Measuring Goal Setting in School-Aged Children: Studying the Effects of Demographic Variables in Regression-Based Norms

A. Toornstra¹ · P. P. M. Hurks¹ · W. Van der Elst² · K. Massar¹ · G. Kok¹ · L. M. G. Curfs³

Received: 27 February 2020 / Revised: 27 February 2020 / Accepted: 16 March 2020 / Published online: 26 April 2020
© The Author(s) 2020

Abstract

The aim of the study was to establish demographically representative norms for tasks measuring goal setting, and more specifically planning and reasoning in children. Three tasks were administered to $n = 195$ Ukrainian children aged 5.10 to 14.5 years old: the Spatial Working Memory (SWM), the Stockings of Cambridge (SOC) test, and the Naglieri Nonverbal Ability Test (NNAT). Main outcome per test was accuracy: i.e., the total number correct for the SOC and NNAT, and the total amount of incorrect responses for the SWM. Correlations among accuracy measures varied from -0.51 to 0.60 , indicating these tasks measure related but at the same time unique constructs. Higher age was associated with more accurate test performances on all outcome measures. On the NNAT, we found a curvilinear association between age and accuracy, indicating that younger children's NNAT accuracy scores increased more with age compared with older children. We found a cubic age effect on accuracy for the SWM and SOC: i.e., test scores were relatively stable at younger and older ages, with a curvilinear increase in test scores in the other age groups. Demographically corrected norms were calculated and presented per test. These indicated that sex was not associated with accuracy scores on any of the tests. Last, a higher level of parental education (LPE) was associated with higher accuracy scores, but only on the NNAT. We conclude that demographic variables in norm analyses enhance insight in the scores and allow for application in clinical settings and research.

Keywords Planning · Reasoning · Spatial Working Memory (SWM) · Stockings of Cambridge (SOC) · NNAT · Continuous norming

Introduction

Goal setting is an important skill for daily life. Anderson (2002, p. 74) defined goal setting as “the ability to develop new concepts, as well as the capacity to plan actions in advance and approach tasks in an efficient and strategic manner.” Goal setting for example includes the ability to derive new notions (i.e., reasoning), planning steps to achieve end goals and applying this plan efficiently. As such, deficits in goal setting may lead to inadequate planning, and inadequate

execution of plans and limited conceptualizing of novel ideas while e.g. trying to solve a problem. The number of instruments available to measure goal setting in children is relatively scarce and suitable norm data are limited. However, interpretation of test scores of a specific sample is only meaningful when compared with suitable norm data.

The current study aims to establish such norms for a sample of Ukrainian schoolchildren for three tests measuring goal setting: namely, the Spatial Working Memory (SWM) test and the Stockings of Cambridge (SOC), which are both subtests from the Cambridge Neuropsychological Test Automated Battery (CANTAB®; Cambridge Cognition 2012) and the Naglieri Nonverbal Ability Test (NNAT; Naglieri 2003), which are described in detail below. We employ a relatively novel and promising method for norming, which is called continuous norming (Van Breukelen and Vlaeyen 2005; Van der Elst et al. 2011). This method includes demographic variables of the sample to evaluate their influence on the scores and to increase the meaningfulness of the norms (Heaton et al. 2003).

✉ A. Toornstra
toornstra.a2@outlook.com

W. Van der Elst
wim.vanderelst@gmail.com

¹ Faculty of Psychology and Neuroscience (FPN), Maastricht University, Maastricht, Netherlands

² Hasselt University, Hasselt, Belgium

³ MUMC, Maastricht University, Maastricht, Netherlands

We will now briefly describe the tests, which were all non-verbal and, apart from the NNAT, computer administered. This procedure also allowed testing of illiterate or low-literate children. In the SWM task, participants see boxes on the screen, and the task is to find as fast as possible tokens hidden behind some of the boxes (see Fig. 1 for an example). A box can be opened by touching it. Applying a process of elimination enhances token search efficiency and is as such believed to measure planning ability.

The SOC is a computerized adaptation of the Tower of London task (Ozonoff et al. 2004; Shallice 1982). The SOC measures the ability to plan and to execute a multi-step solution to a problem. Participants see colored circles (“balls”) in the bottom half of the screen that need to be moved to “stockings” until a pattern is created that is identical to the example pattern displayed in the top half of the screen (see Fig. 2 for an example). The trials increase in level of difficulty by increasing the number of moves needed to solve a problem.

Lastly, the NNAT consists of pattern matrices. Participants need to identify the correct missing part from five response options shown in order to complete the target drawing (see Fig. 3 for examples). This test is believed to measure reasoning, which falls within the functional domain of goal setting (Anderson 2002; Giessman et al. 2013; Naglieri and Ford 2003).

Within the domain of goal setting, the three tasks can be conceptualized as well-defined problems, i.e., with a clear start and final state (as opposed to poorly defined problems; Eysenck and Brysbaert 2018), which the child needs to solve. Reasoning in solving a problem is thought to be enhanced if the nature of the problem, the begin and end states, and the methods for a resolution are clear from the start. This well-defined analytical problem solving is thought to differ from creative problem solving (Lavric et al. 2000) in that the representations require relatively little novel conceptualization (Knoblich et al. 2001). Well-defined problems in contrast to creative problem solving (the latter may be required in poorly defined problems) may incorporate analogies and/or previous

Fig. 1 SWM on a touch screen: simultaneous view of closed boxes, one open box with a token and the column on the right with a previously found token (this is not a real CANTAB© SWM sample)

Fig. 2 The SOC on a touch screen: a two-move problem. The balls in the bottom “stockings” need to be dragged and moved until they are exactly in the same position as the top half of the screen (this is not a real CANTAB© SOC example)

experience with comparable problems. The demands on working memory during reasoning to solve a problem and planning and executing in these tasks are thus predominantly associated with visual spatial working memory and other executive functions such as planning, error monitoring, cognitive flexibility in adapting strategies after an error, inhibition of prepotent (incorrect) steps in e.g. the SOC, and sustaining attention for the increasing duration of trials with difficulty levels (Anderson 2002; Lavric et al. 2000; Mammarella et al. 2008).

In pediatric populations, norm data available for the tests discussed here have, to the best of our knowledge, been calculated in a traditional way (De Luca et al. 2003; Lehto et al. 2003; Luciana and Nelson 2002; Roque et al. 2011). Calculations in the traditional way tend to be based on e.g. 2-year age categories that are created in a sample, requiring significant sample sizes. Other disadvantages of such subgroups in the data are that sample characteristics of a specific age group may be irregular (e.g., scoring higher than expected for an age bracket), yet become the norm (Van der Elst et al. 2011). Also, children of virtually the same age may be in different age groups, while children who differ almost 2 years may be in the same age category: A child of 12.9 is in the 11 to 12 years old age bracket, whereas a child of 13.1 is in the 13 to 14 years old age group (Van der Elst et al. 2011). To avoid these discussion points in traditional norming, a method called continuous norming has been introduced, which will be applied in this study to collect SWM, SOC, and NNAT normative data for schoolchildren (Van Breukelen and Vlaeyen 2005; Van der Elst et al. 2011).

Continuous norming is arrived at by conducting multiple regression analyses that include several demographic variables, such as age, sex, and level of parental education (LPE; as a proxy for a child’s socio-economic position). These demographic variables have been linked to differences

Fig. 3 NNAT examples of the four different reasoning tasks: 1, pattern completion; 2, reasoning and analogy; 3, serial reasoning; 4, spatial visualization (these are not real NNAT samples)

in neuropsychological test performance, e.g., lower scores in working memory tasks have been associated with lower socioeconomic status (Lange et al. 2006; Strauss et al. 2006; Tallberg et al. 2008). For example, age-related development has been associated e.g. with improved goal setting abilities from 4 to 12 years old (Green et al. 2019; Luciana and Nelson 2002; Roque et al. 2011) and, in pediatric populations, the validity of the SWM, SOC, and NNAT was established in schoolchildren from 5 years and older (Green et al. 2019; Carman and Taylor 2010; Luciana and Nelson 2002; De Luca et al. 2003; Lehto et al. 2003; Rojahn and Naglieri 2006).

