Validation and Normalization of the Tower of London-Drexel University Test 2nd Edition in an Adult Population with Intellectual Disability

Javier García-Alba[1], Susanna Esteba-Castillo[2], Miguel Ángel Castellanos López[1],
Emili Rodríguez Hidalgo[2], Nuria Ribas Vidal[2], Fernando Moldenhauer Díaz[3] and Ramón Novell-Alsina[2]

[1] *Universidad Complutense (Spain)*
[2] *Parc Hospitalari Martí i Julià (Spain)*
[3] *Hospital Universitario La Princesa (Spain)*

Abstract. Despite how important it is to assess executive functioning in persons with Intellectual Disability (ID), instruments adapted and validated for this population are scarce. This study's primary goal was to find evidence for the validity of the ID version of the Tower of London (TOL^{DXtm}) test in persons with mild (ID$_{Mi}$) and moderate (ID$_{Mo}$) levels of ID with Down Syndrome (DS). A multicenter study was carried out. Subjects ($n = 63$, ≥ 39 years old) had DS with mild ($n = 39$) or moderate ID ($n = 24$) with no minor neurocognitive disorder or Alzheimer's disease. Assessment protocol: TOL^{DXtm} for ID, Kaufman Brief Intelligence Test Second Edition (K-BIT II), Cambridge Examination for Mental Disorders of Older People with Down's Syndrome and Others with Intellectual Disabilities (CAMDEX-DS), Weigl's Color-Form Sorting Test (WCFST), Barcelona Test for Intellectual Disability (BT-ID), and the Behavior Rating Inventory of Executive Function (BRIEF-P). The internal consistency (ID$_{Mi}$ and ID$_{Mo}$), factor structure of the different subscales, and relationship between TOL^{DXtm} subscales and other cognitive measures (BT-ID, WCFST, and BRIEF-P) were analyzed. A normative data table with ID population quartiles is provided. TOL^{DXtm} for ID showed a robust one factor structure and coherent associations with other, related neuropsychological instruments. Significant differences between ID$_{Mi}$ and ID$_{Mo}$ on movement-related variables like *Correct (Corr; p = .002)* and *Moves (Mov; p = .042)* were observed, along with good internal consistency values, Corr ($\alpha = .75$), Mov ($\alpha = .52$). Regarding internal consistency, no between-groups differences were observed (all *p*-value > 0.05). The TOL^{DXtm} for ID is thus an instrument, supported by good validity evidence, to evaluate problem-solving and planning in ID. It distinguishes between individuals with mild and moderate ID, and is highly associated with other measures of executive functioning.

Received 24 February 2016; Revised 7 June 2017; Accepted 9 June 2017

Keywords: down syndrome, executive functions, intellectual disability, Tower of London, validity.

The Tower of London (TOL) test (Shallice, 1982) is a neuropsychological instrument that evaluates difficulties with planning and non-verbal problem solving, which are associated with frontal lobe dysfunction, especially in the right dorsolateral prefrontal cortex, cingulate cortex, and basal ganglia (Tirapu, García, Luna, & Periañez, 2012). The planning sequence begins with proposing an objective, mentally rehearsing, applying one's chosen strategy, and finally appraising whether or not the objective was acheived (Tirapu-Ustárroz, Muñoz-Céspedes, Pelegrín-Valero, & Albéniz-Ferreras, 2005).

Shallice (1982) created the original TOL. Most versions utilize two peg boards, each with three vertical pegs arranged with several colored beads. However, some test variants employ other physical objects, like the Monsters and Globes task (Kotovsky, Hayes, & Simon, 1985) and computerized versions like the Cambridge Neuropsychological Test Automated Battery (CANTAB) (Owen, Downes, Sahakian, Polkey, & Robbins, 1990).

The latest version of the TOL was created at Drexel University; as such, it is called the Tower of London-Drexel University 2nd Edition (TOL^{DXtm}) (Culbertson & Zillmer, 2001). It includes a children's version (7 to 15 years old) and an adult version (16 and up). Beginning with the TOL^{DXtm} test, the authors proposed a modification for people with intellectual disability (ID) (Culbertson & Zillmer, 2001) (Figure 2). The present study's objective was to gather evidence for the validity of that version.

Due to the TOL's sensitivity to executive function deficits, numerous studies have deployed it to study different populations: subjects with dementia (Paula, Neves, Levy, Nassif, & Malloy-Diniz, 2012), traumatic brain injury (Cockburn, 1995), patients with schizophrenia (Zhu et al., 2010), people with addictive behaviors (Davydov &

Polunina, 2004), and children with focal frontal brain lesions (Jacobs & Anderson, 2002).

The TOL[DXtm] has various adaptations: for Alzheimer's type dementia (Rainville et al., 2002) and Parkinson's disease (Culbertson, Moberg, Duda, Stern, & Weintraub, 2004). The Neuronorma project has provided normative data in Spanish adult (Peña-Casanova et al., 2009) and young adult (Rognoni et al., 2013) populations.

As for populations with ID, the TOL has been the instrument of choice to study the relation between motor performance and executive functioning (Hartman, Wouwen, Scherder, & Visscher, 2010) as part of cognitive testing of children and adolescents with autism spectrum disorders (Robinson, Goddard, Dritschel, Wisley, & Howlin, 2009), and adults with fragile X syndrome (Moore et al., 2004), Prader-Willi syndrome (Walley & Donaldson, 2005), and Down syndrome (Ball, Holland, Treppner, Watson, & Huppert, 2008).

Though it is considered a standard in evaluating processes like planning, the TOL has certain limitations. A ceiling effect has been observed in young subjects, depending on the scoring method used (Berg & Byrd, 2002). Also, its task complexity makes it unsuitable for populations with cognitive deficit or ID unless it is adapted. Furthermore, the process of administering it needs to be standardized, because variations in the instructions provided, assistance, and learning processes all influence test outcomes (Unterrainer, Rahm, Leonhart, Ruff, & Halsband, 2003). If TOL-type tests had simpler tasks and more flexible rules, it would increase efficacy in the diagnostic process and subsequent implementation of cognitive intervention programs (Ball et al., 2008).

Instruments administered as part of any neuropsychological assessment should draw on normative data from the reference population, and people with ID are no exception. They have specific cognitive needs, and require tests in keeping with their intellectual potential. Though there is a growing variety of neuropsychological tests for the adult population with ID, instruments still need to be adapted for cognitive tests to meet reliability and validity standards. Normative data in the Spanish adult population with ID have not yet been published for any version of the TOL. With that in mind, the present study's objective was to examine reliability and validity evidence for the TOL[DXtm] in the adult population with ID.

The study of the psychometric properties of the TOL[DXtm] using theoretical neuropsychological models (Tirapu et al., 2012), its functioning in different populations, and its precision in the cognitive process it measures should reveal a robust structure and strong relation with other neuropsychological tests, especially tests that measure executive functions related to or involved in planning. Yet not all tests evaluate executive functions via the subject. Some measure executive functions based on parent/teacher reporting, and it would stand to reason that data collected through self-report, from the patients themselves, would differ in certain ways. For that reason, based on the tests employed in this study, we expect to find various factors, made up of variables from executive function measures taken from relatives, neuropsychological variables from strict executive functioning measures administered to patients, and a factor made up of the TOL's weakest predictor variables. Even though we expect to find results partly consistent with the above in light of the literature (Culbertson & Zillmer, 2001), we cannot overlook the type of population utilized, with Down syndrome, so we also anticipate differences from findings in other populations.

