Sustained attention, not procedural learning, is a predictor of reading, grammatical and arithmetic skills in children

Gillian West[1]

David R. Shanks[2]

Charles Hulme[1]

[1] Department of Education, University of Oxford, UK

[2] Department of Experimental Psychology, University College London, UK

Gillian West (corresponding author):

ORCID ID:0000-0002-4413-3601

Department of Education, University of Oxford, UK

Email: gillian.west@education.ox.ac.uk

1

Abstract

The procedural deficit hypothesis claims that impaired procedural learning is a causal risk factor for developmental dyslexia and developmental language disorder. We investigated the relationships between measures of basic cognitive processes (declarative learning, procedural learning and attention) and measures of attainment (reading, grammar and arithmetic) in a large sample of 7- and 8-year-old children. A latent variable path model showed that verbal declarative memory skills predicted attainment but were not significantly related to attention. Procedural learning was only weakly related to measures of attainment and attention assessed during the procedural learning task accounted entirely for its relationship with measures of attainment. Our results challenge the procedural deficit hypothesis of reading and language disorders, but suggest that attentional skills (rather than procedural learning ability per se) may be an important predictor of reading, arithmetic and grammatical skills.

KEYWORDS

Sustained attention, not procedural learning, is a predictor of reading, grammatical and arithmetic skills in children

The procedural deficit hypothesis claims that impaired procedural learning is one cause of the deficits found in developmental dyslexia and developmental language disorder (Nicolson & Fawcett, 2007; Ullman & Pierpont, 2005), as well as mathematics disorder (Evans & Ullman, 2016). The theory subscribes to a dual-process view of memory (Squire, 2004). Declarative memory is claimed to be responsible for storing word-specific knowledge, while unconscious, automatic procedural memory processes underpin the learning of statistically-regular, rule-based aspects of language (Ullman, 2004). The procedural deficit hypothesis sees procedural sequence learning specifically as a critical skill underlying the acquisition of sequential or probabilistically-structured linguistic structures (Christiansen, Dale, Ellefson, & Conway, 2002).

It is important to emphasize here that developmental language disorder and developmental dyslexia are quite distinct disorders, although with considerable overlap and frequent comorbidity (Bishop & Snowling, 2004; Catts, Adlof, Hogan, & Weismer, 2005; Krishnan, Watkins, & Bishop, 2016). We also note that the terminology used to refer to children with oral language difficulties is currently in flux; the studies we review here have typically referred to this group of children as having "specific language impairment" or simply "language impairment", but in the current paper we use the more recently adopted term "developmental language disorder" for this group of children. Both reading and writing are scaffolded on oral language (Fletcher, 2009) and early problems with oral language place children at greater risk of problems with literacy later on. Critically for the theory of a common procedural memory deficit across developmental disorders, however, the procedural memory system allegedly accounts for the seemingly effortless retrieval of grammar-related and phonological information required for fluency in both language and reading (Ullman, 2004; Nicolson & Fawcett, 2007). This theory has also been extended to explain

3

mathematics disorder. Mathematical difficulties are frequently comorbid with dyslexia (Landerl &

Moll, 2010) and arithmetical ability relates to both general verbal ability, phonological processing

(Vukovic & Lesaux, 2013) and reading skill (Korpipää et all., 2017). However, just as in the case of

language and reading skills, arithmetical skills, such as counting, addition and subtraction, become

automatized over time and may depend (in part) upon procedural learning (Van Daal, van der Leij,

& Ader, 2013).

The most widely used measure of procedural learning in studies of the procedural deficit

hypothesis is the Serial Reaction Time task (SRT: Nissen & Bullemer, 1987). This task measures

response times to a stimulus that appears in different locations on a computer display, either

randomly, or according to a covert sequence. Faster response times to trials that conform to the

sequence compared to random trials are taken as evidence of procedural learning. However, results

are inconsistent, with some studies finding evidence of impaired procedural learning on the SRT

task in children with dyslexia and language impairment (e.g. Stoodley, Ray, Jack, & Stein, 2008)

and others finding no evidence of such impairments (e.g. Gabriel, Maillart, Guillaume, Stefaniak, &

Meulemans, 2011). Reviews of this literature have reached contradictory conclusions. For example,

Lum, Ullman, and Conti-Ramsden (2013) reported impaired performance on the serial reaction time

task in dyslexic, compared to age-matched control groups, in a meta-analysis of 14 group design

studies (mean effect size d = .45, 95% CI [.20, .69], p < .001). In contrast, Schmalz, Altoe, and

Mulatti (2017) considered a meta-analysis infeasible owing to the variety of dependent variables

used in different studies and the scarcity of reported condition averages from which to gauge effect

sizes. Their review concluded that the evidence from serial reaction time tasks for a procedural

learning impairment in dyslexia was "underwhelming" (p. 10).