Data on the influence of children's biological sex are still inconclusive with regard to goal setting abilities in children. For instance, Pangelinan et al. (2011) did not find sex differences on the SWM in their sample of 172 children aged 6–13 years old, and Rojahn and Naglieri (2006) did not find sex differences on the NNAT in children from 6 to 9 years old. In 10–13-year-old children however, females scored slightly

higher on the NNAT and there was a slight sex difference in favor of males in 15–16-year-old children. However, these authors deemed the effects too small to apply to norm scores.

Additionally, the demographic variable LPE has not been studied often in relation to goal setting abilities in schoolchildren. LPE is perceived as an estimate for socioeconomic status (SES, White 1982). It has been found to correlate positively with children's higher academic achievement (Hackman and Farah 2009) and cognitive functions such as working memory (Mezzacappa 2004; Naglieri and Ronning 2000a; Noble et al. 2007). These findings underline the relevance to study the influence of age, sex, and LPE on the reasoning and planning (in the domain of goal setting) tasks in this study and to include demographic variables in these analyses attempting to establish norms. However, the amount of studies examining the validity of goal setting tasks such as the SWM and the SOC in pediatric populations is still limited and often do not include multiple demographic variables.

Table 1 Basic demographic data

	N	Age (mean in years)	Age SD	Sex		Level of parental education (LPE)	
				Female	Male	Low	High
SWM	169	9.89	2.43	78	91	17	152
SOC	152	10.02	2.43	69	83	14	138
NNAT	195	9.74	2.52	95	100	24	171

Table 2 Means and standard deviations (SD) for the SWM, the SOC, and the NNAT outcome measures (observed scores)

	Mean	Std. deviation	N	Minimum of the sample	Maximum of the sample
SWM total errors	54.65	21.30	169	2	96
SOC problems solved in minimum moves	5.08	3.09	152	0	11
NNAT	19.85	8.93	195	0	48

In the current study, our aim is to fill part of the gaps in the existing literature on the influence of demographic variables on goal setting tasks. In the statistical method of regression analyses (Van der Elst et al. 2011), the categorical variables sex (male/female) and LPE (high/low) and the continuous variable age are simultaneously included. This means that age as a continuous variable will be applied to each data point in the sample (Bechger et al. 2009). Both the option of including several demographic variables in the analyses and the application of each variable to all data (and not a specific subcategory, e.g., a 2-year age group) make these regressions more adequate and individualized (Van der Elst et al. 2011; Toornstra et al. 2020). Further, the second aim of the research is the current sample, which consists of schoolchildren in Ukraine and is therefore different compared with the mostly Western norm samples of the respective tests. Non-verbal tests are thought to be suitable for measurement in different populations but may still not be completely culture free (e.g., in their design), thus supporting the need to establish the normal range for this sample (Rosselli and Ardila 2003). Norms for these measures have, to the best of our knowledge, not been established for Ukraine and hence are needed in order to interpret scores in a meaningful manner (Lange et al. 2006). The aim of the present study was to assess the effect of demographic variables and establish a normative data range on the SWM, SOC, and NNAT, for a sample of $n = 195$ schoolchildren in Ukraine, aged 5.10 to 14.5 years old.

Methods

Participants

The sample of $n = 195$ children was enrolled in local schools in Ukraine which provides a 9-year compulsory education up until the children are in general 15 years old (Table 1). Age was approximately evenly distributed except for the youngest and oldest age groups. Per school, all children were invited and 80.5% participated. The LPE was based on the level of education completed by the parent(s) in a household; low described education up until grade 9 (similar to primary and middle schools) and high applied to grade 10 (high school or polytechnic) and above, according to the Ukrainian educational system (Ukraine Channel, 2017). Education is compulsory

up until and including grade 9. The highest level was assigned as the mean for the household in the 6.5% of the cases showing a discrepancy between parents' (or caregivers') level of education (in line with e.g. Koball and Jiang 2018). Please see Table 1 for an overview of all background information for the current sample.

Procedure and Instrument

Researchers contacted schools to partake in the study, after which responding school managements invited parents to information meetings about this study. Information and consent letters were provided by the researchers and the school management and then taken home. Testing started after parents and children had given informed consent letters to their class teachers.

The SWM and SOC, as part of the Cambridge Neuropsychological Test Automated Battery (CANTAB®, Cambridge Cognition 2012), and the NNAT were administered individually in a separate room. CANTAB®-trained testers administered the SOC and SWM on a HP Pavilion TS sleek-book laptop with a 15-in. touch screen. The paper and pencil NNAT was administered on a different day within the same weeks with the provided trial booklet. Two trained researchers administered tests. Instructions followed the pre-described scripts and practice items in the school's language of instruction. All tests were terminated after their respective termination rules (see "Materials"). The research ethics committee of the Faculty of Psychology and Neuroscience of Maastricht University, The Netherlands, approved this study. All data were obtained in compliance with ethics regulations of the WMA declaration of Helsinki (2013). School management received debriefing by anonymous group reports per class.

Table 3 Pearson zero-order correlations for the SWM, the SOC, and the NNAT outcome measures

	1	2	3
1 SWM total errors	1		
2 SOC problems solved in minimum moves	−0.53**	1	
3 NNAT	−0.51**	0.60**	1

** Correlation is significant at the 0.01 level (2-tailed)

Table 4 Final regression models for the SWM, the SOC, and the NNAT

Test score	<i>N</i>	Variable	<i>B</i>	<i>p</i>	<i>T</i>	Stand. deviation (residual)	<i>R</i> ²
SWM total errors (Box-Cox transformation, λ 1.40)	169	(Constant)	121.04	0.000	0.63		
		Age ²	0.17	0.000	−1.59		
		Age ³	−0.0069	0.000	1.63		
		Age ³ (LNAge)	0.001	0.000	2.36	47.27	0.27*
SOC problems solved in minimum moves (Box-Cox transformation, λ 1.60)	152	(Constant)	−2.57	0.000	−0.86		
		Age ³	0.00009	0.000	3.64		
		Age ³ (LNAge)	−0.000017	0.000	−3.54	4.20	0.21*
NNAT	193**	(Constant)	14.96	0.000	9.99		
		Age_C	0.16	0.000	9.43		
		Age_C ²	−0.002	0.002	−3.19		
		LPE	7.53	0.000	5.13	6.61	0.41*

* $p < 0.000$. ** N is lower due to removal of 2 outliers

Materials

First, SWM is a computerized self-ordered search task, which requires planning (Cambridge Cognition 2012). Lowe and Rabbitt (1998) found a moderate reliability (0.6–0.68) on the CANTAB© SWM.

In this task, blue tokens need to be found in boxes that appear on the screen. Once a token has been found, this token is placed in an empty column on the right side of the screen (Fig. 1). Tokens are hidden according to pre-set rules unknown to the participant: The box where a token was found will not have a token again for the duration of this particular trial. Once a token has been found, a continued search from that box working in a systematic way is deemed most effective (Owen et al. 1990). Each trial starts with a similar screen with boxes and the trials increase in difficulty by increasing the number of tokens that need to be found to a maximum of 8

(thus increasing the chance of errors as more boxes need to be remembered that contained a token). The main outcome measure of our study is the number of errors (SWM total errors); i.e., fewer errors are thought to reflect an efficient search strategy and are therefore seen as a measure of accuracy (Cambridge Cognition 2012). The break-off point for the whole test is based on too many total errors. Total administration time is on average 10 min.