Method

Participants

The sample was comprised of 63 adult participants (≥ 39 years; 33 men and 30 women) with Down syndrome (DS) resulting from trisomy 21 and confirmed by karyotype, and mild (ID_{Mi}) and moderate (ID_{Mo}) levels of ID according to DSM-5 criteria (ID_{Mi} $n = 39$, ID_{Mo} $n = 24$). Participants were selected from hospital units: the Adult Down Syndrome Unit (ADSU) in Internal Medicine at La Princesa University Hospital (Madrid); and the Specialized Mental Health Unit for Adults with ID (SMHU-ID) at Martí i Julìa Park Hospital (Girona). The units provided lists of possible candidates from among their patients, then participants were selected through simple random sampling. The research protocol was approved by the Research Ethics Committee at each institution. All participants (subjects and their guardians) signed the appropriate informed assent forms (subjects) or guardian consent forms.

All subjects had to meet the following inclusion criteria, and none of the exclusion criteria.

Inclusion criteria

People with DS confirmed by karyotype (including mosaicisms and translocations); aged ≥ 39 years with ID_{Mi} and/or ID_{Mo} according to DSM-5 criteria; with no symptomatology of minor neurocognitive disorder (DSM-5) or Alzheimer's disease according to the dementia criteria established in CAMDEX-DS (which includes diagnostic criteria from the DSM-IV and ICD-10); receiving no prescription drug treatment that could interfere with the objectivity of data collected in clinical assessments.

Exclusion criteria

Patients with clinical hypo/hyperthyroidism or uncontrolled B_9 or B_{12} hypovitaminosis; altered consciousness

(delirium); severe uncorrected sensory alteration (auditory/visual) that would impede proper test completion; minor or major neurocognitive disorder; and not providing informed consent/assent in writing.

Materials

Kaufman Brief Intelligence Test, Second Edition (KBIT-2; Kaufman & Kaufman, 2004)

This instrument yields an intellectual quotient. The test is widely utilized for research in people with ID because it gives a standard base score of 40. Reliability coefficients (matrices, $\alpha = .83$; verbal knowledge, $\alpha = .87$). The KBIT-2 has not been translated into Spanish. This study only utilized the Matrices part, which has no verbal component, and as its authors stipulate in the manual, it can be used without the Verbal Knowledge component to obtain an IQ.

Adaptative Behavior Scale-Residential and Comunity-second edition (ABS-RC:2; Nihira, Leland, & Lambert, 1993)

This scale evaluates adaptive behavior. Internal consistency (part one, adaptive behavior, $\alpha = .91$; part two, 'problem' behaviors, $\alpha = .80$). This study utilized the Spanish-language version by Medina-Gómez and García-Alonso (2010).

Cambridge Examination for Mental Disorders of Older People with Down's Syndrome and Others with Intellectual Disabilities (CAMDEX-DS; Esteba-Castillo, Novell, Vilá, & Ribas, 2014)

This diagnostic instrument detects dementia in people with DS and other causes of ID, internal consistency ($\alpha = .93$).

Barcelona Test-Intellectual Disability (BT-ID; Esteba-Castillo, Caixas, Deus, & Peña-Casanova, 2015)

This neuropsychological test battery was adapted and validated in Spanish for adults with ID. Of its constituent subtests, this study utilized the following: Inverse Digits (working memory), Planning and Organization, Resisting Interference, Verbal Execution, and Semantic Fluency with Animals. Internal consistency by area (α orientation $= .87$; α attention $= .85$; α working memory $= .91$; α language $= .96$; α praxis $= .87$; α memory $= .74$; α executive functions $= .85$; α visual-construction $= .83$).

Weigl Color Form Sorting Test (WCFST; Goldstein & Scheerer, 1941)

This assesses the ability to categorize across two dimensions: it means ignoring a dominant dimension (color) to categorize based on a second, less dominant dimension (shape). It allows experimenters to offer external clues to see if performance improves. The WCFST has demonstrated sensitivity in detecting brain damage, and it correlates strongly with other tests of frontal performance and cognitive deficit (MMSE $r = .67$, $p < .0001$; total CAMCOG-R $r = .72$, $p < .0001$; CAMCOG-R executive functions subtests $r = .65$, $p < .0001$), showing it can help diagnose cases of cognitive alteration (Hobson, Meara, & Taylor, 2007). The WCFST does not have a verbal component, so the original version was utilized, applying Strauss and Lewin's (1982) criteria for test administration and correction.

Behavior Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2000)

In this case, we used the parent form, the BRIEF-Parents (BRIEF-P), specifically a Spanish translation edited by the original authors (Gioia, Isquith, Guy, & Kenworthy, 2016). This interview with parents evaluates eight domains of executive functions: Inhibit, Shift, Emotional Control, Initiate, Working Memory, Plan/organize, Organization of Materials, and Monitor." Internal consistency (teacher form, $\alpha = .80$; parent form, $\alpha = .98$), test-retest reliability (teacher form, $r = .88$; parent form, $r = .82$).

Tower of London-Drexel University: 2nd Edition (TOLDXtm; Culbertson & Zillmer, 2001)

We utilized the TOLDXtm adaptation for people with ID. It consists of two wooden peg boards (one for the patient, one for the examiner) with three vertical pegs of differing length, and three beads (one red, one green, and one blue). The maximum number of moves per item or problem is 20, and the maximum time is 120 seconds. The main difference between the two versions of the TOLDXtm (for children, adults) and the version of the TOLDXtm for ID is the minimum number of moves per problem, and their complexity. The minimum number of moves shifts from 3 to 7 on the children's version (7–15 years old), and 4 to 7 on the adult version (≥ 16 years old). In the version for people with ID, meanwhile, the minimum number of moves shifts from 3 to 4. That noticeably simplifies task execution. It is not until the seventh item that examinees first encounter a problem with a minimum of 4 moves, so the first six problems provide training. Test-retest reliability coefficients: total moves ($r = .81$, $p < .005$); total time violations ($r = .79$, $p < .005$); total rule violations ($r = .42$, $p < .05$). Sensitivity and specificity indices, .76 and .81, respectively.

This test collects the following measures and scales, as a function of moves, time, and violations of the instructions: *a)* correct (Corr): number of problems solved with the minimum number of moves specified; *b)* total moves (Mov): the sum of excess moves used on each item. A *move* was understood as taking a bead off

a peg and placing it on another, either at the bottom, or on top of another bead; *c)* initiation time (IT): the time lapsed between the signal to start executing, and making the first move; *d)* total time (TT): the time lapsed between initation and stopping the stopwatch; *e)* execution time (ET): difference between total time and initiation time; *f)* violation type I (VTI): times when the maximum beads per peg was exceeded; *g)* violation type II (VTII): number of times more than one bead was moved in a single turn; *h)* number of times examinee had to be reminded of the premise (PV).

The original test presents a scoresheet on which to report test performance: total correct, total moves, total initiation time, total execution time, total time, and total type I and type II violations. The test is over when in two consecutive problems, more than 20 moves are made or the maximum time is exceeded.

Relation between the tests utilized

The subtests and tests chosen to gather evidence of validity involve tasks related to the executive functions (Esteba-Castillo & García-Alba, 2015): the WCFST requires set shifting, which is the ability to change strategy; the BRIEF-P gives two dysexecutive indexes, one cognitive (inhibit, shift, emotional control, initiate, and working memory) and one behavioral (plan/organize, organization of materials, and monitor); the subtests (of the executive component) of the BT-ID require one to summon processes such as working memory, planning and organization, resisting interference, and verbal fluency.

