Infant and Child Development
Inf. Child. Dev. (2013)
Published online in Wiley Online Library
(wileyonlinelibrary.com). DOI: 10.1002/icd.1833

Working Memory and Dynamic Measures of Analogical Reasoning as Predictors of Children's Math and Reading Achievement

Claire E. Stevenson[a,*], Catharina E. Bergwerff[a,b], Willem J. Heiser[a] and Wilma C. M. Resing[a]
[a]*Leiden University, Leiden, The Netherlands*
[b]*VU University Amsterdam, Amsterdam, The Netherlands*

Working memory and inductive reasoning ability each appear related to children's achievement in math and reading. Dynamic measures of reasoning, based on an assessment procedure including feedback, may provide additional predictive value. The aim of this study was to investigate whether working memory and dynamic measures of analogical reasoning are unique predictors of children's concurrent and subsequent reading and math achievement. School children (N = 188, M = 7.1 years, SD = 11 months) were administered a dynamic test of analogical reasoning comprising a pretest-training-posttest design. Pretest performance measures static reasoning ability, whereas posttest performance and feedback-needs during training are considered dynamic measures. Verbal and visuo-spatial working memories were assessed prior to dynamic testing. Performance on national reading and math achievement tests were gathered at two time points within one school year. A multilevel mixed-effects model indicated that verbal (but not visuo-spatial) working memory and dynamic reasoning measures formed unique predictors of concurrent and subsequent achievement in math and reading. Verbal working memory efficiency and performance on a dynamic test of analogical reasoning were both positively related to math and reading achievement in children in kindergarten, first grade and second grade. Dynamic assessment, in addition to working memory assessment, may be useful for educational psychologists when attempting to gauge children's future school performance. Copyright © 2013 John Wiley & Sons, Ltd.

*Correspondence to: Dr. Claire E. Stevenson, Faculty of Social Sciences, Institute of Psychology, Department of Methodology & Statistics, Leiden University, Wassenaarseweg 52, PO Box 9555, 2300 RB, Leiden, The Netherlands. E-mail: cstevenson@fsw.leidenuniv.nl

Key words: figural analogies; dynamic testing; school performance; analogical reasoning

INTRODUCTION

Working memory and inductive reasoning ability appear related to children's school achievement (e.g. Goswami, 1991; Krumm, Ziegler, & Buehner, 2008) and measures in both fields are employed by school psychologists to assess children's ability to learn. In the current study, the predictive value of working memory for academic success was compared to the predictive value of dynamic measures of inductive reasoning. There are several approaches to working memory, defined as the ability to process, store and retrieve information, but the cognitive model by Baddeley and Hitch is probably the most applied within the field of cognitive science. Within their model, four subsystems can be distinguished; the central executive and its slave systems, the phonological loop and the visuo-spatial sketchpad (Baddeley & Hitch, 1974), and the episodic buffer (Baddeley, 2000). The phonological loop appears essential for language acquisition and control of behaviour (Baddeley, 2003; Baddeley, Gathercole, & Papagno, 1998), whereas the visuo-spatial sketchpad plays an important role in explicit motor sequence learning (Bo & Seidler, 2009).

Working memory is considered essential to learning and related to children's school achievement (e.g. Alloway & Alloway, 2010; Alloway & Passolunghi, 2011; Krumm et al., 2008). The efficiency of working memory appears to be of good predictive value for academic success in general (e.g. Bull, Espy, & Wiebe, 2008; Gathercole, Pickering, Knight, & Stegmann, 2004; Holmes, Gathercole, & Dunning, 2009; St Clair-Thompson & Gathercole, 2006). Campbell, Dollaghan, Needleman, and Janosky (1997) and Weismer et al. (2000) found that verbal working memory ability is largely related to the development of language skills. Mainly verbal, but also visuo-spatial working memory have been found to predict reading achievement (Nevo & Breznitz, 2011, 2013). Verbal and visuo-spatial working memories were found to be powerful predictors of future math skills (Alloway & Alloway, 2010). Furthermore, verbal working memory appeared to be a good predictor of reading and math skills in children with reading disabilities (Gathercole, Alloway, Willis, & Adams, 2006) and reading and spelling in children with intellectual disabilities (Henry & Winfield, 2010). We therefore expected to find that verbal and visuo-spatial working memories provide a separate contribution to the prediction of reading and mathematics ability.

Much of learning in school is considered to be a form or analogical reasoning, in which knowledge about a familiar situation or object is used to learn about a new similar one (e.g. Goswami, 1992). Analogical reasoning is frequently assessed using classical analogy problems, such as the matrices used in the present study (see Figure 1). The ability to solve such matrices, often considered a measure of fluid reasoning, has been shown to be a good predictor of school achievement in both the reading (Ferrer et al., 2007; Stanovich, Cunningham, & Feeman, 1984) and math domain (Primi, Ferrão, & Almeida, 2010; Taub, Keith, Floyd, & McGrew, 2008).

Dynamic measures of analogical reasoning may, however, be an additional or perhaps better predictor of school performance. The repeated measures in dynamic testing allow us not only to measure reasoning ability at a certain point in time, but also assess learning at a micro-level, which may provide insight into learning and achievement at a macro-level (Siegler, 2006; Stevenson, 2012). Dynamic testing, often contrasted with static tests such as traditional cognitive ability measures, is an