The SOC (Cambridge Cognition 2012) is a computerized task of planning and reasoning. Lowe and Rabbitt (1998) found a moderate reliability (0.6–0.68) on an older version of the CANTAB© SOC.

The SOC test procedure is as follows: Two displays with three colored balls each are shown on a screen. The balls in the lower half of the screen need to be moved until they form an exact copy of the example pattern in the upper part of the screen (see Fig. 2). The balls can only be moved according

Fig. 4 Trajectory of the regression-based scores for the SWM, the SOC, and the NNAT (high LPE) accuracy scores by age

to specific rules, e.g., only one ball can be moved at a time, and a ball at the bottom cannot be moved if there is one on top. The number of moves is seen as a measure of the ability to plan, with fewer moves indicating better planning ability. The problems increase in difficulty with respect to the necessary minimum number of moves, from 2 to 5 moves, to copy the image at the top of the screen. Each of these trials starts with a new screen. The outcome measure “Problems solved in minimum moves” reflects the amount of times a trial was successfully completed, using the minimum amount of moves. The best possible score is 12 (i.e., 12 problems solved using the minimum number of moves). The test break-off point is when three consecutive problems registered as too many moves (the exact amount is dependent on the problem, i.e., 7 moves for a problem that can be solved in 3 moves). Total administration time is on average 13 min.

Last, the NNAT is an untimed one-on-one task on paper (Naglieri 2003). It purports to measure reasoning.

The test consists of four different forms of reasoning in increasingly difficult trials and thus involves cognitive flexibility and goal setting (see Fig. 3). Proof of the test’s validity for children was found in correlations of the group version of the NNAT with well-established tests, such as the Raven’s Progressive Matrices (RPM) test, as well as reliability from 5 years old onward (Naglieri 2003; Naglieri and Ronning 2000b; Raven 2000). The procedure is as follows: One page shows a two-colored matrix with a part left blank. The correct missing part needs to be chosen from 5 response options underneath (see Fig. 3). The outcome measure is the summation of correct responses (maximum possible is 72). Total administration time of this untimed paper test depends on the individual student; the break-off point is four consecutive incorrect trials.

Statistical Analyses

First, means and standard deviations are calculated for accuracy scores of all tests (see Table 2). Then, the observed means of the Ukrainian schoolchildren are compared with the observed means of the Western norm samples for the SWM, the SOC, and the NNAT respectively by evaluating the overlap between the confidence intervals (CIs) of the two samples. This comparison is done based on the Western age subgroups, and subgroups that are the same compared with the Ukrainian sample, which means the oldest and youngest children are excluded. Next, Pearson zero-order correlations are calculated between the Ukrainian SWM, SOC, and NNAT observed scores (Table 3). Regression models are established for each outcome measure of the SWM, SOC, and NNAT, with the demographic variables age, sex, and LPE, and their two-way

interactions (e.g., age $\times$ LPE). Sex was dummy coded as 0 = female and 1 = male. LPE was dummy coded as 0 = low and 1 = high. Age is centered (Age_C = calendar age in months – mean age of the sample 116.44 months), and quadratic (age_C²) and cubic (age_C³) age are included to evaluate curvilinear effects, i.e., scores increase less with age in older children (Van der Elst et al. 2011). Plotted observed scores and the predicted scores are visually inspected for potentially more complex curvilinear patterns. Fractional polynomial (regression) models with exponents are used to achieve the best possible fit capturing complex curvilinearity in the data. The demographic variables (age, sex, LPE) and their two-way interactions are included (Van der Elst et al. 2016). A restricted set of 7 exponents {–2, –1, –0.5, 0, 0.5, 1, 2} is used, which is typically adequate to capture complex relationships between variables (Van der Elst et al. 2016). We establish which demographic variables are predictive for each of the outcome measures of the SWM, SOC, and NNAT. The full models are reduced in a stepwise hierarchical way by removing the least significant predictor if its *p* value is higher than 0.05, and Akaike’s Information Criterion (AIC) is used to identify the two best powers (i.e., the lower the better the fit) for the models.

Next, the assumptions of regressions analysis are evaluated for each model, that is the normal distribution of residuals, the absence of heteroscedasticity, multicollinearity, and influential cases (Field 2009; Van der Elst et al. 2006). The Kolmogorov-Smirnov test is applied to check the normal distribution of the residuals. The occurrence of multicollinearity is tested by calculating the Variance Inflation Factors ($VIF \leq 10$) while Cook’s distance and standard deviation of residuals $> |3|$ are used to identify influential cases (Fisher et al. 2014). The residuals are plotted and visually inspected. Last, the standardized residuals are grouped in quartiles based on the predicted scores. Levene’s test is used to evaluate heteroscedasticity in the standardized residuals based on these quartile groups using $p = 0.05$.

In the last step, the normative data are calculated using the expected scores for each participant based on the final models. The residuals are calculated (=observed score – expected score). The residuals are standardized using the standard deviation (SD) of the residuals of the analyzed sample (= residual/SD (residual) of the normative sample; Van der Elst et al. 2011). Last, the residuals are converted to percentile values based on the normal distribution if the normality assumption (as evaluated in the Kolmogorov-Smirnov test) is met. If this normality assumption is not met, the empirical distribution of the standardized residuals of the sample is used. Making these calculations per child in practice is tedious so tables with these converted scores are in the Appendix to support ease of use. An alpha level of 0.05 will be applied to avoid type 1 errors. All calculations are made in SPSS version 24 and R.

Results

The means and SD for the SWM, the SOC, and the NNAT outcome measures (observed untransformed scores of the Ukrainian sample) are presented in Table 2. The tests were part of a test battery and especially younger children started to drop out in the latter part of this battery consisting of the SWM and the SOC (i.e., not starting the test), while others dropped out due to absenteeism and/or illness (SWM $n = 169$ and SOC $n = 152$ children).

The results of the correlation analysis between the SWM, SOC, and NNAT outcome measures are presented in Table 3. All accuracy measures correlated significantly with each other in the expected direction, meaning that higher accuracy on one is linked to higher accuracy on the other (i.e., lower is better on SWM errors, hence the negative correlations with NNAT and SOC problems solved in minimum moves scores).

Next, the final regression models ($p < 0.01$) are shown in Table 4. The NNAT, the SWM, and the SOC are all curvilinear models. Box-Cox transformations were applied to SWM total error and SOC scores, because preliminary analyses suggested heteroscedasticity with the untransformed scores (Osborne 2010). Also, outliers (i.e., 2 for the NNAT), which are scores with a standardized residual $SD > 3.0$, were removed. After these transformations and removal of outliers, the assumptions of regression analyses were met for the final models, i.e., Kolmogorov-Smirnov values $p \geq 0.05$, except for SWM total errors $p = 0.000$. All values of the Levene's statistic $p \geq 0.05$; Cook's distance values ≤ 0.01 ; and all Variance Inflation Factors ≤ 1.02 (except for SWM total errors, VIF 6.4, which is below the cutoff of 10 that is indicative of multicollinearity). There were only significant main effects and no significant interaction effects. The SWM total errors did not meet the normality assumption. Transformations of the standardized scores to percentiles were therefore based on the distribution of the standardized residuals (Zhou 1998).

