**Final report**

# Analysing of validity and reliability of TALi DETECT

## Contents




**Publisher:**

TALi Health Pty Ltd.



Correspondence concerning this article should be addressed to:
Dr. Azadeh Feizpour; E-mail: afeizpour@talihealth.com.

## 1. Executive Summary

A comprehensive data analysis showed that TALi DETECT has an excellent factor structure with a three-factor model comprising of Selective Attention, Sustained Attention and Cognitive Flexibility. The model overall provides excellent fit to the data, both when examined at a single point in time and for a repeated assessment over time.

This study provides evidence of the reliability and stability of TALi DETECT over time. The test-retest reliability is very strong between all factors except cognitive flexibility. Using the current set of measures, TALi has adopted this three-factor model as its attentional model for TALi DETECT score and report generation platform meanwhile development work has commenced to increase the reliability of the cognitive flexibility factor.

TALi DETECT shows acceptable correlation for most subtests with other standard measures of childhood attention and is capable of reliably showing whether a new participant is above or below the average of the reference group in terms of their attentional abilities.

The outcomes of analysis provide the necessary evidence to conduct further studies in clinical paediatric populations. Such studies will enable TALi DETECT to show the point at which a person's different pattern of attentional abilities could be considered problematic, or diagnostic of a specific disorder.

## 2. Introduction

TALi DETECT is a gamified assessment tool designed to assess the attentional capabilities of millions of children entering the education system annually. The gamified subtests are based on well-established, standardised cognitive assessments that have been modified for use with young children. While TALi DETECT is not a formal diagnostic tool for identifying developmental delays or disorders, it is designed to triage children with early indication of attention vulnerabilities to appropriate early interventions.

## 3. Project Objectives

The aim of this project was to evaluate the validity and reliability of TALi DETECT as a screening tool for early childhood attention vulnerabilities. The project assessed whether:

- TALi DETECT has a reliable factor structure for differentiating the core domains of attention (construct validity): Selective Attention, Sustained Attention and Executive Attention.
- TALi DETECT produces consistent results over time (test-retest reliability).
- TALi DETECT is comparable to existing performance-based measures of cognitive attention (convergent validity)

## 4. Methodology

### 4.1 Participants

This project involved 8 Victorian Primary Schools, 6 Victorian Kindergarten and Childcare Centres, 61 Teachers and Principals, 351 Parents/guardians, and 351 Children. The sample consisted of 179 males and 172 females. All parents provided written informed consent. The project was approved by Monash University Human Research Ethics Committee and the Department of Education and Training.

Children were eligible if they were:

(a) in participating preschool, preparatory/foundation or Year 1 classes;

(b) fluent in English;

(c) did not have an intellectual disability based on parent reports and confirmed by IQ Composite scores >70 on the Kaufman Brief Intelligence Test at baseline (KBIT-2; Kaufman & Kaufman, 2004); and

(d) did not have a diagnosed neurodevelopmental disorder such as ADHD.

Of the 351 children, 11 were excluded from analysis due to ineligibility. The final analysed sample therefore consisted of 340 children aged 4 to 8 years.

## 4. Methodology (continued)

### 4.2 Measures

All assessments were conducted by trained researchers. Assessment sessions with children involved the completion of TALi DETECT (20-30 minutes) and an existing computerised assessment of attention (Test of Everyday Attention for Children; TEA-Ch 2: 30-40 minutes).

