Validity Performance in the AULA Nesplora Test

Fidel Rebon, Iñigo Altuna, Amaya Lobo, Elena Salillas, Gema Climent

MOTIVATION FOR THE STUDY

It is an acclaimed necessity that performance tests in any domain include an integrated performance validity subtest as suboptimal effort is a stumbling block to accurate assessment, especially in compensation-seeking individuals (Bashem, 2022). As has been advocated for the past two decades**, the use of psychometric indicators is the most valid approach to identify neuropsychological response validity**, as well as independent effort measures and embedded validity indicators (EVI). Employing an EVI constructed by the test authors, or by the owners of the commercial operation, rather than relying on external validators, represents a methodological advance, as it reduces the level of inference involved in determining the validity of a given test score (Erdodi, 2014). Analysing patterns of performance *within the instrument* to establish the credibility of the response set seems to be a logical approach in assessing performance validity, although for more demanding domains highly sensitive measures may be needed in addition to the embedded ones (Bussé, 2012).

AULA Nesplora (*Climent & Banterla, Nesplora 2011*) is a CPT for the **support of the diagnosis of attentional processes** with 260 items distributed in two tasks: a first Xno, very stimulating, and an X task, less stimulating. The theoretical paradigm is based on Sergeant and Barkley (Servera, 2005) and their theories on attention regulation. The test is normalised for subjects aged 6 to 16 years (Iriarte & al. 2012).

The motivation of our current study is the implementation of an EVI in the AULA report, following the guidelines of the *American Academy of Clinical Neuropsychology Consensus Conference Statement on the Neuropsychological Assessment of Effort, Response Bias, and Malingering* and the criteria both from other CPTs and the literature, using a random sample of anonymised test runs for matching and verifying that the implemented EVI can extract and detect test runs that do not conform to the inclusion/exclusion guidelines. In this way, we can help professionals detect test runs that do not meet the established criteria and may mean that the test has not been executed properly.

HYPOTHESIS AND OBJECTIVE

The Performance Validity to be considered in the EVI is **a measure of unusual patterns of AULA performance** that are not typically seen in children with attention problems or suspected attention problems, hence the large clinical sample collected. Conditions that may signal unusual performance in children are: **oppositional behaviour, test maladaptation, NON-COMPREHENSION OF THE TEST**, or simply **non-cooperation on the part of the child**. Extreme cases of attention problems may also be seen in these performances.

Although the professional evaluator monitors the intentions, symptom validation schemas, and the effort put into the test, it is the obligation of those who design and create the tests to offer ease of detection and to incorporate an EVI in the test itself, especially when there is sufficient clinical data with the necessary psychometric indicators to establish at least a warning note to the evaluator, so that he/she can activate or increase his/her vigilance in this respect.

METHODOLOGY

Environment:

Using the **AULA Nesplora** Virtual Reality test (*Climent and Banterla, 2012*).

- Virtual reality headset with motion sensors and headphones.
- Classroom scenario and embedded task.
- Hyperstimulation and hypostimulation; visual and auditory

The **sample** consisted of **1235 clinical test runs** with 1129 unique subjects, anonymised and authorised by our clinics and research collaborators.

... by employing the following statistical techniques ...

- Data cleaning, **elimination of incongruent cases** under neuropsychological supervision.
- **Stratification** of cases **by means of the test scales** (gender and age).
- **Determine technique for obtaining performance outliers** (outliers - multivariables) in the whole sample by stratification: measurement of distances (*Cook, Mahalonobis*), clustering (*hierarchical, kmeans, diana, model, pam, clara, agnes*), statistical tests (*Grubbs's test, Dixon's test, Rosner's test*).
- Calculation of the **Mahalonobis** distance to determine the approximations of the performances and use a **Chi-square** with a probability of 99% to determine those cases that are outliers.
- **Correlation of the distances** obtained (**Pearson**) with the variables selected for the study in order to know which have the greatest influence when making the cut-off.
- **Establishment of cut-off points** for each variable.
- **Peer review** under expert (neuropsychological) criteria of the cases excluded by means of the cut-off points.
- **Readjustment of the cut-off points** if necessary.