None of the participants obtained the maximum or lowest possible score (lowest, i.e., 0 errors on the SWM) on any of the outcome measures. The best scores achieved were 11 (out of 12) on the SOC by 2%, 2 errors on the SWM by 0.6%, and 48 (out of 72) on the NNAT by 0.5% of the participants (no ceiling effects). Last, the CIs ($p = 0.01$) for the grand mean scores showed overlap for the SWM (CI Ukrainian sample ranges between 40.82 and 56.92, and 18.96 and 47.99 for the CANTAB norm group; Cambridge Cognition 2014). The CIs of the grand mean scores showed minimal overlap between the Ukrainian sample and the CANTAB norm group on the SOC (CI 3.99–6.55 and 6.06–8.75 respectively). There was no overlap between the CIs of the grand mean NNAT scores (CI Ukrainian children 15.69–21.31 and CI 22.80–30.45 in the NNAT norm sample; Naglieri 2003), so the score ranges differed between the two samples. These differences

highlight the importance of having normative data for Ukrainian children.

Normative Procedure

Based on the established regression models, norms for the SWM, the SOC, and the NNAT outcome measures are calculated for the Ukrainian sample by applying the procedure described above (norm tables in Appendix). Box-Cox transformations were reverted back to untransformed scores (Osborne 2010). For example, presume that an 11-year-old child completed 20 items correctly on the NNAT and the level of education of the child's parents is high. The regression model for the NNAT is $\text{Constant} + B_1(\text{Age_C}) + B_2(\text{Age_C}^2) + B_3(\text{LPE})$ with the model presented in Table 4. This is $\text{Constant} + 0.16 \times (\text{age child in months} - \text{average age sample}) + (-0.002 \times \text{Age_C}^2) + (1 \times 7.53)$, that is $14.96 + [0.16 \times (132 - 116.44)] + (-0.002 \times 242.11) + 7.53 \times 1 = 24.5$. The residual is computed, which is $20 - 24.5 = -4.5$. In the third step, the residual is standardized $-0.68 (-4.5/6.61)$. The standardized residual is turned into a percentile based on the standard normal cumulative distribution. A standardized residual of -0.68 corresponds with a percentile value equal to 23. This means that 23% of the population of 11-year-old children with high LPE obtain a NNAT score that is equal or lower than 20. The NNAT score of this child is therefore in the lower part of the normal range (5th–95th percentile). Lastly, the NNAT outcome measure showed a quadratic age effect, which indicates a curvilinear development with age. The SWM and SOC scores showed cubic age effects, that is a trajectory with multiple curves. The cross-sectional developmental trajectories of all scores are in Fig. 4. The NNAT and SOC show in principle increasing accuracy between the ages of 5.5 to 13.5 years old and then start to level off. Figure 4 also shows development of scores with age for the outcome measure SWM total errors in which a lower score is better. Several bends in the regression line are typical for a cubic age effect as in SWM total errors, which appears to stabilize from 13 years old. Norm tables are in the Appendix.

Discussion

The aim of this study was to establish demographically corrected pediatric norms for three tests measuring goal setting in school-aged children in Ukraine. Anderson (2002) defined goal setting as “the ability to develop new concepts, as well as the capacity to plan actions in advance and approach tasks in an efficient and strategic manner” (p. 74). We hence looked more specifically at planning and reasoning. The normal range of performance for Ukrainian children on these tasks was established by applying an auspicious method called

continuous norming. This different approach to norming included simultaneously taking into consideration the impact of continuous and categorical demographic variables on test performance. Demographic variables included in this study are age, sex, and LPE (Van Breukelen and Vlaeyen 2005; Van der Elst et al. 2011). This method results in more suitable estimates of norm scores for reasons considered in “Introduction.” The results of this study will be discussed next.

The current results show first of all that accuracy scores on the administered tests correlated in the expected directions. These correlations support the notion that reasoning and planning (including executing the planned steps to achieve a goal) occur in the same domain (EF domain), as hypothesized by Anderson (2002). The correlations are on the other hand moderate to strong (but not very strong, i.e., strong > 0.90), which indicates differences in underlying constructs in the tests.

Second, we found age-related changes on test performance on all tests; i.e., we found that on average older children made fewer errors on the SWM and solved higher amounts of problems on the SOC and the NNAT. This is in line with research by Pangelinan et al. (2011) who reported that age in typically developing children from 6 to 12.9 years old was associated with increasing accuracy on the SWM. Furthermore, the NNAT showed in principle curvilinear development with age, and stabilizing scores in the oldest children. The SOC and the SWM showed cubic development with age, i.e., some stabilizing in scores at lower and higher ages (Fig. 4). Development of working memory, a key component underlying planning as required for executing the SWM, SOC, and NNAT, reportedly has a (curvi)linear development with age from early childhood (Anderson 2002; De Luca et al. 2003; Gathercole et al. 2004; Luciana and Nelson 2002; Naglieri 2003). However, Crescentini et al. (2012) showed that planning also differs from memory (i.e., areas of brain activation involved) in a Tower of London task (comparable with the SOC). This is in line with Anderson (2002) who posed that planning and reasoning are related executive functions in the domain of goal setting, yet different from other functions such as information processing (which includes memory). Therefore, in principle, increased accuracy found with increasing age may be associated with planning and reasoning (and thus goal setting), which was also shown in the relatively fewer studies on normative data in children on the SWM and SOC (Green et al. 2019; De Luca et al. 2003; Lehto et al. 2003; Luciana and Nelson 2002; Roque et al. 2011).

Third, the scores of the Ukrainian sample were compared with the Western norm to examine whether test scores overlapped (Rosselli and Ardila 2003). The comparison was limited to traditional 2-year age subgroups, which was the only available demographic variable for both populations (Cambridge Cognition 2014; Naglieri 2003). The CIs of the grand mean for the SWM overlapped between the Ukrainian

sample and the CANTAB norm sample, which was also, but to a lesser degree, the case for the SOC scores. This indicates a comparable performance in planning, and goal setting in Ukrainian scores and Western scores (Lehto et al. 2003). The NNAT however showed no overlap in CIs between the Ukrainian sample compared with the Western mean scores. Additionally, 0.5% in the Ukrainian sample achieved the highest score on the NNAT, so there were no test ceiling effects. Taking this together, the NNAT scores appear to indicate a lower mean compared with the Western population. Caution is warranted as these comparisons were only based on descriptive data (CIs) and did not include the demographic variables sex and LPE, which would need to be included in future statistically more rigorous analyses assessing cultural differences in test scores.

Fourth, we found neither sex effects nor an age by sex interaction on any of the tests. This is in line with Luciana and Nelson (1998) who did not find sex differences on the accuracy scores of the SOC and SWM. Sex differences were found in a study on a large sample of the NNAT, but Rojahn and Naglieri (2006) deemed effect sizes negligible and were subsequently not included in the standardized norms, thus supporting the findings here.

The last finding is an effect of the demographic variable of LPE on the NNAT, linking higher NNAT accuracy to higher LPE. LPE is associated with the socio-economic status of families, which in turn is linked to life outcomes such as cognitive development of children (Bradley and Corwyn 2002). Bronfenbrenner and Morris (2006) hypothesized that planning is an example of a cognitive ability that is subject to environmental influences, i.e., a child that experiences modeling and learning of planning from e.g. the adults in his/her life may be more likely to use it when a situation requires. Lohman and Gambrell (2012) however found that children of lower SES performed better on the NNAT. Lohman and Gambrell (2012) link these results to test design differences, i.e., describing the NNAT and the Raven's as figural reasoning tests as opposed to most other non-verbal reasoning tests. The latter are thought to use, among others, pictures and line drawings which may be associated with other constructs in testing (e.g., language).