DOI: 10.1002/icd

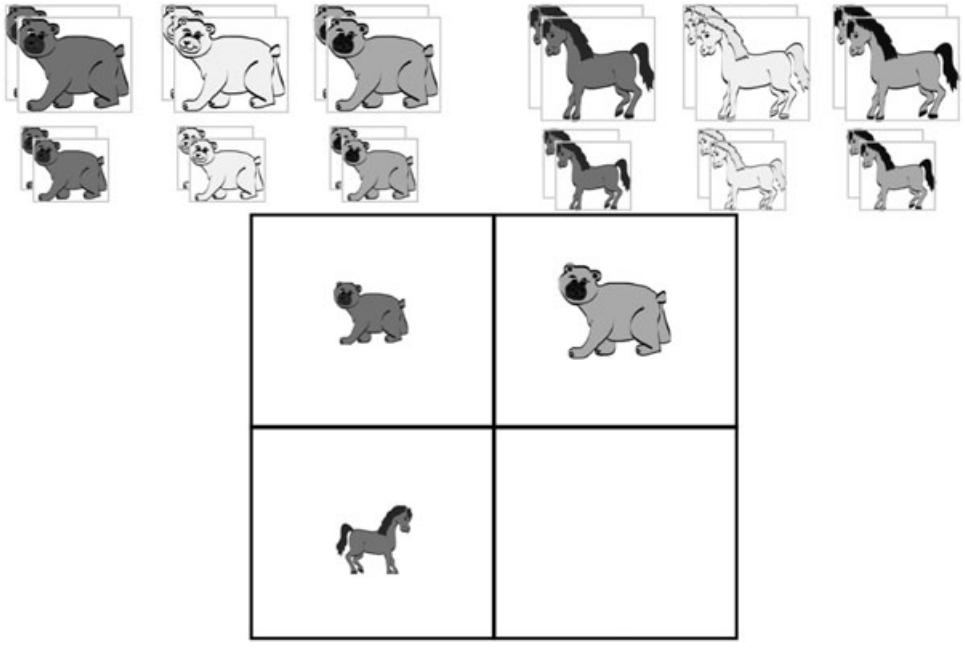

Figure 1. Example item from AnimaLogica.

assessment method in which feedback is incorporated into the procedure in order to facilitate learning and gain insight into learning efficiency and instructional-needs (Elliott, Grigorenko, & Resing, 2010).

The purpose of the current study was to investigate whether dynamic reasoning measures are additional predictors of children's present and future school achievement in reading and math, while taking the predictive value of both working memory and static reasoning ability into account. It is hypothesized that dynamic reasoning measures provide additional predictive value, as Krumm et al. (2008) found that working memory could not predict school success beyond reasoning ability. Similarly, Rohde and Thompson (2007) reported that working memory did not account for additional variance in academic achievement after controlling for general cognitive ability.

Earlier research has shown that dynamic measures of reasoning, stemming from dynamic tests, can provide additional predictive value of school achievement in reading (e.g. Fuchs, Compton, Fuchs, Bouton, & Caffrey, 2011; Swanson, 2011), math (e.g. Beckmann, 2006; Jeltova et al., 2011; Sittner Bridges & Catts, 2011) and other school achievement topics such as geography (Hessels, 2009). For example, in dynamic tests comprising a pretest-training-posttest design, performance change from pretest to posttest was demonstrated to be a unique predictor of children's reading skills (Swanson, 2011) and math achievement (Stevenson, Hickendorff, Resing, Heiser, & de Boeck, 2013). Furthermore, 'feedback-needs', i.e. the amount of feedback a child required to independently solve tasks during the training phase of a dynamic test, and posttest scores have been found to explain additional variance in the prediction of children's math and reading achievement scores (Resing, 1993). In the present study, both posttest performance and 'feedback-needs' are included as dynamic measures to predict children's school achievement.

However, not all studies show advantages of dynamic measures in predicting school achievement (e.g. Coventry, Byrne, Olson, Corley, & Samuelsson, 2011;

DOI: 10.1002/icd

Thatcher Kantor, Wagner, Torgesen, & Rashotte, 2011). Furthermore, investigations of the predictive value of dynamic measures do not always take working memory into account, which is important as working memory if often described as strongly related to inductive reasoning ability (e.g. Cho, Holyoak, & Cannon, 2007; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). Therefore, the present study aimed to extend these findings by comparing the predictive value of dynamic measures of inductive, analogical reasoning with that of verbal and visuo-spatial working memory.

Current Study

Many educational psychologists are interested in the future academic achievement of pupils and in the factors that influence academic performance so that instruction can be adapted to an individual's educational needs. Even though, in the past, researchers have found that either working memory or dynamic measures of inductive reasoning are related to math and reading scores, it remains unclear if these factors each have a unique contribution in the prediction of academic success. In the current study, primary school children were dynamically tested on a figural analogies task. Their performance on national school assessments of reading and math was collected at two time points within one school year. We tested the hypotheses that working memory, static and dynamic measures of figural analogical reasoning each form unique predictors of children's concurrent and subsequent reading and math achievement.

METHOD

Participants

One hundred and eighty eight children were recruited from five elementary schools in the south-west of the Netherlands. The sample consisted of 88 boys and 100 girls, with a mean age of 7 years, 1 month (SD = 11 months), from kindergarten, first grade and second grade (group two, three and four in the Dutch school system). The socio-economic status of the participants was determined by the educational level of the parents. Of all participants, 84 per cent had a middle to high socio-economic status, while 6 per cent had a low socio-economic status (determined by a low parental educational level) and 10 per cent had a very low socio-economic status (determined by a very low parental educational level). These percentages are comparable to the percentages of socio-economic status among the Dutch population (88, 8 and 4 per cent, respectively) (Central Bureau of Statistics, 2011). Written informed consent for children's participation was obtained from the parents. These children participated in a previous study on dynamic testing (Stevenson, Hickendorff et al., 2013) and were selected for the current study based on the availability of additional school achievement data. At Time 1, school achievement data was available for 188 children. At Time 2, six participants attended a different school, and therefore their school achievement scores could not be retrieved. Reading and math scores at Times 1 and 2 were not available for all participants because schools in the Netherlands may choose to administer only one rather than both of the subjects. This study was approved by the psychology ethics review committee of Leiden University.

Design and Procedure

The children were administered with a dynamic test of analogical reasoning, a verbal and a visuo-spatial working memory task and scholastic achievement tests in

DOI: 10.1002/icd

reading and math at halfway through the school year (Time 1). At the end of the school year, (Time 2), 6 months after Time 1, the reading and math achievement tests were administered again. The scholastic achievement tests were administered by the child's teacher in the classroom.

The dynamic test and working memory tasks were administered at Time 1 in five weekly sessions individually in a quiet room at the child's school. Each session lasted approximately 20 minutes. The first session consisted of verbal and visuo-spatial working memory tasks. The AnimaLogica dynamic test comprised a pretest-training and training-posttest design and was administered for the remaining four sessions. All instructions were provided according to standardized protocols by educational psychology students trained in the procedures (see Stevenson, Heiser, & Resing, 2013).