RESULTS

108 test runs do not pass the filter. Approx. **9% of cases**.

Of which:

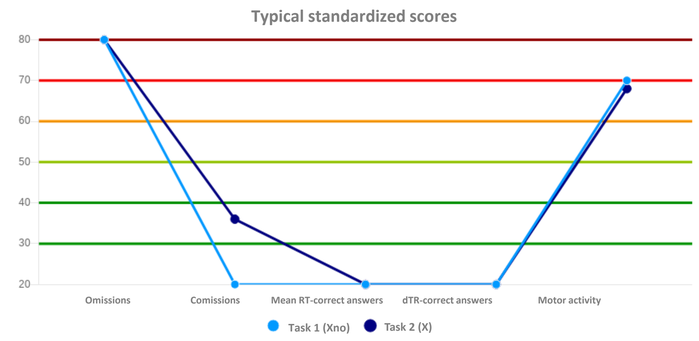
- 8 are under 6 years old.
- 26 aged 6 years.
- 18 aged 7 years.
- 17 aged 8 years.
- 12 aged 9 years.
- 9 aged 10 years.
- 5 aged 11 years.
- 13 aged from 12 to 16 years.

FILTERING VARIABLES
t_correctreactime_mean
t_commissionreactime_mean
x_commission_n
xno_commission_n
x_omission_n
xno_omission_n
diff_time_mean

The filter detects that the required score has not been reached in task Xno for 27 subjects, and in task X for 49 subjects.

24 subjects performed anomalously **on both tasks** and 7 of these 24 subjects did not perform any clicking in the whole test.

... graph of a report with atypical results in its variables ...



Only the clinician can determine both the diagnosis and the cause of the need for monitoring or the need for repetition.

The ultimate decision on the validity or otherwise of the execution must always be made by the assessing professional.

The implemented algorithm is an anomalous performance classifier, which can give an idea that the test has been taken without interest, quickly without facing it, or that some of the tasks have not been understood.

It is necessary and useful to avoid false positives and performances that cannot be classified as "normal".

Although it is not very common to suspect non-credible performances in children, cases have been reported in adolescents and adults*. In any case, recommendations for integrating performance validity also include children's measures.

CONCLUSIONS

AULA is sometimes administered to children under 6 years of age; of the sample of 1129, 22 subjects are under 6 and **8 of them are filtered by the EVI**, which indicates that AULA at these ages is not only does not have a normative but that it **is likely that the screening may be indicative in some children but effectively not in all of them**.

76 subjects are only "invalidated" in one task, not in both, which may correspond to a lack of understanding of the task in question. The fact that the filter affects the second task more (49) may also be indicative of a lack of regulation rather than a lack of understanding of the second task.

The following paragraph has been added to the **AULA report**: *For the filtering of this assessment, an EVI has been used, the use of which in children is recommended according to the American Academy of Clinical Neuropsychology Consensus Conference Statement on the Neuropsychological Assessment of Effort, Response Bias, and Malingering.*

This ratio (EVI) shows performance problems during the administration of the test that assesses whether problems of incongruence are detected that could affect the results before a clinical diagnosis is considered.

Option 1

```
diff_time_mean = round((x_commissionreactime_mean -
x_correctreactime_mean)/1000, 2),
x_correctreactime_s =  round(sqrt(x_correctreactime_var)/1000, 2)

if(
    (t_correctreactime_mean < 150 &
     t_commissionreactime_mean < 150 & x_commission_n > 14
&   xno_commission_n > 4) |

    (xno_omission_n > 130  & x_omission_n > 27) |
    (xno_commission_n > 27 & x_commission_n > 130) |
 (diff_time_mean > 75 & diff_time_mean > x_correctreactime_s)
)
In the case of (name) this assessment meets the requirements to be
considered to be supervised in both parts of the test.
else(){
In the case of (name), this assessment meets the requirements to be
considered adequate in both parts of the test.
}
You may retest from (experiment_date, + two weeks).
```