Procedure

Subjects who met all the inclusion criteria and none of the exclusion criteria took part in the study. They were administered all the tests described above. First, their ID levels were determined (K-BIT II, ABS-RC: 2). Then, we determined if cognitive deficit resulting from dementia was present or absent (CAMDEX-DS and CAMCOG). Last, the neuropsychological test batteries were administered (BT-ID, BRIEF-P, WCFST, and TOL^{DXtm}). Assessments were conducted by psychologists specialized in the neuropsychology of ID, in rooms at the SMHU-ID at Martí i Julià Park Hospital (Girona) and the Hermanos García Noblejas Medical Specialty Center (La Princesa University Hospital, Madrid). In Madrid as well as Girona, subjects were always evaluated by the same professional, in both cases by staff highly specialized in the neuropsychological assessment of ID.

Data analysis

We computed descriptive statistics for all sociodemographic variables, and for the scales of the TOL^{DXtm}.

In addition to the eight scales proposed by Culbertson and Zillmer (2001) and discussed in the instruments section, we added a new variable called *Hits.* For its sheer simplicity, the authors decided it would be interesting to include in the study. This variable is to the number of problems correctly solved out of the ten on the TOL^{DXtm}, regardless of the number of moves used. This variable was included in our analyses.

The descriptive data tables provide means, standard deviations, indicators of skewness and kurtosis (calculated as b_1 and b_2), and corrected item-total correlations on each measure in the 10 problems, for the sample all together as well as separately by group, ID_{Mi} and ID_{Mo}. We ran the appropriate comparisons between groups; in this case, since scores were not normally distributed, we opted for the Mann-Whitney test.

Internal consistency and factor structure

For all TOL^{DXtm} scales, in the total sample as well as the two subgroups, we determined internal consistency using Cronbach's alpha coefficient. We also compared the equality of alpha coefficients for independent groups (Feldt, Woodruff, & Salih, 1987) using the groups in this study (ID_{Mi} and ID_{Mo}). The factorial composition of the test's 10 problems was examined through Exploratory Factor Analysis. For each scale, we calculated the number of optimal factors to extract according to classic parallel and MAP criteria, which are widely used in psychometric test validation. The results include goodness of fit indicators (RMSEA, RMSR, and TLI) obtained by extracting a single factor, for each scale. The Ordinary Least Squares (OLS) method of factor extraction was used throughout, and we worked from Pearson correlation matrices. For the variables Hits and Corr, since they were dichotomous, we worked with tetrachoric correlation matrices.

Relationship between the TOL^{DXtm} and other neuropsychological tests

We examined relations between the TOL^{DXtm} scales and other cognitive measures through Exploratory Factor Analysis. We worked with the Pearson correlation matrix, calculated using total scores on all the scales and tests in the study. As above, we opted for OLS estimation and Promax rotation. The number of factors to extract was determined by the aforementioned parallel, MAP, and VSS algorithms. The results section presents the pattern matrix, correlations between factors, and the following goodness of fit indexes: RMSEA, RMSR, and TLI.

Normative data table

To provide a possible clinical interpretation of test scores, a frequency table is included, with each of the

nine scales of the TOL^{DXtm} represented. Given the sample available, we chose to present quartiles only.

All analyses were conducted using *R* statistical software with the *psych* and *cocron* packages.

Results

Difficulty and consistency of scales on the test

Though TOL items were constructed to progressively increase in difficulty, that was not apparent in the results (see Figure 1 and Appendices A through I). Instead we observed a staggered structure in which difficulty decreased from the first item to the second, increased in problems 3 and 4, then dipped again only to increase on item 7, and again on 10. This behavior was more or less generalized to all scales, but was less pronounced on Hits, which showed a more gradual rise in difficulty.

The meaning of these shifts, probably related to the cognitive components of each task, is explored in the Discussion section.

In general, and especially on the violations scales, there were notable discrepancies between indices of skewness and kurtosis. That was due to small sample size, and to the heterogeneity of patients with DS. With those considerations in mind, we elected to use non-parametric tests.

Total scores on the three movement-related scales (Corr, Mov, and Hits) showed significant differences in score between the ID_{Mi} and ID_{Mo} groups (Table 1), with the ID_{Mo} group performing worse across the board: Corr ($p < .05$), Mov ($p < .05$), and Hits ($p < .01$). Its values of internal consistency, gauged by Cronbach's alpha, were found to be suitable, with the highest value found on the Hits scale ($\alpha = .89$), followed by Corr ($\alpha = .75$), and Mov having the lowest of the three ($\alpha = .52$). Tests of the equality of alpha did not suggest differences between the Mild and Moderate groups on any scale, showing the test to be equally reliable in the two groups: Mov ($p = .216$), Corr ($p = .066$), and Hits ($p = .159$).

The time scales (initiation, execution, and total) had similar or greater values of internal consistency than previous scales (Table 2): initiation ($\alpha = .84$), execution ($\alpha = .75$), and total ($\alpha = .77$). Meanwhile, though times were longer in the Moderate group than the Mild
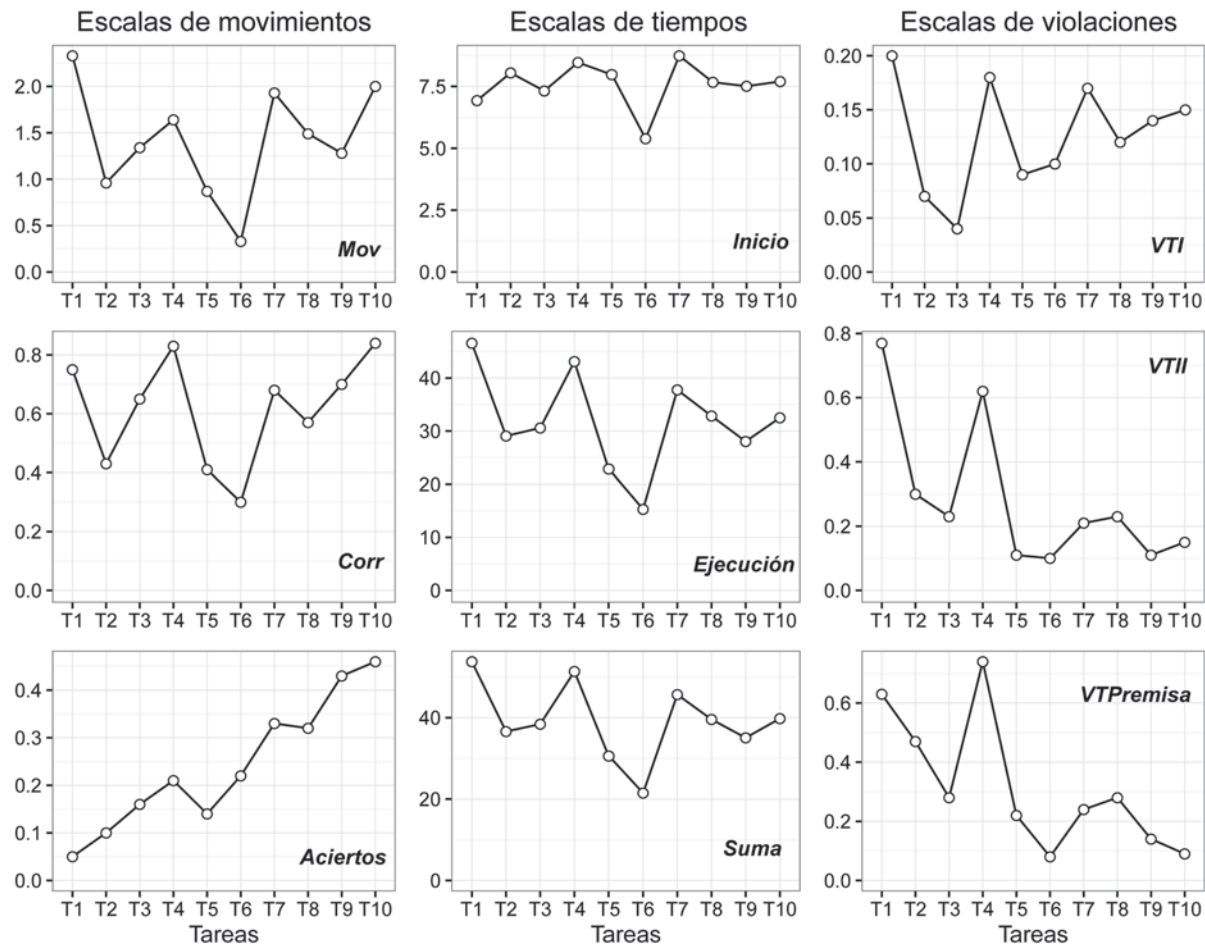
Figure 1. Graph of mean scores on the scale's 10 tasks in terms of moves, time, and violations.