Another limitation of studies in this field is that most have used deterministically-structured

SRT tasks, which use a fixed sequence of locations. However, these tasks have been criticized for

not fully dissociating procedural and declarative learning, because the fixed sequence structure leaves open the possibility that explicit learning is involved (Shanks & Johnstone, 1999). For example, participants on such tasks display above chance performance in post-task "generation" tests and, therefore, cannot be said to be completely unaware of the structure of the sequences learned (Shanks, Green & Kolodny, 1994). Deterministically-structured tasks also take a measure of procedural learning towards the end of the task, with the corresponding risk of fatigue effects confounding results. Versions that alternate sequenced and random locations throughout the task (Hedenius, et al., 2011; 2013; Howard, Howard, Japikse, & Eden, 2006) and probabilistically-structured (Desmottes, Meulemans, & Maillart, 2016a; 2016b; Desmottes, Maillart, & Meulemans, 2017; Henderson & Warmington, 2017) versions have also been developed, which use more complex sequential structures that render the sequence less discernable. The proportion of unaware participants on these more complex tasks is far higher, yet they still show significant procedural learning (Cleeremans & McClelland, 1991; Howard, Howard Jr, Japikse, DiYanni, Thompson, & Somberg, 2004; Schvaneveldt & Gomez, 1998). The risk of fatigue effects in such tasks is also reduced as they enable a measure of procedural learning to be taken across the whole task. For these reasons, in the current study we used a probabilistically structured SRT task that minimizes the possible role of declarative memory processes.

A review of previous research suggests there may be a number of reasons why the results from studies investigating the relationship between SRT tasks and measures of reading and language skills are so mixed.

1. The majority of studies use extreme group designs, comparing children with dyslexia or developmental language disorder with an age-matched control group with normal language skills. Extreme group designs are subject to a number of methodological limitations (Preacher, Rucker, MacCallum, & Nicewander, 2005), such as regression to the mean for extreme scores,

misclassification into groups, inflated effect sizes and decreased reliability. Sample sizes are also typically small, further compounding these issues. Studies investigating individual differences in procedural (or statistical) learning and reading in large unselected samples are relatively scarce, although less prone to such methodological limitations (e.g. Kidd & Kirjavainen, 2011; Schmalz, Moll, Mulatti, & Schulte-Korne, 2019; West, Vadillo, Shanks, & Hulme, 2018).

2. Reliability of the SRT task may be poor. Difference scores, such as those used to measure procedural learning on the SRT task tend to be unreliable (Lord, 1958; Overall & Woodward, 1975). Additionally, performance on procedural learning tasks (such as the SRT task) typically reflects a host of perceptual, motoric and other cognitive processes, as well as procedural learning. The relative contribution of procedural learning to overall performance on such tasks, therefore, tends to be low which means that the reliability of such a measure of procedural learning will be low too (Ostergaard, 1998). Previous studies have not reported the reliability of the serial reaction time tasks used, but in a recent study we found that the reliability of a 500 trial probabilistic SRT task in a large sample of children was extremely low (test-retest reliability $r = .21$; West et al., 2018).

3. The putative poor reliability of the SRT tasks typically used in studies of children with dyslexia or developmental language disorder may in part be due to their length. Procedural learning tasks in children are typically kept short in an effort to maintain motivation and avoid fatigue (e.g. Perlant & Largy, 2011). However, task length is usually an important determinant of reliability (Nunally & Bernstein, 1994) and the small number of trials used in many SRT tasks with children is a cause for concern.

4. Finally, self-regulatory mechanisms develop with age (Putzke, Williams, Adams, & Boll, 1998), so unreliable procedural learning scores may be a particular issue in children, where levels of sustained attention may vary and influence learning. This is potentially a serious issue in studies of

children with dyslexia or developmental language disorder, since these disorders are frequently comorbid with attentional difficulties (Ebert & Kohnert, 2011; Finneran, Francis, & Leonard, 2009) and sustained attention has been suggested as a potential predictor of the development of reading skills (Rabiner, Coie, & The Conduct Problems Prevention Research Group, 2000), as well as language (Ebert, & Kohnert, 2011) and mathematical ability (Anobile, Stievano, & Burr, 2013).