There are some limitations in the current study that need to be addressed. First, the number of parents with a low LPE group was smaller than the high LPE group. In multiple linear regressions, each data point of the whole sample is used (i.e., relation of the demographic variables to the dependent variable of the test score); thus, fewer data are needed to arrive at statistically valid results (Van Breukelen and Vlaeyen 2005). Similarly, these regressions allow for a smaller overall sample, which supports these analyses despite higher dropout rates in the SWM and SOC which were in the last half of this test battery. Secondly, the number of regressions computed might increase the chance of type I errors. Although it cannot be stated with absolute certainty that some relationships are not

chance findings (Piechatek et al. 2009), the discussion shows that the findings in our study are supported by (neuro)psychological literature. Furthermore, discussing the difference between short-term memory and working memory is not within the scope of this study (Faw 2003). Discussions here focus on the ability of keeping information in memory for a short while and to manipulate that information according to task demands, which aligns with Anderson's (2002) conceptualization of working memory.

Conclusion

Demographically corrected norms were established for three tests measuring planning and reasoning (in the domain of goal setting) in children, i.e., the SWM, the SOC, and the NNAT. Demographic variables (age, sex, LPE) were associated with test scores on these tests, but not all were associated with the

scores in this sample. Including demographic variables is therefore important when calculating norm scores (Mitrushina et al. 2005). Two tests showed differences in score trajectories compared with Western norms (i.e., cubic age effects), which supports our contention to examine differences between populations in relation to test scores (in line with Rosselli and Ardila 2003). Providing the norm tables and establishing the normal range for different populations especially for functions that are difficult to evaluate (e.g., planning and reasoning and thus goal setting) is important for use in clinical and research settings.

Compliance with Ethical Standards The research ethics committee of the Faculty of Psychology and Neuroscience of Maastricht University, The Netherlands, approved this study. All data were obtained in compliance with ethics regulations of the WMA declaration of Helsinki (2013).

Appendix

Table 5 Normative data for the SWM total errors (accuracy) stratified by age

Percentile	Age in years																		
	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10	10.5	11	11.5	12	12.5	13	13.5	14	14.5
99	55	56	55	54	52	49	46	43	39	35	32	27	24	21	20	20	23	28	35
98	56	56	55	54	52	50	47	43	39	35	33	27	24	21	20	21	23	28	35
97	56	56	56	55	53	50	47	44	40	36	33	28	25	22	21	22	24	29	36
96	57	57	56	55	53	51	48	44	41	37	34	29	25	23	22	22	25	29	36
95	57	57	57	56	54	51	48	45	41	37	35	29	26	24	22	23	25	30	37
94	58	58	57	56	54	52	49	46	42	38	35	30	27	24	23	24	26	31	37
93	58	58	58	56	55	52	49	46	42	38	35	30	27	25	23	24	26	31	38
92	58	59	58	57	55	53	50	46	43	39	36	31	28	26	24	25	27	32	38
91	59	59	59	58	56	53	50	47	43	39	37	32	29	26	25	26	28	32	39
90	59	59	59	58	56	53	50	47	43	39	37	32	29	26	25	26	28	33	39
89	59	59	59	58	56	54	51	47	44	40	37	32	29	27	26	26	28	33	39
88	59	60	59	58	56	54	51	48	44	40	37	32	29	27	26	26	29	33	40
87	60	60	59	58	56	54	51	48	44	40	38	33	30	27	26	27	29	33	40
86	60	60	60	59	57	54	51	48	44	41	38	33	30	28	27	27	29	34	40
85	60	60	60	59	57	55	52	48	45	41	38	33	30	28	27	27	30	34	40
84	60	61	60	59	57	55	52	49	45	41	39	34	31	28	27	28	30	34	41
83	61	61	61	59	58	55	52	49	45	42	39	34	31	29	28	28	31	35	41
82	61	61	61	60	58	55	53	49	46	42	39	34	31	29	28	28	31	35	41
81	61	62	61	60	58	56	53	50	46	42	40	35	32	29	28	29	31	35	42
80	61	62	61	60	58	56	53	50	46	42	40	35	32	30	28	29	31	35	42
75	63	63	62	61	60	57	54	51	48	44	41	37	34	31	30	31	33	37	43
50	70	71	70	69	67	65	62	59	56	53	50	46	43	41	40	41	43	46	52

Table 5 (continued)

	Age in years																		
Percentile	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10	10.5	11	11.5	12	12.5	13	13.5	14	14.5
25	78	78	78	77	75	73	70	67	64	61	59	55	52	50	49	50	51	55	60
20	79	79	79	78	76	74	71	68	65	62	60	56	53	52	51	51	53	56	62
19	79	79	79	78	76	74	71	69	65	62	60	56	53	52	51	51	53	56	62
18	80	80	79	78	77	75	72	69	66	63	61	57	54	52	51	52	54	57	62
17	80	80	80	79	77	75	73	70	66	63	61	57	55	53	52	52	54	58	63
16	80	80	80	79	77	75	73	70	67	63	61	57	55	53	52	52	54	58	63
15	80	80	80	79	77	75	73	70	67	63	61	57	55	53	52	53	54	58	63
14	80	81	80	79	78	75	73	70	67	64	62	58	55	53	52	53	55	58	63
13	81	81	81	79	78	76	73	70	67	64	62	58	55	54	53	53	55	58	64
12	81	81	81	80	78	76	73	70	67	64	62	58	55	54	53	53	55	58	64
11	81	81	81	80	78	76	73	71	67	64	62	58	56	54	53	53	55	59	64
10	81	81	81	80	78	76	74	71	68	64	62	58	56	54	53	53	55	59	64
9	82	82	82	80	79	77	74	71	68	65	63	59	57	55	54	54	56	59	65
8	82	82	82	81	79	77	74	71	68	65	63	59	57	55	54	54	56	60	65
7	82	82	82	81	79	77	75	72	69	66	63	59	57	55	54	55	57	60	65
6	82	83	82	81	79	77	75	72	69	66	64	60	57	55	55	55	57	60	65
5	83	83	82	81	80	78	75	72	69	66	64	60	58	56	55	55	57	61	66
4	83	84	83	82	81	79	76	73	70	67	65	61	59	57	56	56	58	61	67
3	84	84	83	82	81	79	76	73	70	67	65	61	59	57	56	57	58	62	67
2	85	85	84	83	82	80	77	74	71	68	66	62	60	58	57	58	59	63	68
1	86	86	86	85	83	81	79	76	73	70	68	64	62	60	59	60	61	65	70

Table 6 Normative data for the raw score SOC problems solved in minimum moves stratified by age

	Age in years																		
Percentile	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	10.5	11.0	11.5	12.0	12.5	13.0	13.5	14.0	14.5
99	6	6	6	7	7	8	8	8	9	9	9	9	9	9	9	9	9	9	9
98	5	6	6	7	7	7	8	8	8	8	9	9	9	9	9	9	9	9	9
97	5	6	6	6	7	7	7	8	8	8	8	9	9	9	9	9	9	9	8
96	5	5	6	6	7	7	7	8	8	8	8	9	9	9	9	9	9	9	8
95	5	5	6	6	6	7	7	7	8	8	8	8	9	9	9	9	9	8	8
94	5	5	5	6	6	7	7	7	8	8	8	8	9	9	9	9	8	8	8
93	5	5	5	6	6	7	7	7	8	8	8	8	8	9	9	9	8	8	8
92	4	5	5	6	6	6	7	7	7	8	8	8	8	8	8	8	8	8	8
91	4	5	5	6	6	6	7	7	7	8	8	8	8	8	8	8	8	8	8
90	4	5	5	6	6	6	7	7	7	8	8	8	8	8	8	8	8	8	8
89	4	5	5	5	6	6	7	7	7	8	8	8	8	8	8	8	8	8	8
88	4	5	5	5	6	6	7	7	7	7	8	8	8	8	8	8	8	8	8
87	4	4	5	5	6	6	6	7	7	7	8	8	8	8	8	8	8	8	8
86	4	4	5	5	6	6	6	7	7	7	8	8	8	8	8	8	8	8	7
85	4	4	5	5	6	6	6	7	7	7	7	8	8	8	8	8	8	8	7
84	4	4	5	5	6	6	6	7	7	7	7	8	8	8	8	8	8	8	7
83	4	4	5	5	5	6	6	7	7	7	7	8	8	8	8	8	8	8	7
82	4	4	5	5	5	6	6	7	7	7	7	8	8	8	8	8	8	8	7

