**Online Appendices at: www.paulvangeert.nl/articles\_appendices.htm**

**Appendix I: Preschoolers' Learning Behavior Questionnaire**

Name of child:..................................................................................................................

Date of birth:....................................................................................................................

Sex:...................................................................................................................................

Name rater:.......................................................................................................................

Date: ................................................................................................................................

Fill out the category (unresponsive, reflective or impulsive) that fits best with the preschool child's learning behavior, according to the regular learning behavior during a task, such as making puzzles, building blocks or drawing. Read the definitions of the three concepts carefully in order to circle the best fitting category. Please circle only one category.

The preschool child regularly works concentrated on a task. However, the child does start only with help or with a stimulus. The child waits until someone encourages or stimulates him or her to take action.

***Unresponsive***

The preschool child regularly works concentrated on a task and regularly completes a task. The preschool child takes initiative to start a (new) task and does what has been requested.

***Reflective/Normal***

The preschool child is easily distracted, stops relatively fast during a task and after that starts a new task. The child regularly gives up relatively fast and is very active.

***Impulsive***

**Appendix II: Combining and comparing multiple mouse behavior characteristics in Excel**

**Effect sizes**

First, differences in the height of each standardized mouse characteristic between two groups of children were calculated, for example, the difference between average reaction time of the group of ASD children and the average reaction time of the group of ADHD children was calculated. Next, an average of multiple mouse behavior (MB) characteristics was calculated to calculate the average effect size of multiple characteristics.

**Significance of differences**

To examine differences in MB profiles in different types of children (e.g. ADHD, ASD etc.), we used Excel, see Table 1.

Table 1: Empirical and reshuffled fictive data of one MB characteristic per group in Excel

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Empirical data | Reshuffled data |
| Participant | Diagnosis | Reaction time | Average reaction time per diagnosis | Reaction time | Average reaction time per diagnosis |
| 1 | ADHD | 0.148637 |  | 0.241512 |  |
| 2 | ADHD | -0.05314 | 0.047748 | 0.148637 | 0.195074 |
| 3 | Comorbid | -0.17227 |  | -0.23593 |  |
| 4 | Comorbid | -0.07613 |  | -0.2956 |  |
| 5 | Comorbid | -0.12984 | -0.12608 | -0.0927 | -0.20808 |
| 6 | ASD | -0.23593 |  | -0.03634 |  |
| 7 | ASD | -0.14642 |  | -0.10198 |  |
| 8 | ASD | -0.0927 |  | -0.12917 |  |
| 9 | ASD | 0.241512 |  | 0.14537 |  |
| 10 | ASD | 0.587005 |  | -0.17227 |  |
| 11 | ASD | -0.15968 |  | 0.268706 |  |
| 12 | ASD | -0.04365 |  | 0.088338 |  |
| 13 | ASD | 0.074404 |  | 0.074404 |  |
| 14 | ASD | 0.088338 |  | -0.07613 |  |
| 15 | ASD | 0.268706 |  | -0.02308 |  |
| 16 | ASD | -0.02308 |  | -0.14642 |  |
| 17 | ASD | -0.03634 |  | 0.451833 |  |
| 18 | ASD | 0.14537 | 0.051348 | -0.04365 | 0.023048 |
| 19 | EL | 0.451833 |  | -0.05314 |  |
| 20 | EL | -0.12917 |  | -0.12984 |  |
| 21 | EL | -0.2956 |  | -0.07545 |  |
| 22 | EL | -0.07545 |  | -0.15968 |  |
| 23 | EL | -0.10198 | -0.03007 | 0.587005 | 0.033778 |

Table 2: Comparing MB characteristics per group in Excel

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  EL compared to ADHD | Higher reaction time | Lower missing go's | Lower repeats | Lower no go's | Lower go's | Lower errors |  Sum |
| EL  | =if(average shuffled data >= average empirical data,1,0) | =if(average shuffled data <= average empirical data, 1,0) | **Etc.** | **..** | **..** | **..** | =if(sum (whole row)=6,1,0 |
| ADHD | = if (average shuffled data <= average empirical data, 1,0) | =if(average shuffled data >= average empirical data,1,0) | **Etc.** | **..** | **..** | **..** | =if(sum (whole row)=6,1,0 |
|  |  |  |  |  |  |  | =if(sum(whole column)=6,1,0 |
| **Results of Monte Carlo Analysis** |  | EL | ADHD | Combination |
|  |  |  |  | Average | 0.003 | 0 | 0 |
|  |  |  |  | # of simulations | 1000 | 1000 | 1000 |

**Explanation:**

First, we calculated averages per MB characteristic per group (based on residuals of empirical data, since the data were corrected for Composite Standard Scores (fourth column, Table 1).

Then we randomly reshuffeled the empirical data and calculated averages per type of child (sixth column, Table 1) based on the randomly permuted data. Random permutation is carried out by means of the reshuffle procedure in Poptools which is an add-in in Excel.

Next, we analyzed whether there were significant differences between the four types of children. For example, we examined whether the EL children on average show better MB skills than the ADHD children. To analyze this, we first had to score whether the reshuffled average was smaller or larger than the empirical average.