Instruments

Automated Working Memory Assessment (AWMA, Alloway, 2007)

Listening recall. In this verbal memory span task, the child heard a sequence of spoken sentences (e.g. 'bicycles can walk') and was asked to convey whether the sentence was true or false immediately following the sentence (e.g. 'false') and then to repeat the first word of each of the presented sentences (e.g. 'bicycles', …). The sequence length began with a single sentence and increased by one sentence after four correct trials within a block of six trials. The task was terminated if less than four 'first words' were correctly recalled. Scores were based on correctly recalled first words in the right order. The total number of correctly recalled sequences was scored. Scores were automatically standardized by the computer program to a mean of 100 and a standard deviation of 15 for each age band. The test-retest reliability of the Listening Recall subtest is 0.79 (Alloway, Gathercole, & Pickering, 2006).

Spatial span. In this visuo-spatial memory span task, the child was presented with sequences of two shapes that were either facing the same (i.e. rotated but not mirrored) or opposite (i.e. rotated and mirrored) direction. Immediately after each presented pair, the child was asked to convey whether the shapes were facing the same or the opposite direction. After each sequence of shape pairs, the child was asked to recall and point to the locations of the red dots that were displayed next to each right-hand shape. The sequences of shape pairs increased by one after four correct trials within a block of six trials (Alloway, Gathercole, Kirkwood, & Elliott, 2008). The task was terminated if less than four 'red dot locations' were correctly recalled. The total number of correctly recalled sequences was scored. Scores were automatically standardized by the computer program to a mean of 100 and a standard deviation of 15 for each age band. The test-retest reliability of the Spatial Span subtest is 0.82 (Alloway et al., 2006).

AnimaLogica: dynamic test of figural analogical reasoning

AnimaLogica is a computerized dynamic test of analogical reasoning for children (Stevenson, 2012). The figural analogies comprised of 2×2 matrices with familiar animals as objects (see Figure 1). The animals changed horizontally or vertically by colour, orientation, size, position, quantity or animal species. The children had to construct the solution using a computer mouse to drag and drop animal figures representing the six transformations into the empty box in the lower left or right quadrant of the matrix. A maximum of two animals were present in each cell of the analogy. These were available in three colours (red, yellow and blue), two sizes (large and small) and six species (lion, elephant, horse, camel, dog and bear). The orientation (facing left or right) could be changed by clicking the figure. The child could drag the

DOI: 10.1002/icd

animal or animals to the right position (on top, in the centre or on the foot of the matrix cell). Quantity was specified by the number of figures placed in the empty box.

Pretest and posttest. The tests consisted of 20 items of varied difficulty. The pretest and posttest contained item isomorphs – comprising the same transformations and difficulty, but different animals and colours. Cronbach's measures of internal consistency for the pretest and posttest for dataset this sample was based on were $\alpha = 0.90$ and $\alpha = 0.91$ respectively (N = 255) (Stevenson, Hickendorff et al., 2013). The item difficulty for the pretest ranges from 0.02 to 0.60 and for the posttest from 0.12 to 0.84. With regard to construct validity, the Rasch-scaled pretest scores correlated strongly with the Raven Standard Progressive Matrices (Raven, Raven, & Court, 2004), $r = .60$ (N = 253) in a separate sample of 7 year olds (Stevenson, 2012).

Training. The training consisted of the same figural analogy matrices. The ten training items did not occur in the pretest and posttest. Two training methods were applied: graduated prompts or outcome feedback (Stevenson, Hickendorff et al., 2013). Half of the participants (96 children) received a training based on graduated prompts, while the other half (92 children) received a training based on outcome feedback. The graduated prompts method (e.g. Campione & Brown, 1987; Resing & Elliott, 2011) consisted of stepwise instructions and began with general, metacognitive prompts, such as focusing attention, followed by cognitive hints, emphasizing the transformations and solution procedure, and ended with step-by-step scaffolds to solve the problem. A maximum of five prompts were administered. Outcome feedback training also allowed for five attempts to correctly solve each item. However, the children were only told if their solution was correct or incorrect and received motivational comments. For both forms of training, the examiner proceeded with the next item after a correct solution or five attempts. The reliability of the training items scale for the aggregated dataset was $\alpha = 0.84$ (N = 379).

Scoring. AnimaLogica provides three scores: (i) pretest, (ii) posttest and (iii) training. The correct/incorrect constructions of the figural analogies on the pretest and posttest were used to compute pretest and posttest ability scores on an item response theory scale (IRT, Embretson & Reise, 2000). IRT rather than a classical test theory (CTT) measure (i.e. proportion correct) was used because we measured individual performance growth over time. CTT scores are sensitive to bottom and ceiling effects, which can be further biassed due to the correlated nature of individual pretest and posttest scores (Embretson & Reise, 2000). The multidimensional Rasch estimates were reliably computed using an aggregated dataset (N = 514) with the lme4 package (Bates, Maechler, & Bolker, 2010) for R (e.g. De Boeck et al., 2011).

The training score quantifies the amount of feedback (max. five attempts per item) required by the child to solve each of the training items (Stevenson, 2012). The correct/incorrect solution of each attempt for each training item was used to compute a latent training score on a partial-credit IRT scale (e.g. Wang & Heffernan, 2013) using the generalized Partial Credit Model (gPCM, Muraki, 1992). A latent training score rather than a classical test theory (CTT) measure (i.e. total number of prompts) was used to account for differences in difficulty and prompt effectiveness between the items provided during training. The gPCM estimates were computed for the aggregated dataset (N = 379) using the ltm package for R (Rizopoulos, 2006). Although training type affected AnimaLogica outcomes, the two forms of training (outcome feedback and graduated prompts) did not differ in their predictive value on academic achievement and were thus collapsed onto one scale in this study.