A few studies examining reading ability or dyslexia and statistical learning have included independent measures of sustained attention (e.g. von Koss Torkildsen, Arciuli, & Wie, 2019; van Witteloostuijn, Boersma, Wijnen, & Rispens, 2019) but so far have found no evidence of a relationship between statistical learning and attention. However, no studies have yet investigated the potential effects of variable levels of sustained attention on the procedural learning task itself, in spite of findings that manipulating levels of attention paid to statistical learning tasks does have downstream effects on learning (Baker, Olson, & Behrmann, 2004; Toro, Sinnett, & Soto-Faraco, 2005). The serial reaction time task is a long and attentionally-demanding task. It is, therefore, important to examine whether variations in attention on the task may explain any relationship with reading and language ability.

In summary, impaired procedural learning has been suggested as a possible causal risk factor for both developmental dyslexia and developmental language disorder, but research to date has produced inconsistent results. The current study uses a concurrent correlational design with a large sample of children unselected for ability. This design is not subject to the methodological limitations of extreme group designs. Furthermore, we will also assess the extent to which variations in both procedural and declarative memory skills correlate with attainment in reading, language and arithmetic. We assess procedural learning using an SRT task with a large number of trials in an attempt to ensure that the task is reliable. By testing all participants on two separate occasions, we can also assess the test-retest reliability of this task. This is particularly important

given that the reliability of this task typically has not been reported in studies in this area, and in a previous study was found to be low (West et al., 2018). We relate measures of language-related attainment (reading, grammar and arithmetic) to performance on the SRT task. We also include measures of verbal declarative memory. This is necessary in order to assess whether procedural learning and declarative learning are separable constructs and whether they relate differentially to reading or language problems. If a procedural learning impairment is a causal risk factor for both developmental dyslexia and developmental language disorder, we would expect a correlation between procedural learning on the SRT task and measures of attainment in both reading and language. Finally, we use a novel procedure of rating each child's attention to the SRT task on both test occasions. This allows us to investigate both the extent to which variable attention levels on the SRT task may influence procedural learning and whether attention per se, as well as or instead of performance on the SRT task itself, is a reliable correlate of reading and language skills.

Method

A concurrent correlational design was used to explore the relationships between procedural and declarative memory skills, attention and attainment in 7- and 8-year-old children. Ethical approval for the study was granted by the UCL Research Ethics committee and complied with the Declaration of Helsinki. Head teachers gave written informed consent for children to take part, and the parents of each child were given the option of withdrawing their child from the study before it began. Children were told that they could withdraw from the study at any time, should they wish to.

Participants.

One hundred and twelve children from four state primary schools in London, UK, unselected for ability, took part in the study (mean age = 92. 11 months; SD = 5.71; range 83 – 105

months; 53 girls and 59 boys). This age group was considered old enough to complete the task battery and is in the same age range as samples in several studies of procedural learning in children with developmental language disorder (e.g. Lum, Gelgic, & Conti-Ramsden, 2010; Lum & Bleses, 2012; Hsu & Bishop, 2014; Jimenez-Fernandez, Vaquero, Jimenez, & Defior, 2011). Children in participating schools were enrolled on an opt-out basis, with only 5 children across all four schools opting out. Age standard scores on the TOWRE Sight Word Efficiency subtest ranged from 76 to 136 and on the Phonemic Decoding Efficiency subtest from 68 to 133, indicating a very wide range of reading ability in the sample.

Tasks and testing procedures.

Testing took place across three sessions. Tasks were administered in the same fixed order to all children by the same experimenter. Children first completed a battery of attainment measures during a whole-class testing session followed by the individual tests. Much evidence suggests that scores on different language tests load on a single factor (Klem, Melby-Lervåg, Hagvet, Lyster, Gustafsson, & Hulme, 2015; Tomblin and Zhang, 2006). Therefore, because our main focus was on reading, and because of time constraints, we used the Test of Reception of Grammar (TROG-2) as our sole measure of language ability. This test assesses grammatical skills, which it has been claimed are particularly dependent of procedural learning (Ullman & Pierpoint, 2005).

Test of reception of grammar (TROG-2: Bishop, 2003). Children matched spoken sentences to one of four pictures. The 80-item task was customized to allow for group administration. Children marked pictures in a response booklet that corresponded to sentences spoken by the experimenter.

Test of word and non-word reading efficiency (TOWRE-2: Torgesen, Wagner, & Rashotte, 1999). Children read aloud as many words (or non-words) as possible in 45 seconds. These tests were administered individually.

Test of basic arithmetic and number skills (TOBANS: Brigstocke, Moll, & Hulme, 2016). These timed group-administered tests assessed fluency in addition, subtraction, and multiplication. They had no reading requirement, with all instructions read aloud to the children.

Two individual sessions for each child were then scheduled a week apart to complete the following declarative and procedural memory measures.