On all graphs, higher values indicate higher task difficulty; on the scales Corr and Hits, the proportion of misses is presented.
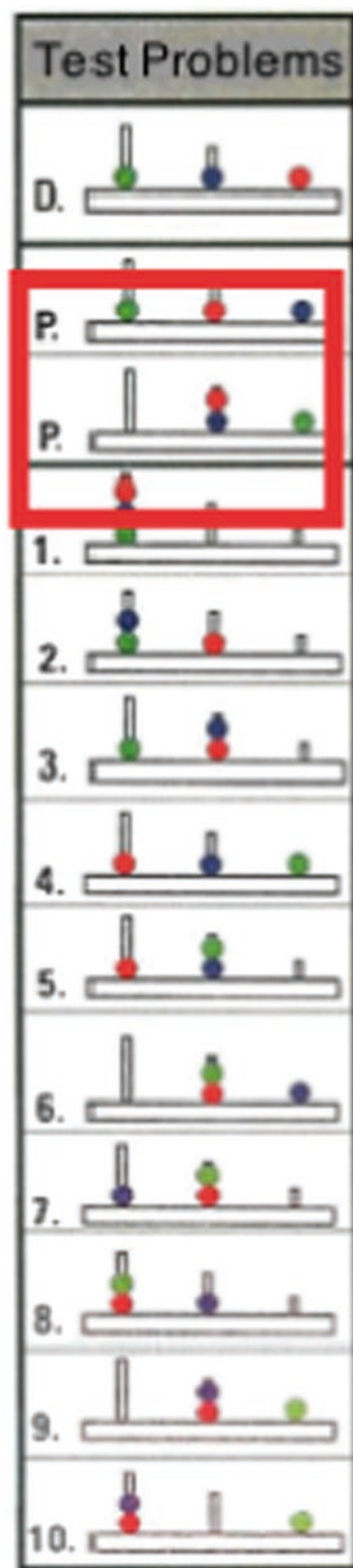
Figure 2. TOL^DX: 2nd Edition-Test Items for Mentally Challenged Populations. Adapted from Culbertson & Zillmer (2001).

group, no statistically significant differences were detected: initiation ($p = .147$), execution ($p = .475$), and total ($p = .581$). Initiation time did reveal significant differences between groups in terms of alpha values – Mild ($\alpha = .73$), Moderate ($\alpha = .91$), $p < .01$ – but that is not especially worrisome since their values were still high. On the other two scales, differences in alpha were not found between groups; execution ($p = .150$), total ($p = .200$).

The third set of scales, violations, are peculiar in that violations decrease as one advances through problems (as in a Pareto distribution). These generally presented poorer psychometric properties than the scales above (Table 3). Their alpha indexes were lower: VTI ($\alpha = .42$), VTII ($\alpha = .70$), and PV ($\alpha = .67$); and none yielded between-groups differences: VTI ($p = .670$), VTII ($p = .311$), and PV ($p = .406$). Some alpha values could not be calculated for lack of subjects with data.

The test's factor structure

In terms of factor structure, the TOL^DXtm test was built with a single dimension. Confirmatory factor analysis could not be conducted due to particularities of the sample, but Table 4 is included to present parallel and MAP algorithms and descriptive data from an exploratory factor analysis of just one factor. The MAP algorithm indeed proposed a single-factor solution for all scales; but the other algorithm yielded less uniform results, especially for the scales Mov, initiation time, and the three violations scales. Goodness of fit indicators for one-factor extraction were not very acceptable on any scale, with all values of RMSEA and RMSR above .1. Corr was the exception; it had little explained variance (24%), but it presented an RMSEA less than .05 (RMSEA = .03). The scale that explained the highest percentage of variance was Hits, at 44%, while Mov and VTI explained the least variance, at 17%.

Relation between TOL^DXtm test scales and other neuropsychological tests

Table 5 presents the sample's descriptive statistics for the remaining tests in the study, and Table 6 conveys the results of exploratory factor analysis, which was done using total scores from the nine scales of the TOL^DXtm, and said tests. The correct number of factors to retain, according to parallel and MAP criteria, was three, with 43% of variance explained and the following goodness of fit values: RMSR = .09; TLI = .057; and RMSEA = .15. Interpreting the matrix of factor loadings reveals a first factor (F1) associated with the scales of the BRIEF-P, a second factor (F2) comprised of variables from the BT-ID, WCFST, and two movement-related variables from the TOL^DXtm (Corr and Hits), and a third and final factor (F3) that combines all

Table 1. *Descriptive Statistics for Total Scores on Movement-related Scales*

	n	Mean	Sd	Sk	Ku	α
Hits						
ID All	63	7.59	2.84	−1.18	0.30	0.89
ID Mild	39	8.26	2.41	−2.11	4.34	0.85
ID Moderate	24	6.50	3.18	−0.28	−1.51	0.91
	[a]$U = 605$; $p = 0.048$*			[b]$F_{(23, 38)} = 1.66$; $p = .159$		
Corr						
ID All	63	3.84	2.54	0.04	−1.02	0.75
ID Mild	39	4.62	2.23	−0.20	−0.64	0.63
ID Moderate	24	2.58	2.55	0.72	−0.58	0.81
	[a]$U = 683$; $p = .002$**			[b]$F_{(23, 38)} = 1.94$; $p = .066$		
Mov						
ID All	63	10.63	8.07	1.04	0.86	0.52
ID Mild	39	10.13	7.32	0.67	−0.41	0.39
ID Moderate	24	11.46	9.27	1.16	0.82	0.61
	[a]$U = 440.5$; $p = .042$*			[b]$F_{(23, 38)} = 1.56$; $p = .216$		

Sd: Standard deviation; Sk: Index of skewness; Ku: Index of kurtosis; alpha: Value of Cronbach's α; [a] Mann-Whitney's contrast for ID Mild and ID Moderate; [b] Test for equality of alpha coefficients between ID Mild and ID Moderate.

Table 2. *Descriptive Statistics for Time Scales*

	n	Mean	Sd	Sk	Ku	α
Initiation time						
ID All	63	71.40	36.36	0.17	−0.48	0.84
ID Mild	39	65.28	30.65	−0.19	−0.53	0.73
ID Moderate	24	81.33	42.95	0.05	−1.16	0.91
	[a]$U = 365$; $p = .147$			[b]$F_{(23, 38)} = 3.00$; $p = .002$**		
Execution time						
ID All	63	243.21	117.78	0.35	0.06	0.75
ID Mild	39	232.31	112.15	0.10	−0.37	0.70
ID Moderate	24	260.92	126.83	0.53	−0.06	0.82
	[a]$U = 417$; $p = .475$			[b]$F_{(23, 38)} = 1.66$; $p = .150$		
Total time						
ID All	63	299.37	131.55	0.28	0.40	0.77
ID Mild	39	289.49	122.56	−0.10	0.08	0.73
ID Moderate	24	315.42	146.28	0.54	0.01	0.83
	[a]$U = 428.5$; $p = .581$			[b]$F_{(23, 38)} = 1.58$; $p = .200$		

Sd: Standard deviation; Sk: Index of skewness; Ku: Index of kurtosis; alpha: Value of Cronbach's α; [a] Mann-Whitney's contrast for ID Mild and ID Moderate; [b] Test for equality of alpha coefficients between ID Mild and ID Moderate.