- TALi DETECT consists of 7 game-based assessment tasks which measure children's cognitive attentional ability.
- Seven subtests of the Test of Everyday Attention for Children Second Edition, (TEACh-2J; developed based on Manly et. al 2001) which is designed for children aged 5 to 7 years, measured cognitive attention processes.

| TALi DETECT subtest | Shortened version | Literature-based Cognitive Task |
|---|---|---|
| Feed Whiz | Speed | Simple Reaction Time Task |
| Look for Lobsters | Seek | Visual Search Task |
| Scan the Sky | Scan | Visual Search Task |
| Find the Frog | Monitor | Divided attention task |
| Don't Pat the Pig | Focus | SART (sustained attention to response task: high Go, low No-Go, fixed order) |
| Make Jam | Inhibit | Stop Signal Task (SST) |
| Sort the Blocks | Switch | Dimensional Change Card Sort (DCCS) |

To assess whether TALi DETECT was a reliable assessment of attention, a second assessment of TALi DETECT was completed on average 2 weeks after the first assessment, for 50% of the included sample (156 children).

### 4.3 Hypothesised 3-factor model

In the following table, the Simple Reaction Time task was considered as a baseline assessment of visuomotor speed and was used as control variable in the structural equation modelling explained in the Results sections. Furthermore;

- The latent variable of Selective Attention was hypothesised to be estimated by the observed variables from TALi DETECT subsets of "Look for Lobsters" and "Scan the Sky";
- The latent variable of Sustained attention was hypothesised to be estimated by the observed variables from TALi DETECT subsets of "Find the Frog" and "Don't Pat the Pig";
- The latent variable of Executive attention (alternatively referred to as Attentional Control) was hypothesised to be estimated by the observed variables from TALi DETECT subsets of "Make Jam" and "Sort the Blocks".

The important variable of SSRT (stop-signal reaction time) couldn't be estimated from the Make Jam (SST) subtest of DETECT due to low number of stop levels. Therefore, this subtest was excluded from the final analysis.

## 4. Methodology (continued)

### 4.4 Approach to Analysis

All analyses were performed using the structural equation modelling (Asparouhov, T., & Muthén, B., 2009, West, S. G. et al., 2012) package, Mplus 8.4 (Muthén & Muthén, 2017). Firstly, the data were inspected and prepared (including data transformation) for later data analysis. Then, the first wave of the data was used to compare different factor models to determine the optimal one. For each model, model fit statistics, internal consistency, and discriminant validity statistics were investigated. Next, the test-retest reliability was assessed using the optimal factor model. In this stage, measurement invariance across time was also assessed. Finally, using the available data, the convergent validity of TALi DETECT was examined.

## 5. Results

### 5.1 Construct Validity

Five different models were tested by a biostatistician expert in high-level statistical modelling[1]:

1. The original hypothesised measurement model
2. A 1-factor model
3. A bi-factor measurement model
4. A measurement model based on exploratory structural equation modelling
5. A modified measurement model, based on a diagnosis of problems from the original measurement model

Among the examined models above, the most theoretically plausible model with the best fit indices (model 5) is presented below.

In the diagrams below, the squares represent the observed variables (i.e., the score for an individual on each of the TALi DETECT subtests). The circles represent the latent factor (i.e., Selective, Sustained and Executive Attention). The lines with directional arrows represent factor loadings (i.e., the extent to which each latent variable is associated with each observed variable). The lines with bidirectional arrows represent correlations between factors.