Table 6 (continued)

Percentile	Age in years																		
	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	10.5	11.0	11.5	12.0	12.5	13.0	13.5	14.0	14.5
81	4	4	5	5	5	6	6	7	7	7	7	8	8	8	8	8	8	8	7
80	4	4	4	5	5	6	6	6	7	7	7	8	8	8	8	8	8	7	7
75	3	4	4	5	5	5	6	6	7	7	7	7	8	8	8	8	8	7	7
50	2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	7	7	6	6
25	0	1	2	2	3	3	4	4	5	5	5	6	6	6	6	6	6	6	5
20	0	1	1	2	3	3	4	4	5	5	5	5	6	6	6	6	6	5	5
19	0	0	1	2	3	3	4	4	4	5	5	5	6	6	6	6	6	5	5
18	0	0	1	2	2	3	4	4	4	5	5	5	6	6	6	6	6	5	5
17	0	0	1	2	2	3	3	4	4	5	5	5	5	6	6	6	5	5	5
16	0	0	1	2	2	3	3	4	4	5	5	5	5	6	6	6	5	5	5
15	0	0	1	2	2	3	3	4	4	5	5	5	5	5	6	5	5	5	5
14	0	0	1	1	2	3	3	4	4	5	5	5	5	5	5	5	5	5	5
13	0	0	0	1	2	3	3	4	4	4	5	5	5	5	5	5	5	5	5
12	0	0	0	1	2	3	3	4	4	4	5	5	5	5	5	5	5	5	4
11	0	0	0	1	2	2	3	3	4	4	5	5	5	5	5	5	5	5	4
10	0	0	0	1	2	2	3	3	4	4	4	5	5	5	5	5	5	5	4
9	0	0	0	1	2	2	3	3	4	4	4	5	5	5	5	5	5	5	4
8	0	0	0	1	1	2	3	3	4	4	4	5	5	5	5	5	5	5	4
7	0	0	0	0	1	2	3	3	4	4	4	5	5	5	5	5	5	4	4
6	0	0	0	0	1	2	2	3	3	4	4	4	5	5	5	5	5	4	4
5	0	0	0	0	1	2	2	3	3	4	4	4	4	5	5	5	4	4	4
4	0	0	0	0	0	1	2	3	3	3	4	4	4	4	5	4	4	4	4
3	0	0	0	0	0	1	2	2	3	3	3	4	4	4	4	4	4	4	3
2	0	0	0	0	0	0	1	2	2	3	3	4	4	4	4	4	4	4	3
1	0	0	0	0	0	0	0	1	2	2	3	3	3	4	4	4	3	3	3

Table 7 Normative data for the raw NNAT score, stratified by age and level of parental education: LPE high

Percentile	Age in years																		
	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	10.5	11.0	11.5	12.0	12.5	13.0	13.5	14.0	14.5
99	25	27	29	31	32	34	35	36	37	38	39	40	40	41	41	41	41	41	40
98	23	25	27	29	30	32	33	35	36	37	37	38	39	39	39	39	39	39	39
97	22	24	26	28	29	31	32	33	35	35	36	37	37	38	38	38	38	38	38
96	21	23	25	27	28	30	31	33	34	35	35	36	37	37	37	37	37	37	37
95	20	22	24	26	28	29	31	32	33	34	35	35	36	36	36	37	37	36	36
94	20	22	24	25	27	29	30	31	32	33	34	35	35	36	36	36	36	36	35
93	19	21	23	25	27	28	30	31	32	33	34	34	35	35	35	35	35	35	35
92	19	21	23	24	26	28	29	30	31	32	33	34	34	35	35	35	35	35	34
91	18	20	22	24	26	27	29	30	31	32	33	33	34	34	34	35	34	34	34
90	18	20	22	24	25	27	28	29	31	32	32	33	33	34	34	34	34	34	34
89	17	20	21	23	25	26	28	29	30	31	32	33	33	33	33	34	34	34	33
88	17	19	21	23	25	26	28	29	30	31	32	32	33	33	33	33	33	33	33
87	17	19	21	23	24	26	27	28	30	30	31	32	32	33	33	33	33	33	33
86	16	19	21	22	24	26	27	28	29	30	31	32	32	33	33	33	33	33	32

Table 7 (continued)

Percentile	Age in years																		
	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	10.5	11.0	11.5	12.0	12.5	13.0	13.5	14.0	14.5
85	16	18	20	22	24	25	27	28	29	30	31	31	32	32	32	33	32	32	32
84	16	18	20	22	23	25	26	28	29	30	30	31	32	32	32	32	32	32	32
83	16	18	20	22	23	25	26	27	28	29	30	31	31	32	32	32	32	32	31
82	15	17	19	21	23	24	26	27	28	29	30	31	31	31	31	32	32	31	31
81	15	17	19	21	23	24	26	27	28	29	30	30	31	31	31	31	31	31	31
80	15	17	19	21	22	24	25	27	28	29	29	30	31	31	31	31	31	31	31
75	14	16	18	20	21	23	24	25	27	27	28	29	29	30	30	30	30	30	30
50	9	11	13	15	17	18	20	21	22	23	24	24	25	25	25	26	26	25	25
25	5	7	9	11	12	14	15	17	18	19	19	20	21	21	21	21	21	21	21
20	4	6	8	10	11	13	14	15	17	17	18	19	19	20	20	20	20	20	20
19	4	6	8	9	11	13	14	15	16	17	18	19	19	20	20	20	20	20	19
18	3	5	7	9	11	12	14	15	16	17	18	18	19	19	19	20	20	19	19
17	3	5	7	9	11	12	13	15	16	17	18	18	19	19	19	19	19	19	19
16	3	5	7	9	10	12	13	14	16	16	17	18	18	19	19	19	19	19	18
15	2	5	7	8	10	12	13	14	15	16	17	18	18	19	19	19	19	19	18
14	2	4	6	8	10	11	13	14	15	16	17	17	18	18	18	19	18	18	18
13	2	4	6	8	9	11	12	14	15	16	16	17	18	18	18	18	18	18	18
12	2	4	6	7	9	11	12	13	14	15	16	17	17	18	18	18	18	18	17
11	1	3	5	7	9	10	12	13	14	15	16	16	17	17	17	18	18	17	17
10	1	3	5	7	8	10	11	13	14	15	15	16	17	17	17	17	17	17	17
9	0	3	5	6	8	10	11	12	13	14	15	16	16	17	17	17	17	17	16
8	0	2	4	6	8	9	10	12	13	14	15	15	16	16	16	16	16	16	16
7	0	2	4	5	7	9	10	11	12	13	14	15	15	16	16	16	16	16	15
6	0	1	3	5	7	8	9	11	12	13	14	14	15	15	15	15	15	15	15
5	0	1	3	4	6	8	9	10	11	12	13	14	14	15	15	15	15	15	14
4	0	0	2	4	5	7	8	9	11	11	12	13	13	14	14	14	14	14	14
3	0	0	1	3	4	6	7	9	10	11	11	12	13	13	13	13	13	13	13
2	0	0	0	2	3	5	6	7	9	9	10	11	11	12	12	12	12	12	11
1	0	0	0	0	1	3	4	6	7	8	8	9	10	10	10	10	10	10	10

Table 8 Normative data for the raw NNAT score, stratified by age and level of parental education: LPE low