DOI: 10.1002/icd

Standardized scholastic achievement tests

The children took part in biannual scholastic achievement assessments administered in January and June of each school year (Janssen, Verhelst, Engelen, & Scheltens, 2010; Jongen, Krom, van Onna, & Verhelst, 2011; Koerhuis & Keuning, 2011; Lansink & Hemker, 2012). These multiple-choice tests are widely used at primary schools in the Netherlands for the purpose of tracking children's performance on school subjects. The standardized tests are adapted to the educational curriculum for each grade, thus higher grades often encounter different or more complex problems than lower grades. For math, test versions differ per grade. For reading, there are two types of tests that measure technical reading skills for first and second grades: 3-minute test and reading technique and tempo (Jongen et al., 2011). These tests are strongly correlated and considered to measure a similar construct, as the main difference is that one involves reading out loud, whereas the other uses silent reading (Jongen et al., 2011). Schools may choose which test to administer to which class. The tests provide raw data (i.e. number of correct) and ability scores. The ability scores are based on test difficulty and the number of items but are not comparable for different test types. In order to compare the children's progression over time, we converted the ability scores on each of the test types to Z-scores using the population Means and Standard deviations reported per grade in the technical reports (Janssen et al., 2010; Jongen et al., 2011; Koerhuis & Keuning, 2011; Lansink & Hemker, 2012).

Statistical Models

Statistical models

Multilevel models were used to analyse which variables best predicted the children's school achievement across Times 1 and 2 for reading and math. This type of model was chosen because of the hierarchical structure of the data (test scores over two time points nested in children nested in test-type nested in schools) and that missing data was present as schools are allowed to choose which tests to administer and therefore for some children at certain time points only math scores and in other cases only reading scores were available. In the two sets of models (one for reading and one for math), the dependent variable was the test and grade-appropriate norm-referenced achievement Z-score at Times 1 and 2. Please note that the Mean and Standard deviation of the Z-scores were those of the national dataset per test version; therefore, in our models, we examine the children's performance relative to their peers as it is not possible to directly compare progression across different tests (e.g. pre-reading skills in kindergarten versus Technical Reading test A for grades 1 and 2 versus Technical Reading test B for grades 1 and 2).

Math and reading achievement were modelled separately. The multilevel models (Kreft & de Leeuw, 1998) were fitted using the lme4 package for R Statistical Software (Bates, Maechler, & Bolker, 2010). The model fitting steps were as follows. First, we determined the best model for the nested random effects. We found that a three-level model of measures over time nested within children nested within type of test administered per school (school–test-type) best fit the data. The highest level (school–test-type) accounted for differences in grades (because children in the different grades were administered different types of tests) and schools (type of test administered to a particular grade could differ between schools) and thus also classroom (each classroom was administered the same tests). Because these variables were present in one level, we were able to avoid a more complex model (such as classrooms nested in schools), yet still have a valid model to test our research questions that focused on the child level of the model. Random intercepts and slopes for the child

DOI: 10.1002/icd

level and school–test-type were conceptually required, resulting in the following basic model for reading and math:

$$y_{tij} = \beta_0 + \beta_1 + u_{0ij} + u_{1ij} + v_{0j} + v_{1j} + e_{tij} \quad \text{(M1)}$$

Where Y_{tij} denotes the reading or math score at time t ($t = 0$ or 1) for child i administered a particular grade-appropriate test at a particular school indicated by index j. The overall average intercept is represented by β_0, and the fixed effect (slope) of time is represented by β_1. The random intercepts and slopes over time per child nested within a specific school–test-type are represented by u_{0ij} and u_{1ij}, respectively; v_{0j} and v_{1j} represent the random intercepts and slopes from Time 0 to Time 1 per school–test-type and e_{tij} represents the residual error.

Second, in order to test our hypotheses, we added possible predictors of academic achievement as fixed effects at the child level one-by-one to model M1. The following predictors were tested: (i) verbal working memory score; (ii) visuo-spatial working memory score; (iii) AnimaLogica static pretest score; (iv) AnimaLogica dynamic posttest score and (v) AnimaLogica dynamic training score. The predictors were tested in the order of their correlation strength with the dependent variable (see Table 2). Likelihood ratio (LR) tests and comparisons of AIC and BIC fit indices (smaller is better) were used to assess how the model fit of the new, more complex, model compared to that of the previous, simpler, model. If the LR test was significant, then we could statistically infer that the added fixed effect was an additional predictor of reading or math achievement after controlling for random effects (person, test-type and school) and previously entered fixed effects. If the LR test was not significant and/or the AIC and BIC fit indices did not decrease, then we rejected the new model in favour of the less complex model (without the additional predictor). The results of the model comparisons for the fixed effects as well as the final models for both reading and math are presented in the sections on Reading achievement and Math achievement, respectively. The conclusions based on these models were very robust and did not change when different combinations or levels of random effects were chosen.

RESULTS

Descriptive Statistics

The descriptive statistics of each of the variables are presented in Table 1. The reading and math achievement Z-scores means fall slightly above the national average (0.0). The mean (dynamic) reasoning scores and working memory scores also fall slightly above the norm population mean (0.00 and 100, respectively); the moderately large standard deviations indicate much variability between individuals.

The correlations between each of the measures are presented in Table 2. Here, we see that the (dynamic) reasoning measures of the pretest, posttest and training were all strongly correlated. A moderately strong positive correlation was present between the two working memory measures. Also, the correlation between Times 1 and 2 for reading and also for math achievement was strong. The strong correlation of the achievement per subject over time indicates that children continued to perform similarly relative to their peers over the two measures.