---

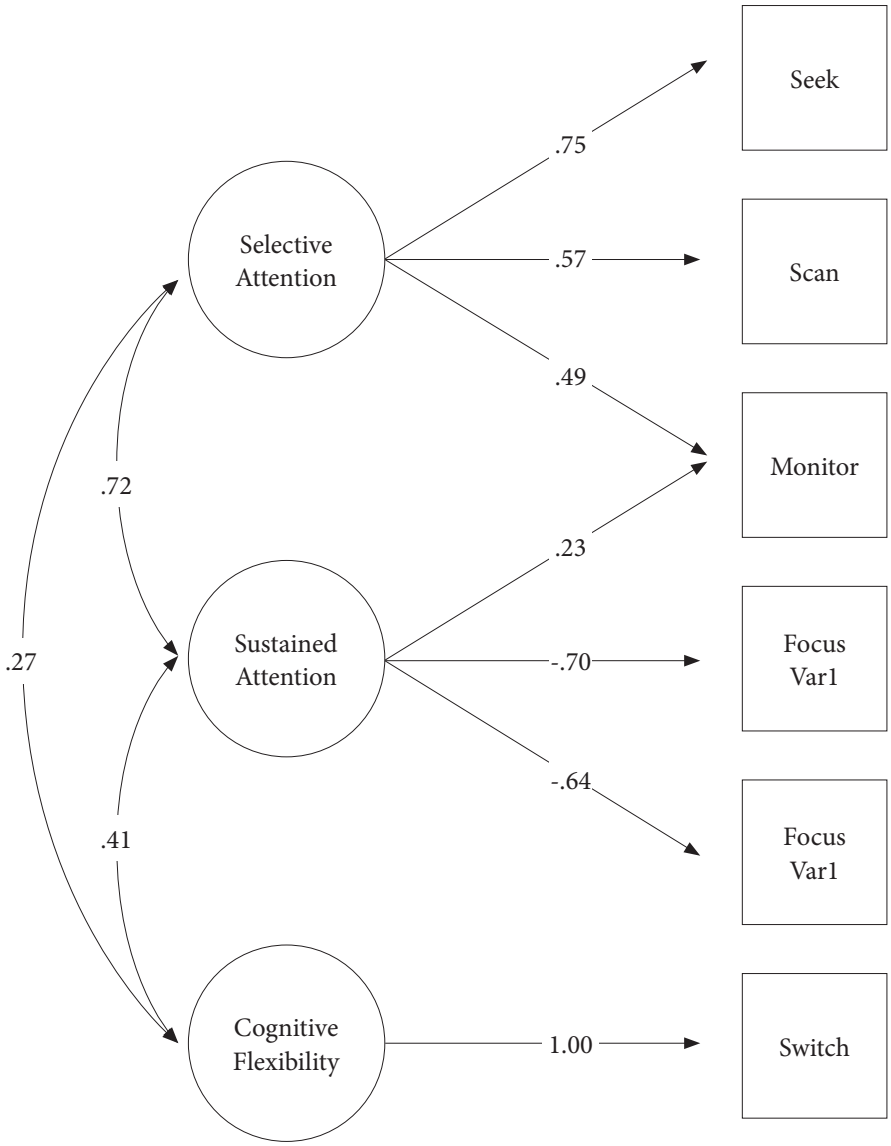
[1]Only a summary of the comprehensive data analysis is presented in this document. A separate full report with all details of this analysis authored by Dr Timothy Bednall (Swinburne University of Technology) is available upon request.

## 5. Results (continued)

### 5.1 Construct Validity (continued)

#### 5.1.1 Model 5: Adjusted Hypothesised Model

Factor analysis showed a better model fit for a Three-Factor model ($\chi^2$(df = 6) = 5.569, p = .473, CFI = 1.000, RMSEA = .000, SRMR = .016) than a General (One-Factor) model ($\chi^2$(df = 27) = 157.675, p = .000, CFI = .797, RMSEA = .119, SRMR = .076). In the Three-Factor model, the Monitor subtask was allowed to cross-load on both the Sustained and Selective attention factors, as hypothetically it could be regarded as both a sustained and selective attention task.




Furthermore, the internal consistency of each factor in the Three-Factor model was acceptable. For Selective Attention, coefficient omega was equal to .63. For Sustained Attention, it was equal to .55. However, it should be observed that this is likely because of the cross-loading of "Monitor", with each latent variable only responsible for a smaller portion of the variance. The factor loadings of the observed variables uniquely associated with each variable were generally strong.

Discriminant validity between the three factors was also promising. The highest correlation was between Selective and Sustained Attention (0.72). These findings suggest that each factor is measuring a unique construct.

