reasoning.

**Progressive Matrices**

Matrix reasoning assessments (e.g., Raven [3]) require the participant to determine which stimulus most logically completes a multi-dimensional pattern. In this version of matrix reasoning, the participant is shown a 3x3 grid (each grid slot has width and height equal to 1/5 the window height) with abstract stimuli in the 8 upper-left slots. The task is to choose which of six possible answer choices best completes the pattern in the grid. The assessment is made up of 17 problems of increasing difficulty that are algorithmically generated from a set of parameters. The 17 problems are divided into three broad problem types: progression matrix, orbital/lateral movement, and Boolean logic. The first 12 trials involve progression matrix rules of increasing complexity. Characteristics of the stimuli that may change include shape, number, color, rotation angle, and size. These patterns may progress across the matrix horizontally, vertically, from upper-left to lower-right diagonal, or from lower-left to upper-right diagonal. Trials 13-15 involve orbital or lateral movement in which square grids or circular orbits are partially filled with elements that progress according to a lateral or rotational movement rule. Trials 16-17 involve Boolean logic in which spatial patterns are combined using Boolean operators such as AND, OR, and XOR. For each problem type, the correct answer is indicated regardless of whether the participant answers correctly. The assessment ends once the participant completes 17 trials or answers three consecutive trials incorrectly. The total number correct is the dependent measure. Matrix reasoning is considered a measure of problem solving and fluid reasoning (often referred to as fluid intelligence).

**Go/No-Go**

In this assessment, the participant must press the space bar in response to a target picture and withhold responding to distractor stimuli. The stimuli are chosen from a set of photos of fruit (apple, watermelon, pear, peach, orange, cantaloupe) and occupy roughly ¼ the height of the window. The identity of the target (e.g., a watermelon) is chosen randomly for a particular run of the assessment and shown to the participant prior to the beginning of the trials. On a given presentation, there is a 50% chance of the stimulus being a target. Each stimulus appears after a random delay varied between 1000 to 3000 msec to discourage anticipatory responding. The participant is told to respond as quickly as possible without making errors. The participant must respond to a “go” trial within 1500 msec. Timing and correctness feedback is given. The assessment ends when a participant responds to ten “go” trials. If a participant makes three errors (responding to “no-go” trials or failing to respond to “go” trials), the assessment is restarted. The dependent measure is the average reaction time on correct trials. This assessment is a measure of response inhibition and speed of processing.

**Arithmetic Reasoning**

This assessment [4] requires the participant to rapidly and accurately solve simple arithmetic problems that are written in words – for example, “Four plus two =”. The answer is input using the number keys on the keyboard (“6” in the current example). For addition and multiplication problems, operands are uniformly sampled from the integers in the range 1-9, and for subtraction and division problems, the second operand and solution are uniformly sampled from the integers in the range 1-9. The addition operator is used in the first five trials. Subsequent trials use addition 50% of the time. In non-addition trials, operators are chosen uniformly from subtraction, multiplication, and division. The assessment lasts 90 seconds, and the total number of correct responses in that time period is the dependent measure. It is considered a measure of problem solving ability.

**Two-Target Search**

In this assessment, the participant must identify two targets, while ignoring two distractors, which are spread across the field of view. In order to partially control for screen size and thus visual angle, prior to taking the assessment, the participant is instructed to hold a credit card-sized object up to the screen and adjust a slider until a representative rectangle matches in size. The display window is scaled accordingly. The participant is briefly presented with four circles containing letters. Each circle is 1 inch tall. Each of the four circles is distributed at a random angle along an invisible circle with a 2.5-inch radius centered in the middle of the screen. The minimum spacing between each stimulus circle is 0.75 inches in order to maximize the spread of visual attention [5]. Two of the letters are designated at the beginning of the session as target letters (e.g., “H” and “T”) selected from a set of six letters of similar shape (“F,” “H,” “K,” “T,” “X,” and “Z”). A fixation cross appears for 1000 msec, after which the four circled letters appear. After a brief presentation, the letters are obscured by visual masking noise made up of all six letters on top of each other (in order to prevent visual afterimage effects). Noise mask durations are equal to stimulus presentation durations. Next, the letters disappear but the circles remain. The participant is instructed to click on the two circles that had contained the target letters. Difficulty is manipulated using an up/down staircase method [6] with exponential scaling. The first trial begins with a 1000 msec presentation time and subsequent trials are adjusted by multiplying presentation time by 10-0.15 after a correct trial and 100.15 after an incorrect trial and rounding up to the nearest 1/30 sec due to screen refresh rate quantization (possible range: 33.333-44666 msec). There are 12 trials total. The threshold presentation time derived from the staircase procedure is the dependent measure. The threshold presentation time is defined as the average of all reversal times (presentation times of trials in which the direction of the staircase switched direction due to answering correctly after a previously incorrect trial or vice-versa) and a potential final presentation time. The final presentation time is defined as the presentation time calculated for what would be the 13th trial, and is included in the average if the final trial was not a reversal trial (possible range: 15.849-63096 msec). This task is designed to be a measure of divided visual attention.