Reading achievement at Time 1 was moderately related to dynamic analogy training performance, weakly related to the dynamic posttest and marginally related to verbal working memory. Each of these relationships was slightly stronger at Time 2. In the case of math achievement, both verbal and visuo-spatial working memories

DOI: 10.1002/icd

Table 1. Descriptive statistics of static and dynamic reasoning test scores, working memory scores (predictor variables) and reading and math achievement (dependent variables)

	N	Min.	Max.	M	SD
AnimaLogica[a]					
Pretest	188	0 (−2.84)	18 (5.00)	3.96 (0.45)	4.38 (1.85)
Posttest	188	0 (−2.77)	19 (2.95)	8.80 (0.24)	5.36 (1.12)
Training[b]	188	0 (−1.79)	50 (3.52)	18.48 (0.29)	13.33 (0.84)
Working memory					
Verbal	188	60	144	106.70	16.7
Visuo-spatial	188	59	140	106.34	20.6
Reading achievement[c]					
Reading time 1	187	−3.66	4.74	0.63	1.36
Reading time 2	157	−2.74	3.84	0.40	1.23
Math achievement[c]					
Math time 1	168	−2.79	3.76	0.38	1.27
Math time 2	162	−2.29	3.64	0.47	1.23

M, mean; SD, standard deviation.
[a]Raw scores and Rasch-based logit estimates (between parentheses) are presented. A person's logit score can be converted to the probability of correctly solving an item of average difficulty on the test using the following formula: probability correct = e^{score} / $(1 + e^{score})$.
[b]Raw score is the number of attempts to solve the training items, so here lower scores are better. The logit score represents ability on a latent scale so here a higher score indicates better performance.
[c]Nationally normed Z-scores were computed and are presented to allow for comparison across grades and test versions.

were moderately positively related at Times 1 and 2; although, the relationship between math achievement and verbal rather than visuo-spatial working memory was slightly stronger. Each of the three (dynamic) reasoning scores was significantly correlated (moderate to strong) with math achievement scores at Times 1 and 2 where the dynamic posttest measure showed the strongest association with math achievement of the three AnimaLogica measures.

Statistical Models of Achievement over Time

Reading achievement

The modelling steps of the hypothesized predictors of reading achievement, which were entered as fixed effects into model M1, are shown in Table 3. Here, we see that (i) the dynamic training score and (ii) verbal working memory (WM) each contributed to improved model fit and could therefore be considered significant additional predictors of reading achievement. Visuo-spatial working memory and the analogical reasoning pretest and posttest scores did not improve model fit and were therefore not considered additional predictors of reading achievement. The final model was based on 343 observations over 187 children for 11 test-type by school combinations. The variance of the random effects per child was $\sigma^2 = 1.08$ for Time 1 and $\sigma^2 = 0.80$ for Time 2. A very strong association between Times 1 and 2 ($r = 0.969$) was found. The school and type of test the child was administered also contributed to the variance in the data with the variance of the random effects per school–test-type combination being $\sigma^2 = 0.47$ at Time 1 and $\sigma^2 = 0.33$ at Time 2. Here, we see a moderate association between Times 1 and 2 ($r = 0.539$). The residual variance was $\sigma^2 = 0.18$ and overall model Deviance was 894.1. The effect of Time showed

DOI: 10.1002/icd

Table 2. Pearson correlations between static and dynamic reasoning test scores, working memory scores (predictor variables) and reading and math achievement (dependent variables). The corresponding N per comparison can be found below the reported *r*

		1	2	3	4	5	6	7	8	9
Dynamic test of analogical reasoning										
1	Pretest[a]	1	0.823**	0.766**	0.442**	0.417**	0.079	0.139+	0.506**	0.448**
		188	188	188	188	188	188	187	157	168
2	Posttest[a]		1	0.794**	0.494**	0.451**	0.155*	0.245**	0.595**	0.522**
			188	188	188	188	187	157	168	162
3	Training[a]			1	0.454**	0.371**	0.242**	0.283**	0.526**	0.487**
				188	188	188	187	157	168	162
Working memory										
4	Verbal				1	0.532**	0.137+	0.181*	0.363**	0.319**
					188	188	187	157	168	162
5	Visuo-spatial					1	0.046	0.123	0.328**	0.278**
						188	187	157	168	162
Reading achievement										
6	Read time 1						1	0.780**	0.467**	0.435**
							187	156	167	161
7	Read time 2							1	0.512**	0.394**
								157	157	156
Math achievement										
8	Math time 1								1	0.736**
									168	162
9	Math time 2									1
										162

+$p < 0.10$; *$p < 0.05$; **$p < 0.01$ (two-tailed)
[a]Rasch-based logit estimates were used.

Table 3. Overview of the model comparisons examining the fixed effects of hypothesized predictors of achievement

Model	Ref. model	Fixed effects	AIC	BIC	-LL	#p	*df*	*LR test*[a] *Λ*
Reading achievement								
M1		Time	931	965	456	9		
M2	M1	+ dynamic reasoning training	918	957	449	10	1	14.25***
M3	M2	+ dynamic reasoning posttest	920	963	449	11	1	0.01
M4	M2	+ verbal WM	916	958	447	11	1	4.30*
M5	M3	+ static reasoning pretest	917	963	447	12	1	1.11
M6	M3	+ visuo-spatial WM	917	963	446	12	1	1.35
Math achievement								
M1		Time	940	974	461	9		
M2	M1	+ dynamic reasoning posttest	870	908	425	10	1	71.61***
M3	M2	+ dynamic reasoning t raining	867	909	423	11	1	4.90*
M4	M3	+ static reasoning pretest	867	913	422	12	1	1.79
M5	M3	+ verbal WM	863	909	420	12	1	6.11*
M6	M4	+ visuo-spatial WM	862	912	418	13	1	2.55

AIC, Akaike information criteria; BIC, Bayesian information criteria; LL, log likelihood; LR, likelihood ratio; WM, working memory.
[a]The LR test comprises a comparison between the new model with # p parameters to the previous model as reference using the Chi-square distribution with *df* degrees of freedom.
***$p < 0.001$, **$p < 0.01$, *$p < 0.05$

that children's reading scores (achievement relative to peers) generally decreased from Time 1 to Time 2 (B = -0.27, SE = 0.20 t = -1.35, p = 0.20); however, this effect was not significant. The dynamic training score was the strongest additional predictor of reading achievement (B = 0.34, SE = 0.11, t = 3.04, p = 0.01) followed by verbal working memory (B = 0.19, SE = 0.09, t = 2.12, p = 0.06). These results indicate that generally children who required less feedback during training and had higher scores on the verbal working memory test also had higher reading achievement scores.