## 5. Results (continued)

### 5.1 Construct Validity (continued)

#### 5.1.2 Model 5 with Control Variables (age, gender and motor-speed)

The effects of age, gender and speed task performance on the three latent variables were examined. Age was positively associated with performance on selective attention ($\beta = .57$, $p < .001$), sustained attention ($\beta = .44$, $p < .001$) and cognitive flexibility ($\beta = .23$, $p < .001$). Females performed slightly better on the cognitive flexibility task ($\beta = .14$, $p = .008$), but gender did not appear to influence performance on the selective or sustained attention tasks. Performance on the speed task (reaction time) was negatively associated with selective attention ($\beta = -.35$, $p < .001$), sustained attention ($\beta = -.29$, $p < .001$), but not cognitive flexibility ($\beta = -.08$, $p = .172$). After the inclusion of the control variables, the correlations between the constructs became smaller. This is likely because each control variable accounted for a substantial proportion of the shared variance between the latent variables.

### 5.2 Reliability

In the next stage, the test-retest reliability (i.e., the correlation between time 1 and time 2) of each factor from Model 5 was assessed.

#### 5.2.1 Measurement Invariance

Before further analyses were undertaken, measurement invariance of Model 5 factor structure was assessed (Widaman, Ferrer, & Conger, 2010). Measurement invariance testing ensures that a factor model is measuring the same construct at Times 1 and 2. Four increasingly restrictive models are tested and compared using model fit statistics. If model fit does not become significantly worse, this outcome provides evidence that measurement invariance assumptions have been met. The chi-square difference test is used to evaluate changes in fit between each model.

The four models included:

1. **Configural invariance model.** A configural invariance model is a measurement model that replicates exactly the same factor structure at Times 1 and 2. All parameters, including factor loadings, intercepts and residual variances are freely estimated.
2. **Weak (metric) invariance model.** The metric invariance model is the same as the Configural model, except corresponding factor loadings are constrained to be equal across Times 1 and 2.
3. **Strong (scalar) invariance model.** The scalar invariance model is the same at the previous model, except the corresponding observed variable intercepts are constrained to be equal across Times 1 and 2. In addition, the factor means at Time 2 are freely estimated (they are presumed to be 0 at Time 1). Scalar invariance is the minimum level required to ensure measurement invariance across time.
4. **Strict invariance model.** The strict invariance model is the same as the previous model, with corresponding residual variances constrained to be equal across Times 1 and 2.

## 5. Results (continued)

### 5.2 Reliability (continued)

#### 5.2.1 Measurement Invariance (continued)

The results of the invariance test are presented below:

| | Configural Invariance | Metric Invariance | Scalar Invariance | Strict Invariance |
|---|---|---|---|---|
| $\chi^2$ | 45.541 | 48.293 | 49.975 | 65.918 |
| DF | 34 | 38 | 41 | 46 |
| *p* | .089 | .122 | .159 | .029 |
| Scaling Factor | 0.953 | 0.987 | 0.989 | 1.021 |
| | | | | |
| $\Delta\chi^2$ | -- | 3.338 | 1.738 | 13.944 |
| ΔDF | -- | 4 | 3 | 5 |
| ΔScaling Factor | | 1.273 | 1.016 | 1.281 |
| Δ*p* | | .503 | .629 | .016 |

As indicated by the non-significant Δp values, the results indicate that scalar invariance was established. This finding indicates that it is possible to compare factor estimates at Times 1 and Times 2, as the measurement properties (factor loadings and observed variable intercepts) do not change over time.

The strict invariance model provided a significantly worse fit to the data. This means that the residual variances of each observed variable were different over time. Inspection of the residual variances indicated that they were generally smaller at Time 2, suggesting a compression in the range of performance. This compression may be due to dropouts of children with performance scores that deviate substantially from the average score. Alternatively, improvement in performance over time may lead to some children obtaining perfect scores at Time 2 on some subtests, thereby compressing the range of values.