**References:**

1. Milner B. Interhemispheric differences in the localization of psychological processes in man. Br Med Bull. 1971;27(3):272-7.
2. Baddeley AD. A 3 min reasoning test based on grammatical transformation. Psychon Sci. 1968;10(10):341-2.
3. Raven J. The Raven's progressive matrices: change and stability over culture and time. Cogn Psychol. 2000;41(1):1-48.
4. Deloche G, Seron X, Larroque C, Magnien C, Metz-Lutz MN, Noel MN, et al. Calculation and number processing: assessment battery; role of demographic factors. J Clin Exp Neuropsychol. 1994;16(2):195-208.
5. Sekuler R, Ball K. Visual localization: age and practice. J Opt Soc Am A. 1986;3(6):864-
6. Levitt H. Transformed up-down methods in psychoacoustics. J Acoust Soc Am. 1971;49(2B):467-77.

## **Appendix IV – Additional Displays**

Display A: the cross comparison between the two QDA auto selection models with and without percent change

|  |  |  |
| --- | --- | --- |
| **Auto selection without percent change** |  | **Auto selection with percent change** |
| active\_days | **kept** | active\_days |
| AR\_1 | **replaced** | AR\_p |
| AR\_2 |
| GI\_1 | **replaced** | GI\_p |
| GI\_2 |
| GR\_2 | **kept** | GR\_2 |
| log\_age\_round | **kept** | log\_age\_round |
| log\_est\_hours | **kept** | log\_est\_hours |
| TTS\_2 | **replaced** | TTS\_p |
| PM\_1 | **changed** | GNG\_2 |
|  | MS\_2 |

It is noted that perhaps the addition of Two target search percent change was preferred over its post score would have made Progressive Matrices pre score colinear and left out. As a result of leaving out Progressive Matrices pre score, Go/Not Go post score and memory span post score are added to compensate for the QDA predictive model. Also, the auto selection without percent change removed memory span post score in a later elimination while auto selection with percent change kept it in.

As predicted in the data overview, grammatical reasoning is a determining factor to distinguish the crossword from the Lumosity group because Lumosity doesn't increase your vocabulary and verbal usage when doing their tests while doing crosswords will.

Display B: the cross comparison between the two QDA auto selection models with and without percent change after PCA in terms of correlated variables

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Auto selection without percent change** | |  | **Auto selection with percent change** | |
| Variable correlated | Times in Top 3 |  | Variable correlated | Times in Top 3 |
| active\_days | 1 | **kept** | active\_days | 2 |
| AR\_1 | 2 | **expanded** | AR\_1 | 2 |
| AR\_2 | 1 | AR\_2 | 1 |
|  |  | AR\_p | 1 |
| GI\_1 | 2 | **changed** | GI\_p | 1 |
| GNG\_1 | 4 | **expanded** | GNG\_1 | 2 |
| GNG\_2 | 5 | GNG\_2 | 2 |
|  |  | GNG\_p | 1 |
| GR\_1 | 3 | **expanded** | GR\_1 | 4 |
| GR\_2 | 3 | GR\_2 | 3 |
|  |  | GR\_p | 4 |
| log\_age\_round | 2 | **kept** | log\_age\_round | 3 |
| log\_est\_hours | 2 | **kept** | log\_est\_hours | 1 |
| MS\_2 | 2 | **changed** | MS\_1 | 1 |
|  |  | MS\_p | 1 |
| PM\_1 | 3 | **kept** | PM\_1 | 1 |
| PM\_2 | 2 | PM\_2 | 1 |
| RMS\_1 | 1 | **expanded** | RMS\_1 | 1 |
| RMS\_2 | 1 | RMS\_2 | 1 |
|  |  | RMS\_p | 2 |
| TTS\_1 | 4 | **expanded** | TTS\_1 | 1 |
| TTS\_2 | 1 | TTS\_2 | 1 |
|  |  | TTS\_p | 2 |

Even though memory span pre score and percent change was preferred over its post score, the pre score and a percentage change could derive the post score. The pair of pre and post scores not expanded to include the percent change was Progressive Matrices. The variables correlated post PCA have similar selection patterns to its non-PCA counterpart.

The full equation for principal component 1 without percent change is Prin1 = 0.390087\*GI\_2 + 0.389261\*GI\_1 + 0.30071\*AR\_2 + 0.299549\*AR\_1 + 0.257981\*GR\_2 + 0.253276\*GR\_1 + 0.225913\*RMS\_1 + 0.220251\*MS\_2 + 0.215588\*MS\_1 + 0.212342\*PM\_2 + 0.211194\*RMS\_2 + 0.199338\*PM\_1 + 0.173887\*TTS\_2 + 0.164312\*TTS\_1 + 0.140878\*GNG\_1 + 0.129273\*GNG\_2 + 0.095933\*active\_days + 0.082578\*log\_est\_hours + 0.013397\*log\_age\_round

The full equation for principal component 1 with percent change is Prin1 = 0.391976\*GI\_1 + 0.38655\*GI\_2 + 0.300705\*AR\_1 + 0.299398\*AR\_2 + 0.257897\*GR\_2 + 0.253397\*GR\_1 + 0.229732\*RMS\_1 + 0.21759\*MS\_2 + 0.217547\*MS\_1 + 0.211529\*PM\_2 + 0.207005\*RMS\_2 + 0.19861\*PM\_1 + 0.173121\*TTS\_2 + 0.164698\*TTS\_1 + 0.144797\*GNG\_1 + 0.12514\*GNG\_2 + 0.094931\*active\_days + 0.081525\*log\_est\_hours + 0.012845\*log\_age\_round + 0.008614\*PM\_p + 0.007067\*TTS\_p + 0.0063\*GR\_p - 0.008749\*AR\_p - 0.009118\*MS\_p - 0.017742\*GI\_p - 0.024455\*RMS\_p - 0.02963\*GNG\_p