TOL^{DXtm} measures except the last two. Considering the matrix of correlations, these factors are independent of one another.

Normative data table

Table 7 was included to facilitate clinical interpretation. It presents minimum, maximum, and quartile scores on the nine scales, in the total sample as well as ID_{Mi} and ID_{Mo} groups separately. All scales (except Corr and Hits) were inverted, so the lower a subject's quartile, the worse the performance. In other words, on every scale, a subject in the first quartile (Q_1) performed worse than someone in the third (Q_3).

Discussion

This study's objective was to examine reliability and validity evidence for the version of the TOL^{DXtm} created for people with ID (Culbertson & Zillmer, 2001): 2nd edition for adults with DS. Of the many existing versions of the TOL, this is the only one designed with problem difficulty appropriate for use in people with ID.

Table 3. *Descriptives Statistics for Violation Scales*

	n	*Mean*	*Sd*	*Sk*	*Ku*	*α*
Type I violations						
ID [All]	63	0.94	1.27	1.14	0.04	0.42
ID [Mild]	39	0.82	1.10	1.16	0.35	0.57
ID [Moderate]	24	1.13	1.51	0.88	−0.90	0.73
	[a]$U = 440$; $p = .670$				[b]$F_{(23, 38)} = 1.59$; $p = .190$	
Type II violations						
ID [All]	63	2.30	2.44	1.07	0.35	0.70
ID [Mild]	39	2.15	2.46	0.87	−0.62	[c]
ID [Moderate]	24	2.54	2.45	1.36	1.49	0.75
	[a]$U = 397.5$; $p = .311$					[b] [c]
Premise violations						
ID [All]	63	2.59	3.06	0.91	−0.46	0.67
ID [Mild]	39	2.49	3.19	1.00	−0.30	[c]
ID [Moderate]	24	2.75	2.88	0.70	−1.05	[c]
	[a]$U = 411$; $p = .406$					[b] [c]

Sd: Standard deviation; Sk: Index of skewness; Ku: Index of kurtosis; alpha: Value of Cronbach's α; [a] Mann-Whitney's contrast for ID [Mild] and ID [Moderate]; [b] Test for equality of alpha coefficients between ID [Mild] and ID [Moderate]. [c] This value could not be calculated due to missing data.

Table 4. *Factor Structure of the TOL's 9 Scales. Indicators of How Many Factors (Parallel and MAP). Results on the Right were Obtained by Extracting One Factor (Explained Variance, RMSEA, RMSR, and TLI)*

					One single factor		
		parallel	Map	Var	RMSEA	RMSR	TLI
Movement-related scales	Mov	2	1	0.17	0.16	0.15	0.24
	Corr	1	1	0.24	0.03	0.08	1.00
	Hits	1	1	0.44	0.19	0.12	0.67
Time scales	IT	3	1	0.41	0.13	0.09	0.81
	ET	1	1	0.31	0.13	0.10	0.72
	TT	1	1	0.34	0.18	0.11	0.57
Violation scales	VTI	5	1	0.17	0.16	0.15	0.21
	VTII	4	1	0.22	0.16	0.14	0.45
	PV	5	1	0.28	0.24	0.16	0.32

Assessing executive functions in people with ID is very important because of how they relate to behavioral aspects, like the behavioral disorders so common in these individuals. However, professionals in this field must make an effort to correct the present dearth of tests designed specifically for this population, and lack of adaptations. To fill that gap would enable precise evaluation and correct reporting of subjects' level of functioning (Esteba-Castillo et al., 2013).

Often, evaluations are made using tests designed for the general population, and applying barometers from the population at large, by mental age, to gauge a subject's performance. That can be risky, and anyway diagnostic decisions should be made with caution (Esteba-Castillo & García-Alba, 2015). Being able to

access normative data for reference is indispensable on any neuropsychological test, and for reference data to be valid and correct, it should represent the population to which the evaluee belongs (Rognoni et al., 2013).

Generally speaking, the psychometric properties of the TOL[DXtm] version for people with ID were satisfactory on all variables. It was also noteworthy in its ability to avoid the floor effect, which tests not adapted for the ID population tend to have.

As some authors have warned in past publications, there is an age effect on TOL performance. Peña-Casanova et al. (2009) observed a clear drop in performance in adults 60 to 70 years old, and much higher performance in young people aged 20 to 30. Although there are no TOL[DXtm]-ID data available in

Table 5. *Descriptive Statistics for the Full Sample on the Study's Remaining Tests*

	n	mean	Sd	Sk	Ku
BRIEF-P:					
Emotional control	52	14.98	4.46	1.05	0.34
Shift	52	13.33	3.58	0.79	0.57
Inhibit	52	13.00	3.72	1.80	3.54
Monitor	52	13.23	3.70	−0.77	1.50
Plan/organize	52	17.46	4.86	−0.27	0.46
Working memory	52	15.58	4.08	0.99	0.56
Initiative	52	13.37	6.12	4.80	27.75
Organization of materials	52	7.17	2.20	2.11	4.53
WCFST:	55	2.29	1.95	0.30	−1.55
BT–ID:					
Resisting interference	62	14.58	11.85	0.70	−0.38
Verbal Fluency with Animals	62	8.53	3.81	0.09	−0.59
Inverse digits	63	0.81	1.26	1.32	0.91
Inverse digits – efficiency	63	0.90	1.58	2.63	9.19
Planning and organization	61	4.25	2.87	0.33	−1.26

Sd: Standard deviation; Sk: Index of skewness; Ku: Index of kurtosis.

the child and/or adolescent population with DS, we do not expect they will perform the same as adults. On that note, the general-population version has two forms, one for children and one for adults. Scores on the original test show that on the children's version (7–15 years), the various age ranges show statistically significant differences on all scales ($p < .001$) except total initiation time ($p < .20$). According to the authors, that is linked to cognitive development of processes needed to solve the problems (Culbertson & Zillmer, 2001).

This study's population has specific features belonging to a very advanced stage of development among people with DS; in many of them, it marks the beginning of cognitive decline, depending on normative aging and the onset of Alzheimer's type dementia (Nieuwenhuis-Mark, 2009). Given the high prevalence of Alzheimer's disease in people with DS (Zigman & Lott, 2007) and frontal lobe neurodegeneration, especially in certain areas (Teipel et al., 2004), it is especially important to administer a test like the TOL^{DXtm}-ID around this age. Therefore, future validations of the TOL^{DXtm}, and similar tools, should account for these development stages and their characteristics before establishing appropriate age cutoffs.

The variables in this version of the TOL^{DXtm} (correct, moves, initiation time, total time, execution time, type I violations, type II violations, premise violations) appeared on the original test, and measure different parameters – moves, times, and violations – to determine a subject's level of performance. We observed that the three parameters do not work in the same way. In general, according to internal consistency indices

and their significance across ID$_{Mi}$ and ID$_{Mo}$ groups, the variables Corr and Mov had the highest values. Total scores on Corr and Mov were significantly different in the ID$_{Mi}$ and ID$_{Mo}$ groups. This suggests the minimum and total moves made in the ten problems clearly differed according to the extent of ID. The ID$_{Mi}$ group scored higher on Corr and lower on Mov than the ID$_{Mo}$ group, both of which indicate higher performance in the ID$_{Mi}$ group. However, those two variables' internal consistency was stable. Thus, both can predict ID levels and show similar levels of internal consistency. This aspect is especially relevant because it reports this TOL version's sensitivity in detecting clearly differentiated levels of intellectual functioning. In that sense, the authors believe this adaptation is significantly suitable.