#### 5.2.2 Test-Retest Reliability

Based on the scalar invariance model, correlations among the latent variables at Times 1 and 2 are presented in the table below. Test-retest reliability is highlighted in yellow.

*Correlations among Selective Attention, Sustained Attention and Cognitive Flexibility*

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| **Time 1** | | | | | | |
| 1. Selective Attention | -- | | | | | |
| 2. Sustained Attention | .74*** | -- | | | | |
| 3. Cognitive Flexibility | .26*** | .41*** | -- | | | |
| **Time 2** | | | | | | |
| 4. Selective Attention | .97*** | .72*** | .34*** | -- | | |
| 5. Sustained Attention | .57*** | .76*** | .47*** | .62*** | -- | |
| 6. Cognitive Flexibility | .32*** | .38*** | .45*** | .31*** | .35*** | -- |
| Mean | .00 | .00 | .00 | .47 | -.15 | .11 |
| SD | 11.15 | .74 | 3.27 | 14.02 | .71 | 3.00 |

Note. *$p < .05$, **$p < .01$, *** $p < .001$

## 5. Results (continued)

### 5.2.2 Test-Retest Reliability

As can be seen in the table, both selective and sustained attention showed very high test-retest reliability over time. In contrast, the cognitive flexibility showed lower reliability. A high reliability in this context would normally be considered at around 0.7 or higher and an intermediate reliability at around 0.6 (Fan et al. 2002).

### 5.3 Convergent validity

### 5.3.1 Correlation between observed variables from DETECT and TEA-Ch2J

Convergent validity can be assessed by correlations between DETECT and a well-established measure of the same construct (attention). Evaluating convergent validity in young children (ages 3–7 years) was challenging because of the scarcity of specific gold standard performance-based measures of targeted constructs appropriate for these ages. For the purpose of this clinical trial, TEA-Ch2J was used which includes subtests that measure selective attention and sustained attention in 5- to 7-year-olds (but not Cognitive Flexibility). Given that TEA-Ch2J composite scores (selective and sustained attention indexes) couldn't be collected through this clinical trial, each DETECT and TEA-Ch2J subtest scores were compared. This test has five activities which are theoretically similar to five of TALi DETECT subtests[2].

Any correlation coefficient greater than 0.3 is sufficient to provide evidence that the two tasks share common variance and are similar (Carlozzi et al. 2017). Correlation of interest are between:

- Speed and Simple Reaction Time
- Seek and Balloon Hunt; Seek and Balloons 5
- Scan and Hide and Seek Visual
- Monitor and Balloon Hunt; Monitor and Balloons 5; Monitor and Hide and Seek Visual
- Focus and SART

| | Balloon Hunt | Balloons 5 | Hide and Seek Visual | Hide and Seek Auditory | Simple Reaction Time | SART |
|---|---|---|---|---|---|---|
| Speed | -0.40** | 0.39** | -0.32** | 0.17 | 0.48** | -0.08 |
| Seek | 0.68** | -0.62** | 0.51** | -0.20 | -0.54** | -0.01 |
| Scan | 0.21* | -0.26** | 0.29** | -0.15 | -0.22* | -0.15 |
| Monitor | -0.50** | 0.49** | -0.42** | 0.21 | 0.55** | 0.08 |
| Focus | 0.07 | 0.05 | -0.06 | 0.14 | 0.04 | 0.31** |
| Switch | -0.17 | 0.17 | -0.07 | 0.05 | 0.12 | 0.16 |

Most subtests showed satisfactory correlation with TEA-Ch2J subtests. Future studies will include more comprehensive measures of cognitive attention to enable establishing convergent validity of latent variables (selective attention, sustained attention and executive attention/cognitive flexibility).

[2] The important variable of SSRT (stop-signal reaction time) couldn't be estimated from the Inhibit (SST) subtest of DETECT due to low number of stop levels. Therefore, this subtest was excluded from the final analysis. Furthermore, please note that the DETECT subset of Switch (a measure of Cognitive Flexibility) doesn't have an equivalent subtest in TEA-Ch2J, therefore a convergent validity analysis for this subtest wasn't feasible.