Math achievement

The steps taken to determine the fixed effects for the most parsimonious model of math achievement are shown in Table 3. Here, we see that (i) the dynamic posttest score; (ii) the dynamic training score and (iii) the verbal working memory (WM) each improved model fit and were therefore considered significant additional predictors of math achievement. Visuo-spatial working memory and the analogical reasoning pretest scores were not considered as additional predictors of math achievement as adding them as fixed effects did not improve model fit. The final model was based on 330 observations over 168 children for ten test-types nested within schools. The variances of the random effects at Time 1 were $\sigma^2 = 0.62$ for children and $\sigma^2 = 0.25$ for test-type within school. The random effects at Time 2 were $\sigma^2 = 0.72$ and $\sigma^2 = 0.20$ for children and test-type within school, respectively. A strong association between children's math achievement scores for Times 1 and 2 (r = 0.786) was found. A moderate to strong correlation was found between Times 1 and 2 for each test-type by school combination (r = *0.*601). Both correlations indicate that children's math achievement relative to peers was relatively stable from

DOI: 10.1002/icd

Time 1 to Time 2. The residual variance was $\sigma^2 = 0.18$, and the model deviance was 838.9. The fixed effect of time was not significant (B = 0.11, SE = 0.15, $t = 0.73$, $p = 0.48$). The dynamic posttest score was the strongest additional predictor of math achievement (B = 0.28, SE = 0.06, $t = 4.61$, $p < 0.001$) followed by the dynamic training score (B = 0.27, SE = 0.14, $t = 2.01$, $p = 0.07$) and verbal working memory (B = 0.19, SE = 0.08, $t = 2.53$, $p = 0.03$). These results indicate that children who had higher scores on the AnimaLogica posttest, required fewer prompts during training and had higher scores on the verbal working memory test most likely also had higher math achievement scores at Times 1 and 2.

DISCUSSION

The main aim of this study was to investigate the unique contributions of working memory, analogical reasoning ability and dynamic measures of analogical reasoning to the prediction of young children's school achievement in reading and math during one school year. Our results are in line with previous work and indicate that children with a more efficient working memory or better performance on a (dynamic) test of analogical reasoning generally obtained higher scores on assessments of math and reading achievement. With regard to working memory, specifically, we found verbal, but not visuo-spatial working memory, to be a unique predictor of reading and math achievement within the course of the school year. Verbal working memory has often been found to be a good predictor of both reading and math achievement in children (e.g. Gathercole et al., 2006; Nevo & Breznitz, 2011, 2013). However, our results with regard to visuo-spatial working memory are somewhat in contrast with previous findings in predicting children's achievement in reading (Alloway & Alloway, 2010; Swanson, 2011) and math (Alloway & Passolunghi, 2011). In our dataset, a weak relationship was certainly present, but this was overshadowed when accounting for performance on a dynamic test of analogical reasoning. Visuo-spatial storage components appear related to fluid reasoning (Hornung, Brunner, Reuter, & Martin, 2011) and learning to reason by analogy (Stevenson, Heiser et al., 2013; Tunteler & Resing, 2010). Thus, perhaps visuo-spatial working memory affects static and dynamic figural analogical reasoning, which in turn is a good predictor of academic success (e.g. Vock & Holling, 2008). Future research should investigate whether visuo-spatial reasoning plays a mediating or moderating role.

Matrix reasoning, as statically measured with the AnimaLogica pretest, on its own was found to be related to children's achievement over the course of the school year. This is supported by earlier research on the predictive value of fluid reasoning scores for math and reading achievement (e.g. Balboni, Naglieri, & Cubelli, 2010; Ferrer & McArdle, 2004; Primi et al., 2010). However, differences in working memory and in dynamic reasoning ability were more substantial predictors of children's achievement. For reading, 'feedback-needs', which represents how much help a child required to solve the training tasks during dynamic testing, was the strongest predictor of achievement over time after accounting for the role of school, classroom and the utilized reading measure. For math, the dynamic posttest measure, i.e. performance on the dynamic test after training, as well as 'feedback-needs', each formed unique predictors of present and subsequent achievement. These findings are in line with previous research on the predictive validity of dynamic testing for reading and math achievement, where posttest and/or training scores are additional or better predictors of statically administered measures (Beckmann, 2006; Jeltova et al., 2011; Resing, 1993; Stevenson, 2012). The contribution of this study lies in that

DOI: 10.1002/icd

dynamic measures of analogical reasoning were shown to be unique predictors of reading and math achievement in addition to verbal working memory while controlling for the hierarchical nature of the data.

A complicating factor in this study was using national tests as achievement measures normed per grade and measurement moment which only allowed analysis of progression relative to peers and not growth in the subject area. This is most likely why initial and subsequent achievement relative to peers is so strongly related in our models and indicates that ranking within the classroom did not change much between the two achievement measurement moments. Ideally, each of the students would be administered the same or parallel reading and math measures at each time point and the predictive value would be assessed on a continuous latent scale. This would also allow for longitudinal measurement of children's progression from one grade to the next and assist in disentangling the differential effects of working memory and dynamic reasoning on aptitude change. However, from our results, we can conclude that both verbal working memory and dynamic reasoning measures (posttest scores for math and training scores for reading and math) provide us with additional information on young children's achievement in reading and math during the school year relative to the national grade-appropriate average in both of these school subjects.

Our finding that measures stemming from a dynamic test of analogical reasoning forms unique additional predictors of math and reading achievement while accounting for differences in working memory efficiency and controlling for possible environment effects, such as school and classroom, adds to the growing evidence of the predictive value of dynamic testing measures in psycho-educational assessment (e.g. Beckmann, 2006; Caffrey, Fuchs, & Fuchs, 2008). Working memory and learning during dynamic testing appear to be separate constructs that each uniquely contribute to young children's school performance. In future studies, it is important to examine whether the predictive value of dynamic testing measures hold in comparison to other executive functions (e.g. cognitive flexibility, inhibition) which have also been shown to play a role in children's reasoning ability (Brydges, Reid, Fox, & Anderson, 2012) and academic attainment (St Clair-Thompson & Gathercole, 2006; Van der Sluis, De Jong, & Van der Leij, 2007).

ACKNOWLEDGEMENTS

We thank Aafke Snelting, Janneke de Ruiter, Nienke Faber, Margreet van Volkom, Isabelle Neerhout, Marit Ruijgrok, Bart Leenhouts, Noraly Snel and Rosa Alberto for their assistance in data collection and coding.