Option 2

```
if(
    (t_correctreactime_mean < 150 &
     t_commissionreactime_mean < 150 & x_commission_n > 14 &
xno_commission_n > 4) |

    (xno_omission_n > 130  & x_omission_n > 27) |
    (xno_commission_n > 27 & x_commission_n > 130) |

    (xno_omission_n > 130  & x_omission_n < 27) | # No entiende tarea 1
    (xno_commission_n > 27 & x_commission_n < 130) | # No entiende
tarea 1
)
In the case of (name) this assessment meets the requirements to be
considered to be supervised in task 1 of the test.
else(){
In the case of (name), this assessment meets the requirements to be
considered adequate in task 1 of the test.
}
You may retest from (experiment_date, + two weeks).
```

Option 3

```
if(
    (t_correctreactime_mean < 180 &
     t_commissionreactime_mean < 180 & x_commission_n > 14 &
xno_commission_n > 4) |
    (xno_omission_n > 130  & x_omission_n > 27) |
    (xno_commission_n > 27 & x_commission_n > 130) |

    (xno_omission_n < 130  & x_omission_n > 27) | # No entiende tarea 2
    (xno_commission_n < 27 & x_commission_n > 130) | # No entiende
tarea 2

In the case of (name) this assessment meets the requirements to be
considered to be supervised in task 2 of the test.
else(){
In the case of (name), this assessment meets the requirements to be
considered adequate in task 2 of the test.
}
In task 2, you can retest from day (experiment_date, + two weeks).
```

BIBLIOGRAPHY

Bashem, J. (2022). *Detecting suboptimal effort in traumatic brain injury assessment*.

Busse, M., & Whiteside, D. (2012). Detecting Suboptimal Cognitive Effort: Classification Accuracy of the Conner's Continuous Performance Test-II, Brief Test of Attention, and Trail Making Test. *The Clinical Neuropsychologist*, *26*(4), 675-687. https://doi.org/10.1080/13854046.2012.679623

Erdodi, L. A., Roth, R. M., Kirsch, N. L., Lajiness-O'neill, R., & Medoff, B. (2014). Aggregating Validity Indicators Embedded in Conners' CPT-II Outperforms Individual Cutoffs at Separating Valid from Invalid Performance in Adults with Traumatic Brain Injury. *Archives of Clinical Neuropsychology*, *29*(5), 456-466. https://doi.org/10.1093/arclin/acu026

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, *23*(7), 1093-1129. https://doi.org/10.1080/13854040903155063

Iriarte, Y., Diaz-Orueta, U., Cueto, E., Irazustabarrena, P., Banterla, F., & Climent, G. (2016). AULA—Advanced Virtual Reality Tool for the Assessment of Attention: Normative Study in Spain. *Journal of Attention Disorders*, *20*(6), 542-568. https://doi.org/10.1177/1087054712465335

Servera Barceló, M. (2005). Modelo de autorregulación de Barkley aplicado al trastorno por déficit de atención con hiperactividad: Una revisión. *Revista de Neurología*, *40*(06), 358. https://doi.org/10.33588/rn.4006.2004364

Sharland, M. J., Waring, S. C., Johnson, B. P., Taran, A. M., Rusin, T. A., Pattock, A. M., & Palcher, J. A. (2018). Further examination of embedded performance validity indicators for the Conners' Continuous Performance Test and Brief Test of Attention in a large outpatient clinical sample. *The Clinical Neuropsychologist*, *32*(1), 98-108. https://doi.org/10.1080/13854046.2017.1332240

*For this particular motivation, work is being done on the EVI for Aquarium and Ice Cream. Suite already has a built-in EVI. In the detail of the EVIs for each test, a secondary gain benefit should be analysed in the forensic field or medication procurement