### 5.4 Developmental Sensitivity

Associations of test scores with age reflect the validity of the tests for measuring cognitive development during childhood. DETECT has good developmental sensitivity. The effects of age on the three latent variables were examined. Age was positively associated with performance on selective attention ($\beta = .57$, $p < .001$), sustained attention ($\beta = .44$, $p < .001$) and cognitive flexibility ($\beta = .23$, $p < .001$).

## 6. Conclusion

A comprehensive data analysis showed that TALi DETECT has an excellent factor structure which simply indicates that based on the participants who were tested on the TALi DETECT, there appears to be 3 main distinct factors that are being assessed across the 6 DETECT subtests.

There was sufficient differentiation between Sustained Attention and Selective Attention factors and much better differentiation between Sustained Attention and Cognitive Flexibility and Selective Attention and Cognitive Flexibility factors.

Two constructs, Sustained Attention and Selective Attention, showed acceptable test-retest reliability (.76 and .97 respectively). However, the third factor (Cognitive Flexibility) showed fairly low test-retest reliability (.45). A slightly lower figure also could be argued for children as they generally show much higher unreliability or variance. However, TALi has commenced development projects to improve the test-retest reliability for the Cognitive Flexibility factor.

The convergent validity of TALi DETECT was in a satisfactory range for the subtests that had direct equivalent in TEA-Ch2J. However, future studies will use additional various performance-based assessments of cognitive attention for further establishing DETECT's convergent validity at observed and latent variable level.

The latent factor of Cognitive Flexibility is currently estimated using only one indicator. Therefore, TALi is prioritising the addition of another cognitive task measuring the same construct (Executive Function/Cognitive Flexibility) to the DETECT battery of tests.

The insights acquired through this large-scale validation trial enables TALi to develop a refined version of DETECT which will be tested in large neurotypical and clinical paediatric populations.

## 7. References


- Asparouhov, T., & Muthén, B. (2009). Exploratory Structural Equation Modeling. Structural Equation Modeling: A Multidisciplinary Journal, 16(3), 397–438. https://doi.org/10.1080/10705510903008204
- Carlozzi, Noelle E. Tulsky, David S. Wolf, Timothy J. Goodnight, Siera Heaton, Robert K. Casaletto, Kaitlin B. Wong, Alex W. K. Baum, Carolyn M. Gershon, Richard C. Heinemann, Allen W. (2017). Construct validity of the NIH Toolbox Cognition Battery in individuals with stroke. Rehabilitation Psychology, 62(4), 443–454.
- Fan, J., Bruce D. McCandliss, Tobias Sommer, Amir Raz, and Michael I. Posner (2002) Testing the Efficiency and Independence of Attentional Networks; *Journal of Cognitive neuroscience* 14:3, 340-347
- Manly, T., Anderson, V., Nimmo-Smith, I., Turner, A., Watson, P., & Robertson, I. H. (2001). The Differential Assessment of Children's Attention: The Test of Everyday Attention for Children (TEA-Ch), Normative Sample and ADHD Performance. Journal of Child Psychology and Psychiatry, 42(8), 1065–1081. https://doi.org/10.1111/1469-7610.00806
- Muthén, L. K., & Muthén, B. O. (2017). Mplus User's Guide (8th ed.). Los Angeles, CA: Muthén & Muthén.
- West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In Handbook of structural equation modeling. (pp. 209–231). New York, NY, US: The Guilford Press.
- Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial Invariance Within Longitudinal Structural Equation Models: Measuring the Same Construct Across Time. Child Development Perspectives, 4(1), 10–18. https://doi.org/10.1111/j.1750-8606.2009.00110.x
- Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): a method of assessing executive function in children. Nature Protocol.