The new variable proposed, Hits, is dichotomous (was the subject able to finish the problem or not, regardless of how many moves they used) and it was observed to operate with very high levels of reliability, higher than the variables Corr and Mov. As in the case of Corr and Mov, Hits was consistent across ID levels; comparing alphas in the ID$_{Mi}$ and ID$_{Mo}$ groups produced no statistically significant differences. Outcomes on this variable suggest that as the problems become more difficult, fewer subjects can finish them; only 64% of sujects with ID$_{Mi}$ and 38% of subjects with ID$_{Mo}$ were able to complete problem 10. Nonetheless, the original test did not include Hits, so it is not counted in the final scores assessing examinee performance. With that in mind, and given its indices of reliability, stable consistency, and the information it provides about examinee performance, future versions of the test should take Hits into account.

Table 6. *Exploratory Factor Analysis Results for the Study's Variables*

	F1	F2	F3	h^2
Emotional control (BRIEF-P)	**0.79**	0.26	–0.01	0.697
Shift (BRIEF-P)	**0.78**	0.03	0.01	0.614
Inhibit (BRIEF-P)	**0.76**	–0.07	–0.08	0.593
Monitor (BRIEF-P)	**0.71**	–0.17	0.03	0.534
Plan/organize (BRIEF-P)	**0.62**	–0.07	–0.05	0.393
Working memory (BRIEF-P)	**0.52**	–0.46	0.00	0.477
Initiative (BRIEF-P)	**0.45**	–0.15	0.05	0.225
Corr (**TOL**$^{\text{DXtm}}$)	–0.18	**0.82**	–0.23	0.755
Hits (**TOL**$^{\text{DXtm}}$)	–0.17	**0.72**	0.23	0.599
Inverse digit (BT-ID)	0.12	**0.71**	–0.06	0.516
Inverse digit – efficiency (BT-ID)	–0.05	**0.63**	–0.01	0.397
WCFST	0.10	**0.49**	–0.23	0.306
Resisting interference (BT-ID)	–0.11	**0.46**	–0.17	0.255
Verbal fluency with animals (BT-ID)	–0.05	**0.40**	–0.02	0.167
Organization of materials (BRIEF-P)	0.15	**–0.28**	0.08	0.110
Planning and organization (BT-ID)	0.11	**0.21**	–0.10	0.067
ET (**TOL**$^{\text{DXtm}}$)	–0.08	0.00	**0.99**	0.995
TT (**TOL**$^{\text{DXtm}}$)	–0.14	0.05	**0.97**	0.958
Mov (**TOL**$^{\text{DXtm}}$)	0.17	–0.10	**0.60**	0.400
VTII (**TOL**$^{\text{DXtm}}$)	–0.20	–0.05	**0.43**	0.226
PV (**TOL**$^{\text{DXtm}}$)	0.10	–0.10	**0.41**	0.190
VTI (**TOL**$^{\text{DXtm}}$)	0.11	–0.19	**0.32**	0.153
IT (**TOL**$^{\text{DXtm}}$)	–0.23	–0.18	**0.31**	0.178
Var	0.15	0.14	0.13	
Cumulative var	0.15	0.29	0.43	

Goodness of fit indicators:		Correlations between factors:		
			F1	F2	F3
RMSEA	0.15	F1	1.00		
RMSR	0.09	F2	–0.09	1.00	
TLI	0.57	F3	–0.06	–0.11	1.00

Each variable's highest factor loading appears in bold; h^2 communality of each scale; Below find the variance explained by each factor, and goodness of fit indicators (RMSR, TLI, and RMSEA).

The test addresses another parameter: time (initiation time, execution time, and total time). Measures of IT across the 10 problems in both groups were stable; differences were not observed except on problem 6, but observing and comparing total IT scores, significant differences did not appear. In the ID_{Mi} group, there was a tendency for subjects, independently of item difficulty and the minimum moves needed to correctly complete it, to take a very similar amount of time to initiate. That is, they initiated the task without prior mental planning. The ID_{Mo} group, meanwhile, took longer to initiate tasks, and also displayed a similar effect as in the ID_{Mi} group: initiation time did not depend on task difficulty, and they did not seem to plan moves ahead of time. Regarding ET, differences were observed between the ID_{Mi} and ID_{Mo} groups such that the ID_{Mi} group executed tasks in far less time. However, that effect only occurred in items 1 to 5; from 6 onward, execution times were similar in the two groups. In a way, it seems the ID_{Mo} group learned over the course of problems 1–5 and, having overcome those, improved performance, gaining speed. Yet looking at total ET's across groups, significant differences were not observed. In IT as well as ET, we observed the same effect as in Corr and Mov: comparing the two groups' internal consistency on both variables, values were highly acceptable but not stable; the same level of consistency did not occur in ID_{Mi} as in ID_{Mo}. Fatigue is an important aspect to consider. Looking at time measures on the whole, it does not seem that fatigue influenced quality or ability in task execution. Execution time seemed more closely related and learning dependent, and thus more dependent on cognitive functioning level than on task difficulty or fatigue.

The third parameter refers to instances of the subject violating norms, and what type of norms. Both groups showed a clear tendency to make fewer violations as

Table 7. *Quartiles for All TOL^{DXtm} Scales, in All Subjects and in the ID$_{Mi}$ and ID$_{Mo}$ Groups*

			Worst score	Q$_1$ (25%)	Q$_2$ (50%)	Q$_3$ (75%)	Best score
Escalas de movimientos	*Mov*	All	38	15	8	4	0
		ID$_{Mi}$	28	14	8	5	0
		ID$_{Mo}$	38	16	9	4	1
	Corr	All	0	2	4	6	9
		ID$_{Mi}$	0	3	5	6	9
		ID$_{Mo}$	0	0	2	4	8
	Hits	All	0	6	9	10	10
		ID$_{Mi}$	0	8	9	10	10
		ID$_{Mo}$	1	4	7	10	10
Escalas de tiempos	*IT*	All	167	97	68	43	0
		ID$_{Mi}$	132	87	68	43	0
		ID$_{Mo}$	167	109	90	45	12
	ET	All	600	328	234	164	0
		ID$_{Mi}$	450	293	222	177	0
		ID$_{Mo}$	600	335	273	159	83
	TT	All	706	376	305	218	0
		ID$_{Mi}$	532	349	305	210	0
		ID$_{Mo}$	706	414	301	225	102
Escalas de violaciones	*VTI*	All	4	2	0	0	0
		ID$_{Mi}$	4	1	0	0	0
		ID$_{Mo}$	4	2	0	0	0
	VTII	All	10	4	2	0	0
		ID$_{Mi}$	8	4	1	0	0
		ID$_{Mo}$	10	3	2	1	0
	PV	All	11	5	1	0	0
		ID$_{Mi}$	11	5	1	0	0
		ID$_{Mo}$	8	5	2	0	0

Low quartile indicates poor performance. Worst and best scores report either the high or low score in the sample, depending on the direction of scoring of the scale.

the test went on, especially fewer instances of VTII and PV. Thus, over the course of the test, examinees incorporate norms and can complete a given task with or without making a violation. In the case of VTI, the effect was weaker; when relaying instructions, strong emphasis should be placed on VTI, because it is hardest for examinees to incorporate. In view of the results, it seems clearly related to problem difficulty.