For example, we hypothesized that *EL children show higher reaction times than ADHD children, but a lower number of errors*. The null hypothesis was that there were no differences between the ADHD and EL children. In the empirical data, we observed that the reaction times and errors indeed differed, corresponding to our hypothesis. However, we still had to test whether the *combined* averages of multiple mouse MB characteristics differed significantly from one another. In order to study the probability of combined differences we used a relatively simple criterion, which is that the null hypothesis model obtains a positive score if it does as well or eventually better than the observed scores. For instance, in case of a *lower* or similar reshuffled average for errors than the empirical average for errors in EL children, this was scored with a '1'. This was positively scored with a '1', since we will have low error scores. Thus, if a particular, randomized run off the null hypothesis model assigns an error score to the simulated EL children which is as low or even lower than the observed error score in the EL children, the null hypothesis is "rewarded" by giving it a score of ‘1’. If the reshuffled score, i.e. the score based on the null hypothesis model, was not lower than the score on the empirical data, this was scored with a '0'.

The scores for the ADHD children were the opposite from the EL children, therefore, in case of a *higher* reshuffled average for errors in ADHD children than the empirical average for errors, this was scored with a '1' (since this can be interpreted as positive) (for formulas, see Table 2).

However, in case of hypothesizing higher reshuffled averages, for example in case of reaction times of EL compared to ADHD children, the reshuffled average had to be higher, and was then scored with a '1'.

In case of comparing multiple MB characteristics, for example six in total, the maximum of the sum of the scores was 'six'. In case of a sum lower than six, this was scored as a ‘0’; otherwise it was scored with a '1'.

If the randomized run of the null hypothesis model received a score of ‘1’ on all the tested variables, it does as well as the observed score, and in this case receives a summary score of ‘1’ (if not it receives a ‘0’ score). By carrying out many such randomized runs, in our case 1000, we can calculate the number of times the null hypothesis receives a summary score of ‘1’. By dividing this number by the number of simulations (1000) we obtain an estimation of the p-value of the null hypothesis model. In case of an average lower than .05, we concluded that the empirical average differed significantly from the reshuffled average. This was conducted for the EL children, for the ADHD children, but also for the combination of the groups (which indicates that both groups differ from each other).

**Appendix III: Explanation procedure Figure 7**



Figure 7

First, we have standardized and centered each variable over the data of the three children (i.e. on the basis of 39 data points for each variable, namely 13 games times three children). By doing so, we have conserved the differences between the children, while making the six variables comparable to one another (each now has an average of zero and a standard deviation of one). Second, we have detrended the data by means of a simple differencing technique, which means that every variable for each individual child is now replaced by a series of different scores between the next in the preceding game (12 such different scores per child). We then replaced the difference scores by distance scores, which are simply the absolute values of the difference scores. For each game, we calculated the average of these distance scores for each individual child. These average distance score provides a representation of the amount of variability during each game (which is compared with its preceding game). Finally, these variability scores were smoothed (Savitzky Golay method, window size 3 consecutive observations).

**Appendix IV: Explanation procedure Figure 3**



First, we have standardized and centered each variable over the data of the four groups (i.e. on the basis of 52 data points for each variable, namely 13 games times four groups). By doing so, we have conserved the differences between the children, while making the six variables comparable to one another (each now has an average of zero and a standard deviation of one). Second, a variable clustering procedure was applied on the six variables (VARCLUS routine available in Tanagra). Three clusters were retained (see Table xx)

Table xx

|  |  |  |  |
| --- | --- | --- | --- |
| **Mouse behaviour characteristic** | **Cluster 1** | **Cluster 2** | **Cluster 3** |
| Errors | 0.91 | 0.01 | -0.09 |
| No go | 0.85 | 0.15 | -0.03 |
| Missing go | -0.18 | -0.17 | 1.00 |
| Reaction time | -0.72 | 0.22 | 0.32 |
| Repeats | 0.06 | 1.00 | -0.17 |
| Go | 0.72 | 0.24 | -0.17 |

The first cluster was dubbed “fast errors”, since it correlated highly with the variables errors, no go and go and correlated negatively with the reaction time. The second cluster correspondence with the variable repeats and the third with the variable missing go.

**Appendix V: Smoothing procedure**

In the case study of individual mouse behavior trajectories, the mouse behavior (MB) data across games were smoothed, in order to reveal the general form of the trajectory while keeping local spurts, plateaus and regressions.

In the case study of individual MB trajectories, the MB data across games were smoothed, performed in Table Curve 2D, version 5.0, using a Savitsky Golay smoothing technique, with the best fit for the data, according to the AI Expert within the program, which was 67.3% for the ADHD and EL child and 62.2% for the ASD child.

**Appendix VI: Min-max graphs procedure**

With min-max graphs, score ranges were plotted for each data point by using a moving window (Verspoor, Lowie & Van Dijk, 2008) to investigate whether there are fluctuations in MB over time. This was conducted with five positions, thus, min (t1…5), min (t2…6), et cetera & max (t1…5), max (t2…6) et cetera.

**Appendix VII: Improvement scores**

The total improvement score is the sum of the improvements on each of the variables, which have been first rescaled to z-scores in order to guarantee similar dimensionality of the variables. Second, by means of a linear regression of z-scores of all mouse characteristics within one group, the difference between the first and last point of the regression line was computed, which defined the change score for the variable. The change scores were turned into improvement scores by reversing the direction of the regression line, depending on an improvement criterion which could be different for each variable and group. For instance, reaction times in ADHD children and go in ASD and comorbid children must increase in order to improve. However, reaction times of ASD children must decrease, i.e. they must learn to react faster, in order to count as an improvement. Thus, improvement is always represented by a curve with positive slope, e.g., improvement as a negative regression line is defined as [- (intercept + slope \* game number)]. The total improvement score for a particular group of children is equal to the sum of improvement scores of the variables.

**Appendix VIII: Additional min-max graphs Figure 6**

The bandwidth of a developmental process (based on the observed values) of each child and its variability within the process.