REFERENCES

Alloway, T. P. (2007). Automated working: memory assessment. London, UK: Pearson Assessment.

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106*(1), 20–29. doi: 10.1016/j.jecp.2009.11.003

Alloway, T. P., & Passolunghi, M. C. (2011). The relationship between working memory, IQ, and mathematical skills in children. *Learning and Individual Differences*, *21*(1), 133–137.

Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2008). Evaluating the validity of the automated working memory assessment. *Educational Psychology*, *28*(7), 725–734.

DOI: 10.1002/icd

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: are they separable? *Child Development*, *77*(6), 1698–1716.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.

Baddeley, A. D. (2003). Working memory and language: an overview. *Journal of Communication Disorders*, *36*(3), 189–208. doi: 10.1016/s0021-9924(03)00019-4

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), The psychology of learning and motivation (Vol. *8*, pp. 47–90). New York: Academic Press

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158–173. doi: 10.1037/0033-295x.105.1.158

Balboni, G., Naglieri, J. A., & Cubelli, R. (2010). Concurrent and predictive validity of the Raven Progressive Matrices and the Naglieri Nonverbal Ability Test. *Journal of Psychoeducational Assessment*, *28*(3), 222–235.

Bates, D., Maechler, M., & Bolker, B. (2010). Lme4: linear mixed-effects models using S4 classes., from http://CRAN.R-project.org/package=lme4

Beckmann, J. F. (2006). Superiority: always and everywhere? On some misconceptions in the validation of dynamic testing. *Educational and Child Psychology*, *23*(3), 35–49.

Bo, J., & Seidler, R. D. (2009). Visuospatial working memory capacity predicts the organization of acquired explicit motor sequences. *Journal of Neurophysiology*, *101*(6), 3116–3125. doi: 10.1152/jn.00006.2009

Brydges, C. R., Reid, C. L., Fox, A. M., & Anderson, M. (2012). A unitary executive function predicts intelligence in children. *Intelligence*, *40*(5), 458–469.

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology*, *33*(3), 205–228.

Caffrey, E., Fuchs, D., & Fuchs, L. S. (2008). The predictive validity of dynamic assessment a review. *The Journal of Special Education*, *41*(4), 254–270.

Campbell, T., Dollaghan, C., Needleman, H., & Janosky, J. (1997). Reducing bias in language assessment: processing-dependent measures. *Journal of Speech, Language, and Hearing Research*, *40*(3), 519–525.

Campione, J. C., & Brown, A. L. (1987). Linking dynamic assessment with school achievement. In C. S. Lidz (Ed.), Dynamic assessment: an interactional approach to evaluating learning potential (pp. 82–109). New York, U.S.A.: Guilford Press.

Central Bureau of Statistics [Centraal Bureau voor de Statistiek]. (2011). Kerncijfers bevolking [Population facts]. Voorburg, the Netherlands: Centraal Bureau voor de Statistiek.

Cho, S., Holyoak, K. J., & Cannon, T. D. (2007). Analogical reasoning in working memory: resources shared among relational integration, interference resolution, and maintenance. *Memory & Cognition*, *35*(6), 1445–1455.

Coventry, W. L., Byrne, B., Olson, R. K., Corley, R., & Samuelsson, S. (2011). Dynamic and static assessment of phonological awareness in preschool: a behavior-genetic study. *Journal of Learning Disabilities*, *44*(4), 322–329.

De Boeck, P., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., & Tuerlinckx, F. (2011). The estimation of item response models with the lmer function from the lme4 package in R. *Journal of Statistical Software*, *39*(12), 1–28.

Elliott, J. G., Grigorenko, E. L., & Resing, W. C. M. (2010). Dynamic assessment: the need for a dynamic approach. In P. Peterson, E. Baker & B. McGaw (Eds.), The international encyclopedia of education (pp. 220–225). Oxford: Elsevier.

Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Mahwah, N. J.: Erlbaum Publishers.

Ferrer, E., & McArdle, J. J. (2004). An experimental analysis of dynamic hypotheses about cognitive abilities and achievement from childhood to early adulthood. *Developmental Psychology*, *40*(6), 935.

Ferrer, E., McArdle, J. J., Shaywitz, B. A., Holahan, J. M., Marchione, K., & Shaywitz, S. E. (2007). Longitudinal models of developmental dynamics between reading and cognition from childhood to adolescence. *Developmental Psychology*, *43*(6), 1460–1473.

Fuchs, D., Compton, D. L., Fuchs, L. S., Bouton, B., & Caffrey, E. (2011). The construct and predictive validity of a dynamic assessment of young children learning to read: implications for RTI frameworks. *Journal of Learning Disabilities*, *44*(4), 339–347.

Gathercole, S. E., Alloway, T. P., Willis, C., & Adams, A.-M. (2006). Working memory in children with reading disabilities. *Journal of Experimental Child Psychology*, *93*(3), 265–281.

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, *18*(1), 1–16.

Goswami, U. (1991). Analogical reasoning: What develops? A review of research and theory. *Child development*, *62*(1), 1–22.

Goswami, U. (1992). Analogical reasoning in children. Hove, UK: Lawrence Erlbaum Associates.

Henry, L., & Winfield, J. (2010). Working memory and educational achievement in children with intellectual disabilities. *Journal of Intellectual Disability Research*, *54*, 354–365. doi: 10.1111/j.1365-2788.2010.01264.x

Hessels, M. G. P. (2009). Estimation of the predictive validity of the HART by means of a dynamic test of geography. *Journal of Cognitive Education and Psychology*, *8*(1), 5–21.

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*(4), 9–15.

Hornung, C., Brunner, M., Reuter, R. A. P., & Martin, R. (2011). Children's working memory: its structure and relationship to fluid intelligence. *Intelligence*, *39*(4), 210–221.

Janssen, J., Verhelst, R., Engelen, F., & Scheltens, F. (2010). Wetenschappelijke verantwoording van de toetsen LOVS Rekenen-Wiskunde voor groep 3 tot en met 8 [Scientific report of Student and educational tracking system of arithmetic-mathematics for grades 1-6]. Arnhem, the Netherlands: Cito.