With respect to subjects' performance on different problems, the authors observed staggered performance in terms of problem solving. The same effect was reported using other versions of the TOL in the population with ID (Masson, Dagnan, & Evans, 2010). There are two clearly defined types, items 1–7 and 8–10, and their differentiation is the result of minimum moves shifting from 3 in 1–7, to 4 in 8–10. However, the data suggest there are certain problems (1, 4, and 7) on which the number of moves, execution times, and violations all increased. On problem 1, it seems to be because it is first, an effect of task novelty. Despite doing practice items beforehand, between those items

and problem 1 there seems to be a jump in difficulty. The authors suggest adding a new practice item, one that is harder than the current ones and more similar to problem 1. In any case, with that factor in mind, it is very important for the examiner to ensure the subject correctly understood the instructions. On item 4, conversely, the "extra" difficulty seems more the result of move composition than any other factor. On 7, the increase in moves, times, and violations seems to be in response to the increase to four moves. Even though examinees solve it, that jump increases task difficulty considerably. The new variable, Hits, follows a pattern most in line with problem difficulty (the proportion of Hits decreases with each item), followed by the variables Corr and Mov.

The variables that indicated significant differences beween the ID$_{Mi}$ and ID$_{Mo}$ groups were Corr, Hits, and Mov (those related to the subject's movements). The ones that did not show significant differences between the two groups were IT, ET, and TT (time related), and VTI, VTII, and PV (violation related). Thus, it is best for

clinicians to use Corr, Hits, and Mov to differentiate between the two groups.

Exploratory factor analysis of the nine TOL^{DXtm}-ID variables indicated that none of the goodness of fit indices were especially satisfactory in explaining the variability in scores, so it seems we must assume that multiple variables and factors were involved in subjects' actions. Nevertheless, the movement-related variables, Hits and Corr, explained the most variance in the one-factor solution (.71 and .46, respectively).

On another note, we examined the extent to which this test is related to others with an executive component. Factor analysis suggested a three-factor structure, which would be quite logical. First, there would be one factor comprised of BRIEF-P variables, a second factor would encompass variables with an executive component on the BT-ID and more predictor variables on the TOL^{DXtm} (Corr and Hits), and the third factor would include the remaining measures on the TOL^{DXtm}. This factor structure suggests the tests had high specificity. BRIEF-P scores seemed more closely related to one another than to the cognitive components they measure, executive. That seems to be because parents' impressions about the domains evaluated by the BRIEF-P were not consistent with outcomes on the executive functioning tests subjects completed (TOL^{DXtm}, BT-ID, and WCFST). The same conclusion could be reached about TOL^{DXtm} variables related to times and violations. The variables most predictive of outcomes on the TOL^{DXtm} and the other tests subjects completed, all assessing executive functions, belong to the second factor. As such, there seems to be a strong functional relationship between subjects' performance on the TOL^{DXtm}-ID, and the remaining executive tests. Analysis of the original test (TOL^{DXtm}) using other tests (intelligence, memory, and executive functions) produced a similar factor structure (Culbertson & Zillmer, 2001). Significant correlations were found with tests involving executive processes ($r = +.57$ to $r = -.54$), but as described above, correlations with intelligence tests were not significant. The original test presented four factors: (F1) tests of frontal performance, (F2) the TOL^{DXtm}'s stronger predictor variables, (F3) memory tests, and (F4) psychometric intelligence tests. While that study did not use the same tests as the present one, the factor structure, at a cognitive functional level, bears strong similarity.

On a practical level, and to help clinically interpret the scores reported here, a barometer was created with quartiles to place subjects in normative groups. Table 7 presents all new score-generating variables on the TOL^{DXtm}-ID test. However, please note that according to the scores obtained, our recommendation is to utilize only the Corr variable. Given the alpha values and scores on the ten problems, that seems best. If an examinee's ID level can be determined, match with ID level. If not, match with ranking out of *All* subjects.

We must emphasize that given the limited number of subjects, the sample could be complemented with future studies to confirm results. That said, this study is part of a much broader longitudinal study which, once complete, could provide an interesting opportunity to analyze test-retest reliability (basic-24 month).

In conclusion, the version of the TOL^{DXtm} test for ID showed sufficient evidence of reliability and validity in adults with Down syndrome for it to be used in clinical practice as well as research. Its high correlations with other tests of frontal performance indicate it is especially suitable for assessing processes like planning, which is very important for people with ID.

Supplementary Material

To view supplementary material for this article, please visit https://doi.org/10.1017/S1138741617000300

References

Ball S., Holland A., Treppner P., Watson P., & Huppert F. (2008). Executive dysfunction and its association with personality and behaviour changes in the development of Alzheimer's disease in adults with Down syndrome and mild to moderate learning disabilities. *British Journal of Clinical Psychology*, 1, 1–29. https://doi.org/10.1348/014466507X230967

Berg W. K., & Byrd D. L. (2002). The Tower of London spatial problem-solving task: Enhancing clinical and research implementation. *Journal of Clinical and Experimental Neuropsychology*, 24, 586–604. https://doi.org/10.1076/jcen.24.5.586.1006

Cockburn J. (1995). Performance on the Tower of London test after severe head injury. *Journal of the International Neuropsychological Society*, 1, 537–544. https://doi.org/10.1017/S1355617700000667

Culbertson W. C., Moberg P. J., Duda J. E., Stern M. B., & Weintraub D. (2004). Assessing the executive function deficits of patients with Parkinson's disease: Utility of the Tower of London-Drexel. *Assessment*, 11, 27–39. https://doi.org/10.1177/1073191103258590

Culbertson W., & Zillmer E. (2001). *The Tower of London-Drexel University*. Chicago, IL: Multi-Health Systems.

Davydov D. M., & Polunina A. G. (2004). Heroin abusers' performance on the Tower of London Test relates to the baseline EEG alpha2 mean frequency shifts. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 7, 1143–1152. https://doi.org/10.1016/j.pnpbp.2004.06.006

Esteba-Castillo S., Dalmau-Bueno A., Ribas-Vidal N., Vilà-Alsina M., Novell-Alsina R., & García-Alba J. (2013). Adaptación y validación del Cambrigde Examination for Mental Disorders of Older People with Down's Syndrome and Others with Intellectual Disabilities (CAMDEX) en población adulta con discapacidad intelectual [Adaptation and validation of the Cambrigde Examination for Mental Disorders of Older People with Down's Syndrome and Others with Intellectual Disabilities (CAMDEX) in an adult population

with intellectual disability]. *Revista de Neurología*, 57, 337–346.

Esteba-Castillo S., & García-Alba J. (2015). Neuropsicología de la discapacidad intelectual [The neuropsychology of intellectual disability]. In A. Enseñat, T. Roig, & A. García, *Neuropsicología pediátrica* (pp. 163–188). Madrid, Spain: Síntesis.

Esteba-Castillo S., Novell R., Vilá A., & Ribas N. (2014). *CAMDEX-DS. Prueba de Exploración Cambridge para la Valoración de los Trastornos Mentales en Adultos con Síndrome de Down o con Discapacidad Intelectual* [Cambridge Examination for Mental Disorders of Adults with Down's Syndrome and Others with Intellectual Disabilities]. Madrid, Spain: TEA Ediciones.

Esteba-Castillo S., Caixas A., Deus J., & Peña-Casanova J. (2015). *Neuropsicología del trastorno del desarrollo intelectual con y sin origen genético* [The neuropsychology of intellectual disability with and without genetic cause]. (Unpublished Doctoral Dissertation). Universidad Autónoma de Barcelona. Barcelona, Spain.