Percentile	Age in years																		
	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	10.5	11.0	11.5	12.0	12.5	13.0	13.5	14.0	14.5
99	17	19	21	23	25	26	28	29	30	31	32	32	33	33	33	34	33	33	33
98	15	17	19	21	23	24	26	27	28	29	30	31	31	31	31	32	32	31	31
97	14	16	18	20	22	23	25	26	27	28	29	29	30	30	30	31	31	30	30
96	13	15	17	19	21	22	24	25	26	27	28	29	29	29	29	30	30	29	29
95	13	15	17	19	20	22	23	24	25	26	27	28	28	29	29	29	29	29	28
94	12	14	16	18	20	21	23	24	25	26	27	27	28	28	28	28	28	28	28
93	12	14	16	17	19	21	22	23	24	25	26	27	27	28	28	28	28	28	27
92	11	13	15	17	19	20	22	23	24	25	26	26	27	27	27	27	27	27	27
91	11	13	15	17	18	20	21	22	23	24	25	26	26	27	27	27	27	27	26
90	10	12	14	16	18	19	21	22	23	24	25	25	26	26	26	27	27	26	26
89	10	12	14	16	17	19	20	22	23	24	24	25	26	26	26	26	26	26	26

Table 8 (continued)

Percentile	Age in years																		
	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	10.5	11.0	11.5	12.0	12.5	13.0	13.5	14.0	14.5
88	10	12	14	15	17	19	20	21	22	23	24	25	25	26	26	26	26	26	25
87	9	11	13	15	17	18	20	21	22	23	24	24	25	25	25	26	26	25	25
86	9	11	13	15	16	18	19	21	22	23	23	24	25	25	25	25	25	25	25
85	9	11	13	15	16	18	19	20	21	22	23	24	24	25	25	25	25	25	24
84	8	10	12	14	16	17	19	20	21	22	23	24	24	24	24	25	25	24	24
83	8	10	12	14	16	17	19	20	21	22	23	23	24	24	24	24	24	24	24
82	8	10	12	14	15	17	18	20	21	22	22	23	24	24	24	24	24	24	24
81	8	10	12	13	15	17	18	19	20	21	22	23	23	24	24	24	24	24	23
80	7	9	11	13	15	16	18	19	20	21	22	23	23	23	23	24	24	23	23
75	6	8	10	12	14	15	17	18	19	20	21	21	22	22	22	23	23	22	22
50	2	4	6	8	9	11	12	13	15	16	16	17	17	18	18	18	18	18	18
25	0	0	1	3	5	6	8	9	10	11	12	13	13	13	13	14	14	13	13
20	0	0	0	2	4	5	7	8	9	10	11	11	12	12	12	13	13	12	12
19	0	0	0	2	4	5	6	8	9	10	11	11	12	12	12	12	12	12	12
18	0	0	0	2	3	5	6	7	9	9	10	11	11	12	12	12	12	12	11
17	0	0	0	1	3	5	6	7	8	9	10	11	11	12	12	12	12	12	11
16	0	0	0	1	3	4	6	7	8	9	10	10	11	11	11	12	12	11	11
15	0	0	0	1	2	4	5	7	8	9	9	10	11	11	11	11	11	11	11
14	0	0	0	1	2	4	5	6	7	8	9	10	10	11	11	11	11	11	10
13	0	0	0	0	2	3	5	6	7	8	9	10	10	10	10	11	11	10	10
12	0	0	0	0	2	3	4	6	7	8	9	9	10	10	10	10	10	10	10
11	0	0	0	0	1	3	4	5	6	7	8	9	9	10	10	10	10	10	9
10	0	0	0	0	1	2	4	5	6	7	8	8	9	9	9	10	10	9	9
9	0	0	0	0	0	2	3	5	6	7	7	8	9	9	9	9	9	9	9
8	0	0	0	0	0	2	3	4	5	6	7	8	8	9	9	9	9	9	8
7	0	0	0	0	0	1	2	4	5	6	7	7	8	8	8	8	8	8	8
6	0	0	0	0	0	1	2	3	4	5	6	7	7	8	8	8	8	8	7
5	0	0	0	0	0	0	1	3	4	5	5	6	7	7	7	7	7	7	7
4	0	0	0	0	0	0	1	2	3	4	5	5	6	6	6	7	7	6	6
3	0	0	0	0	0	0	0	1	2	3	4	5	5	5	5	6	6	5	5
2	0	0	0	0	0	0	0	0	1	2	3	3	4	4	4	5	5	4	4
1	0	0	0	0	0	0	0	0	0	0	1	2	2	2	2	3	3	3	2

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology*, 8(2), 71–82. <https://doi.org/10.1076/chin.8.2.71.8724>.
- Bechger, T., Hemker, B., & Maris, G. (2009). < cito_pok_memorandum_2009-01-3.pdf >. Arnhem, The Netherlands: CITO.
- Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, 53(1), 371–399.
- Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. *Handbook of child psychology*.
- Cambridge Cognition (2012). *Neuropsychological Test Automated Battery (CANTABclipse) manual* C. Cognition (Ed.)