Jeltova, I., Birney, D., Fredine, N., Jarvin, L., Sternberg, R. J., & Grigorenko, E. L. (2011). Making instruction and assessment responsive to diverse students' progress: group-administered dynamic assessment in teaching mathematics. *Journal of Learning Disabilities*, *44*(4), 381–395.

Jongen, I., Krom, R., van Onna, M., & Verhelst, N. (2011). Wetenschappelijke verantwoording van de toetsen Technisch Lezen voor groep 3 tot en met 5 uit LOVS [Scientific report of student and educational tracking system of technical reading for grades 1-3]. Arnhem, the Netherlands: Cito.

Koerhuis, I., & Keuning, J. (2011). Wetenschappelijk verantwoording van de toetsen Rekenen voor kleuters [Scientific report of Math test for kindergartners]. Arnhem, the Netherlands: Cito.

Kreft, I., & De Leeuw, J. (1998). Introducing multilevel modeling. London: UK Sage.

Krumm, S., Ziegler, M., & Buehner, M. (2008). Reasoning and working memory as predictors of school grades. *Learning and Individual Differences*, *18*(2), 248–257.

Lansink, N., & Hemker, B. (2012). Wetenschappelijke verantwoording van de toetsen Taal voor kleuters voor groep 1 & 2 uit het Cito volgsysteem primair onderwijs [Scientific report of Reading test for kindergartners from the Cito primary school tracking system]. Arnhem, the Netherlands: Cito.

Muraki, E. (1992). A generalized partial credit model: application of an EM algorithm. *Applied Psychological Measurement*, *16*(2), 159–176.

Nevo, E., & Breznitz, Z. (2011). Assessment of working memory components at 6 years of age as predictors of reading achievements a year later. *Journal of Experimental Child Psychology*, *109*(1), 73–90.

Nevo, E., & Breznitz, Z. (2013). The development of working memory from kindergarten to first grade in children with different decoding skills. *Journal of Experimental Child Psychology*, *114*(2), 217–228.

Primi, R., Ferrão, M. E., & Almeida, L. S. (2010). Fluid intelligence as a predictor of learning: a longitudinal multilevel approach applied to math. *Learning and Individual Differences*, *20*(5), 446–451.

Raven, J., Raven, J. C., & Court, J. H. (2004). Manual for Raven's Progressive Matrices and Vocabulary Scales. San Antonio, Texas: Harcourt Assessment.

Resing, W. C. M. (1993). Measuring inductive reasoning skills: the construction of a learning potential test. In J. H. M. Hamers, K. Sijtsma, & A. J. J. M. Ruijssenaars (Eds.), Learning

DOI: 10.1002/icd

potential assessment: theoretical, methodological and practical issues (pp. 219–242). Lisse, The Netherlands: Swets and Zeitlinger.
Resing, W. C. M., & Elliott, J. G. (2011). Dynamic testing with tangible electronics: measuring children's change in strategy use with a series completion task. *British Journal of Educational Psychology, 81*(4), 579–605.
Rizopoulos, D. (2006). ltm: an R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*, *17*(5), 1–25.
Rohde, T. E., & Thompson, L. A.. (2007). Predicting academic achievement with cognitive ability. *Intelligence*, *35*(1), 83–92.
Siegler, R. S. (2006). Microgenetic analyses of learning. In W. Damon, R. M. Lerner, D. Kuhn, & R. S. Siegler (Eds.), Handbook of child psychology (5th ed., pp. 464–510). Hoboken, NJ: Wiley Online Library.
Sittner Bridges, M., & Catts, H. W. (2011). The use of a dynamic screening of phonological awareness to predict risk for reading disabilities in kindergarten children. *Journal of Learning Disabilities*, *44*(4), 330–338.
St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology, 59*(4), 745–759.
Stanovich, K. E., Cunningham, A. E., & Feeman, D. J. (1984). Intelligence, cognitive skills, and early reading progress. *Reading Research Quarterly, 19*(3), 278–303.
Stevenson, C. E. (2012). Puzzling with potential: dynamic testing of analogical reasoning in children. Amsterdam: Leiden University.
Stevenson, C. E., Heiser, W. J. H., & Resing, W. C. M. (2013). Working memory as a moderator of training and transfer of analogical reasoning in children. *Contemporary Educational Psychology, 38*, 159–169. doi: 10.1016/j.cedpsych.2013.02.001
Stevenson, C. E., Hickendorff, M., Resing, W. C. M., Heiser, W. J. H., & de Boeck, P. A. L. (2013). Explanatory item response modeling of children's change on a dynamic test of analogical reasoning. *Intelligence*, *41*(3), 157–168.
Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence*, *30*(3), 261–288.
Swanson, H. L. (2011). Dynamic testing, working memory, and reading comprehension growth in children with reading disabilities. *Journal of Learning Disabilities*, *44*(4), 358–371.
Taub, G. E., Keith, T. Z., Floyd, R. G., & McGrew, K. S. (2008). Effects of general and broad cognitive abilities on mathematics achievement. *School Psychology Quarterly, 23*(2), 187.
Thatcher Kantor, P., Wagner, R. K. , Torgesen, J. K., & Rashotte, C. A. (2011). Comparing two forms of dynamic assessment and traditional assessment of preschool phonological awareness. *Journal of Learning Disabilities*, *44*(4), 313–321.
Tunteler, E., & Resing, W. C. M. (2010). The effects of self-and other-scaffolding on progression and variation in children's geometric analogy performance: a microgenetic research. *Journal of Cognitive Education and Psychology, 9*(3), 251–272.
Van der Sluis, S., De Jong, P. F., & Van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence*, *35*(5), 427–449.
Vock, M., & Holling, H. (2008). The measurement of visuo–spatial and verbal–numerical working memory: development of IRT-based scales. *Intelligence*, *36*(2), 161–182.
Wang, Y., & Heffernan, N. T. (2013). Extending knowledge tracing to allow partial credit: using continuous versus binary nodes. Paper presented at the Artificial Intelligence in Education Conference, Memphis, TN, USA.
Weismer, S. E., Tomblin, J. B., Zhang, X. Y., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research*, *43*(4), 865–878.

DOI: 10.1002/icd