Feldt L. S., Woodruff D. J., & Salih F. A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement*, 11(1), 93–103. https://doi.org/ 014662168701100107

Gioia G. A., Isquith P. K., Guy S. C., & Kenworthy L. (2000). Behavior rating inventory of executive function. *Child Neuropsychology*, 6, 235–238. https://doi.org/ 10.1076/chin.6.3.235.3152

Gioia G. A., Isquith P. K., Guy S. C., & Kenworthy L. (2016, October). *Behavior Rating Inventory of Executive Funtion (BRIEF)*. North Florida, FL: PARINC. Retrieved from http://www4.parinc.com/Products/Product.aspx?ProductID=BRIEF

Goldstein K., & Scheerer M. (1941). Abstract and concrete behavior an experimental study with special tests. *Psychological Monographs*, 53, i–151. https://doi.org/ 10.1037/h0093487

Hartman E., Wouwen S., Scherder E., & Visscher C. (2010). On the relationship between motor performance and executive functioning in children with intellectual disabilities. *Journal of Intellectual Disability Research*, 54, 468–477. https://doi.org/10.1111/j.1365-2788.2010.01284.x

Hobson P., Meara J., & Taylor C. (2007). The Weigl Colour-Form Sorting Test: A quick and easily administered bedside screen for dementia and executive dysfunction. *International Journal of Geriatric Psychiatry*, 22, 909–915. https://doi.org/10.1002/gps.1765

Jacobs R., & Anderson V. (2002). Planning and problem solving skills following focal frontal brain lesions in childhood: Analysis using the Tower of London. *Child Neuropsychology*, 8, 93–106. https://doi.org/10.1076/ chin.8.2.93.8726

Kaufman A., & Kaufman N. (2004). *Kaufman Brief Intelligence Test, Second Edition (KBIT-2).* Bloomington, MN: Pearson.

Kotovsky K., Hayes J. R., & Simon H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, 17, 248–294. https://doi.org/ 10.1016/0010-0285(85)90009-X

Masson J. D., Dagnan D., & Evans J. (2010). Adaptation and validation of the Tower of London test of planning and problem solving in people with intellectual disability. *Journal of Intellectual Disability Research*, 54, 457–467. https://doi.org/10.1111/j.1365-2788.2010.01280.x

Medina-Gómez M. B., & García-Alonso I. (2010). *Evaluación de la conducta adaptativa de las personas con discapacidad intelectual. Valoración y usos de la escala ABS-RC:2* [Assessment of adaptive behavior in people with intellectual disability. Evaluation and uses of the ABS-RC2 scale]. (Unpublished doctoral dissertation). University of Burgos, Burgos, Spain.

Moore C. J., Daly E. M., Schmitz N., Tassone F., Tysoe C., Hagerman R. J., … Murphy D. G. M. (2004). A neuropsychological investigation of male premutation carriers of fragile X syndrome. *Neuropsychologia*, 42, 1934–1947. https://doi.org/10.1016/j.neuropsychologia. 2004.05.002

Nieuwenhuis-Mark R. E. (2009). Diagnosing Alzheimer's dementia in Down syndrome: Problems and possible solutions. *Research in Developmental Disabilities*, 30, 827–838. https://doi.org/10.1016/j.ridd.2009.01.010

Nihira K., Leland H., & Lambert N. (1993). *Adaptive Behavior Scale: Residential and Community (ABS-RC:2)*. Austin, TX: Pro-ed.

Owen A. M., Downes J. J., Sahakian B. J., Polkey C. E., & Robbins T. W. (1990). Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia*, 28, 1021–1034. https://doi.org/ 10.1016/0028-3932(90)90137-D

Paula J. J., Neves F., Levy A., Nassif E., & Malloy-Diniz L. (2012). Assessing planning skills and executive functions in the elderly: Preliminary normative data for the Tower of London Test. *Archivos de Neuro-Psiquiatria*, 70, 828–829. https://doi.org/10.1590/S0004-282X2012001000015

Peña-Casanova J., Quiñones-Ubeda S., Gramunt-Fombuena N., Quintana M., Aguilar M., Molinuevo J. L., … Blesa R. (2009). Spanish multicenter normative studies (NEURONORMA Project): Norms for the Stroop color-word interference test and the Tower of London-Drexel. *Archives of Clinical Neuropsychology*, 24, 413–429.

Rainville C., Amieva H., Lafont S., Dartigues J. F., Orgogozo J. M., & Fabrigoule C. (2002). Executive function deficits in patients with dementia of the Alzheimer's type: A study with a Tower of London task. *Archives of Clinical Neuropsychology*, 17, 513–530. https://doi.org/10.1093/arclin/17.6.513

Robinson S., Goddard L., Dritschel B., Wisley M., & Howlin P. (2009). Executive functions in children with autism spectrum disorders. *Brain and Cognition*, 71, 362–368. https://doi.org/10.1016/j.bandc.2009.06.007

Rognoni T., Casals-Coll M., Sánchez-Benavides G., Quintana M., Manero R., Calvo L., … Peña-Casanova J. (2013). Spanish normative studies in young adults (NEURONORMA young adults project): Norms for Stroop Color-Word Interference and Tower of London-Drexel University tests. *Neurologia*, 28, 73–80. https://doi. org/10.1016/j.nrleng.2012.02.004

Shallice T. (1982). Specific impairments of planning. *Philosophical Transaction of the Royal Society*, 298, 199–209. https://doi.org/10.1098/rstb.1982.0082

Strauss H., & Lewin I. (1982). An empirical study of the Weigl-Goldstein-Scheerer color-form test according to a developmental frame of reference. *Journal of Clinical Psychology*, *38*, 367–375. https://doi.org/10.1002/1097-4679(198204)38:2<367::aid-jclp2270380226>3.0.co;2–9

Teipel S. J., Alexander G. E., Schapiro M. B., Möller H. J., Rapoport S. I., & Hampel H. (2004). Age-related cortical grey matter reductions in non-demented Down's syndrome adults determined by MRI with voxel-based morphometry. *Brain*, *127*, 811–824. https://doi.org/10.1093/brain/awh101

Tirapu J., García A., Luna P., & Periañez J. (2012). Evaluación de las funciones ejecutivas [Executive functions assessment]. In J. Tirapu, A. García, M. Rios, & A. Árdila (Eds.), *Neuropsicología de la corteza prefrontal y las funciones ejecutivas* (pp. 197–216). Barcelona, Spain: Viguera.

Tirapu-Ustárroz J., Muñoz-Céspedes J., Pelegrín-Valero C., & Albéniz-Ferreras A. (2005). A proposal for a protocol for use in the evaluation of the executive functions. *Revista de Neurología*, *41*, 177–186.

Unterrainer J. M., Rahm B., Leonhart R., Ruff C. C., & Halsband U. (2003). The Tower of London: The impact of instructions, cueing, and learning on planning abilities. *Cognitive Brain Research*, *17*, 675–683. https://doi.org/10.1016/S0926-6410(03)00191-5

Walley R. M., & Donaldson M. D. C. (2005). An investigation of executive function abilities in adults with Prader-Willi syndrome. *Journal of Intellectual Disability Research*, *49*, 613–625. https://doi.org/10.1111/j.1365-2788.2005.00717.x

Zhu Y., Liu X., Wang H., Jiang T., Fang Y., Hu H., … Zhang K. (2010). Reduced prefrontal activation during Tower of London in first-episode schizophrenia: A multi-channel near-infrared spectroscopy study. *Neuroscience Letters*, *478*, 136–140. https://doi.org/10.1016/j.neulet.2010.05.003

Zigman W. B., & Lott I. T. (2007). Alzheimer's disease in Down syndrome: Neurobiology and risk. *Mental Retardation and Developmental Disabilities Research Review*, *13*, 237–246. https://doi.org/10.1002/mrdd.20163