- Cambridge Cognition. (2014). *Child norms Cambridge Neuropsychological Test Automated Battery (manual)*. Cambridge: Cambridge Cognition Ltd.
- Carman, C. A., & Taylor, D. K. (2010). Socioeconomic status effects on using the Naglieri Nonverbal Ability Test (NNAT) to identify the gifted/talented. *Gifted Child Quarterly*, 54(2), 75–84.
- Crescentini, C., Seyed-Allaei, S., Vallesi, A., & Shallice, T. (2012). Two networks involved in producing and realizing plans. *Neuropsychologia*, 50(7), 1521–1535.
- De Luca, C. R., Wood, S. J., Anderson, V., Buchanan, J. A., Proffitt, T. M., Mahony, K., & Pantelis, C. (2003). Normative data from the Cantab. I: development of executive function over the lifespan. *Journal of Clinical & Experimental Neuropsychology*, 25(2), 242–254.
- Eysenck, M. W., & Brysbaert, M. (2018). *Fundamentals of cognition*, chapter 9 problem solving, p. 356–395. Routledge.
- Faw, B. (2003). Pre-frontal executive committee for perception, working memory, attention, long-term memory, motor control, and thinking: a tutorial review. *Consciousness and Cognition*, 12(1), 83–139.
- Field, A. (2009). Discovering statistics using IBM SPSS statistics. sage.
- Fisher, S. D., Gray, J. P., Black, M. J., Davies, J. R., Bednark, J. G., Redgrave, P., et al. (2014). A behavioral task for investigating action discovery, selection and switching: comparison between types of reinforcer. *Frontiers in Behavioral Neuroscience*, 8, 398. <https://doi.org/10.3389/fnbeh.2014.00398>.
- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, 40(2), 177.
- Giessman, J. A., Gambrell, J. L., & Stebbins, M. S. (2013). Minority performance on the Naglieri Nonverbal Ability Test, second edition, versus the Cognitive Abilities Test, Form 6: one gifted program's experience. *Gifted Child Quarterly*, 57(2), 101–109. <https://doi.org/10.1177/0016986213477190>.
- Green, R., Till, C., Al-Hakeem, H., Cribbie, R., Téllez-Rojo, M. M., Osorio, E., et al. (2019). Assessment of neuropsychological performance in Mexico City youth using the Cambridge Neuropsychological Test Automated Battery (CANTAB). *Journal of Clinical and Experimental Neuropsychology*, 41(3), 246–256.
- Hackman, D. A., & Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends in Cognitive Sciences*, 13(2), 65–73.
- Heaton, R. K., Taylor, M. J., & Manly, J. (2003). Chapter 5 - demographic effects and use of demographically corrected norms with the WAIS-III and WMS-III *Clinical interpretation of the WAIS-III and WMS-III* (pp. 181–210). San Diego: Academic Press.
- Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Memory & Cognition*, 29(7), 1000–1009.
- Koball, H., & Jiang, Y. (2018). Basic facts about low-income children: Children under 9 years, 2016. <https://academiccommons.columbia.edu/doi/10.7916/d8-97p-bf73>.
- Lange, R. T., Chelune, G. J., Taylor, M. J., Woodward, T. S., & Heaton, R. K. (2006). Development of demographic norms for four new WAIS-III/WMS-III indexes. *Psychological Assessment*, 18(2), 174.
- Lavric, A., Forstmeier, S., & Rippon, G. (2000). Differences in working memory involvement in analytical and creative tasks: an ERP study. *NeuroReport*, 11(8), 1613–1618.
- Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: evidence from children. *British Journal of Developmental Psychology*, 21(1), 59–80.
- Lohman, D. F., & Gambrell, J. L. (2012). Using nonverbal tests to help identify academically talented children. *Journal of Psychoeducational Assessment*, 30(1), 25–44.
- Lowe, C., & Rabbitt, P. (1998). Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: Theoretical and practical issues. *Neuropsychologia*, 36(9), 915–923.
- Luciana, M., & Nelson, C. A. (1998). The functional emergence of prefrontally-guided working memory systems in four- to eight-year-old children. *Neuropsychologia*, 36(3), 273–293.
- Luciana, M., & Nelson, C. (2002). Assessment of neuropsychological function through use of the Cambridge Neuropsychological Testing Automated Battery: performance in 4- to 12-year-old children. *Developmental Neuropsychology*, 22(3), 595–624.
- Mammarella, I. C., Pazzaglia, F., & Comolli, C. (2008). Evidence for different components in children's visuospatial working memory. *British Journal of Developmental Psychology*, 26(3), 337–355.
- Mezzacappa, E. (2004). Alerting, orienting, and executive attention: developmental properties and sociodemographic correlates in an epidemiological sample of young, urban children. *Child Development*, 75(5), 1373–1386.
- Mitrushina, M., Boone, K. B., Razani, J., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment*. Oxford University Press.
- Naglieri, J. A. (2003). Naglieri Nonverbal Ability Tests. *Handbook of nonverbal assessment* (pp. 175–189). Springer.
- Naglieri, J. A., & Ford, D. Y. (2003). Addressing underrepresentation of gifted minority children using the Naglieri Nonverbal Ability Test (NNAT). *Gifted Child Quarterly*, 47(2), 155–160.
- Naglieri, J. A., & Ronning, M. E. (2000a). Comparison of White, African American, Hispanic, and Asian children on the Naglieri Nonverbal Ability Test. *Psychological Assessment*, 12(3), 328.
- Naglieri, J. A., & Ronning, M. E. (2000b). The relationship between general ability using the Naglieri Nonverbal Ability Test (NNAT) and Stanford Achievement Test (SAT) reading achievement. *Journal of Psychoeducational Assessment*, 18(3), 230–239.
- Noble, K. G., McCandliss, B. D., & Farah, M. J. (2007). Socioeconomic gradients predict individual differences in neurocognitive abilities. *Developmental Science*, 10(4), 464–480. <https://doi.org/10.1111/j.1467-7687.2007.00600.x>.
- Osborne, J. W. (2010). Improving your data transformations: applying the Box-Cox transformation. *Practical Assessment, Research & Evaluation*, 15(12), 1–9.
- 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(10), 1021–1034. [https://doi.org/10.1016/0028-3932\(90\)90137-D](https://doi.org/10.1016/0028-3932(90)90137-D).
- Ozonoff, S., Cook, I., Coon, H., Dawson, G., Joseph, R. M., Klin, A., . . . Pennington, B. F. (2004). Performance on Cambridge Neuropsychological Test Automated Battery subtests sensitive to frontal lobe function in people with autistic disorder: evidence from the Collaborative Programs of Excellence in Autism network. *Journal of Autism and Developmental Disorders*, 34(2), 139–150.
- Pangelinan, M. M., Zhang, G., VanMeter, J. W., Clark, J. E., Hatfield, B. D., & Haufler, A. J. (2011). Beyond age and gender: relationships between cortical and subcortical brain volume and cognitive-motor abilities in school-age children. *Neuroimage*, 54(4), 3093–3100.
- Piechatek, M., Indlekofer, F., Daamen, M., Glasmacher, C., Lieb, R., Pfister, H., . . . Schütz, C. G. (2009). Is moderate substance use associated with altered executive functioning in a population-based sample of young adults? *Human Psychopharmacology: Clinical and Experimental*, 24(8), 650–665. <https://doi.org/10.1002/hup.1069>.
- Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. *Cognitive Psychology*, 41(1), 1–48.
- Rojahn, J., & Naglieri, J. A. (2006). Developmental gender differences on the Naglieri Nonverbal Ability Test in a nationally normed sample of 5–17 year olds. *Intelligence*, 34(3), 253–260.
- Roque, D. T., Teixeira, R. A. A., Zachi, E. C., & Ventura, D. F. (2011). The use of the Cambridge Neuropsychological Test Automated Battery (CANTAB) in neuropsychological assessment: application in Brazilian research with control children and adults with neurological disorders. *Psychology & Neuroscience*, 4(2), 255.

- Rosselli, M., & Ardila, A. (2003). The impact of culture and education on non-verbal neuropsychological measurements: a critical review. *Brain and Cognition*, 52(3), 326–333.
- Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 298(1089), 199–209.
- Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: administration, norms, and commentary*. USA: Oxford University Press.
- Tallberg, I. M., Ivachova, E., Jones Tinghag, K., & Östberg, P. (2008). Swedish norms for word fluency tests: FAS, animals and verbs. *Scandinavian Journal of Psychology*, 49(5), 479–485.
- Toornstra, A., Hurks, P. P., Van der Elst, W., Kok, G., & Curfs, L. M. (2020). Measuring visual matching and short-term recognition memory with the CANTAB® Delayed Matching to Sample task in schoolchildren: effects of demographic influences, multiple outcome measures and regression-based normative data. *Child Neuropsychology*, 26(2), 189–218.
- World Medical Association. (2013). WMA Declaration of Helsinki-Ethical principles for medical research involving human subjects.
- Van Breukelen, G. J. P., & Vlaeyen, J. W. S. (2005). Norming clinical questionnaires with multiple regression: the Pain Cognition List. *Psychological Assessment*, 17(3), 336–344. <https://doi.org/10.1037/1040-3590.17.3.336>.
- Van der Elst, W., Van Boxtel, M. P., Van Breukelen, G. J., & Jolles, J. (2006). The Stroop color-word test: influence of age, sex, and education; and normative data for a large sample across the adult age range. *Assessment*, 13(1), 62–79.
- Van der Elst, W., Hurks, P., Wassenberg, R., Meijs, C., & Jolles, J. (2011). Animal verbal fluency and design fluency in school-aged children: effects of age, sex, and mean level of parental education, and regression-based normative data. *Journal of Clinical and Experimental Neuropsychology*, 33(9), 1005–1015. <https://doi.org/10.1080/13803395.2011.589509>.
- Van der Elst, W., Molenberghs, G., Hilgers, R. D., Verbeke, G., & Heussen, N. (2016). Estimating the reliability of repeatedly measured endpoints based on linear mixed-effects models. A tutorial. *Pharmaceutical Statistics*, 15(6), 486–493.
- White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, 91(3), 461.
- Zhou, M. (1998). *Empirical distributions*. Retrieved from <http://www.ms.uky.edu/%7Emai/java/stat/EmpDis.html>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.