Word Lists (Cohen, 1997). This measure of verbal free recall from the Children's Memory Scale required children to recall as many words as possible in any order from a list of 10 spoken words (Trial 1). After the first presentation only words omitted by the children were re-presented (Trials 2 – 4). After a distractor list children were asked to recall the initial list again without re-presentation (Trial 5). These trials summed to form the child's Learning Score. Recall was requested once more at the end of the testing session (Trial 6: Delayed Recall) and during the second individual testing session a week later (Trial 7: Consolidation). The testing lapse for the consolidation measure was not the same for all children owing to scheduling constraints (93% of children were assessed with a lag of between 6 and 8 days and 7% of children with a lag of 2 or 5 days), but there was no significant difference in consolidation scores for those with a smaller or greater testing interval.

Serial Reaction Time Task (SRT: Nissen & Bullemer, 1987). A probabilistic task, based on the procedure introduced by Schwaneveldt and Gomez (1998), was used to examine non-verbal procedural spatial sequence learning.

Children sat in front of a laptop connected to an Xbox Gamepad controller. On each trial a yellow smiley face stimulus was displayed on the screen in 1 of 4 locations according to the sequence probabilities of the task. The locations formed a diamond pattern that matched the pattern of buttons on the Gamepad controller. The children were told to press the button that corresponded to the position of each stimulus as quickly as possible.

Children were given a short break after every 200 trials and given general encouragement to focus accuracy and maintain motivation. The program recorded response times and accuracy, requiring the child to press the correct button before proceeding to the next trial. The interval between trials was 250ms. The task took 20-25 minutes to complete. Faster response times for probable compared to improbable trials were taken as evidence of procedural learning.

In order to examine test-retest reliability two serial reaction time tasks were created. For the first task two second order conditional 12-item sequences were taken from Shanks, Wilkinson, and Channon (2003): sequence A – 314324213412; sequence B – 431241321423. Both sequences repeated each location 3 times, each time preceded by a different location; each sequence had one reversal (242 or 343) and there were no repeated locations. The second task used sequences with an identical second order conditional structure to those in the first task, taken from Schwaneveldt and Gomez (1998: sequence A: 121342314324; sequence B: 123413214243). In all other ways the two tasks were identical.

Task length was set at 1000 trials following a pilot of a 1500 trial SRT task with 46 undergraduates, since test-retest reliability for the first 1000 trials of the pilot task was high ($r = .70$) and showed only a small improvement when computed over all 1500 trials (overall pilot reliability $r = .71$).

The task began with 10 practice trials, with equal probabilities of each sequence occurring. There were then 10 blocks of 100 trials (NB: for the purposes of analysis, these 10 blocks were subsequently divided into 5 epochs, each containing 200 trials). Each block started with a randomly chosen bigram, e.g. 32. The next location either followed that bigram in sequence A (with a probability of .9, i.e., 4), or in sequence B (with a probability of .1, i.e., 1). Hence the transitions in sequence A were highly probable, while those in sequence B were correspondingly improbable. This process then continued with the new most recent bigram and so on to the end of the block.

Attention during the SRT task. In order to quantify the level of attention paid by each child to the SRT tasks a 9-point rating scale was used.

1 = Did not complete Epoch

2 = Poor attention throughout; continual stopping, talking, difficulty with restarting

3 = Distracted multiple times throughout with talking and short pauses (7 or more)

4 = Some talking and short pauses throughout, getting worse towards the end (4 – 6)

5 = Occasional chatting and short pauses (3 or less)

6 = Good attention, fading slightly towards the end (2 or fewer short pauses towards the end)

7 = Good attention throughout with occasional comments and no real pauses

8 = Good attention throughout

9 = Accurate, highly focussed on the task

An attention score based on this scale was allocated to each epoch of both tasks by the experimenter during testing and averaged for each task.

Scoring and analysis.

First, performance on the SRT tasks was analysed using mixed effects regression models (hierarchical linear models), in which both items and participants were treated as crossed random effects (see Baayen, Davidson, & Bates (2008) for a detailed explanation of this type of model).

Procedural learning scores for each child were then derived from performance on each of the two SRT tasks separately. The measure chosen to represent procedural learning was the overall proportional mean difference in RT between the probable and improbable sequences. This measure controlled for children's baseline speed of response by dividing the difference between their speed on improbable and probable trials by their overall response speed (improbable sequence RT – probable sequence RT) / ((improbable sequence RT + probable sequence RT)/2). A positive score, therefore, reflects procedural learning. Finally, latent variable path analysis was used to assess the relationship between measures of procedural and declarative learning, attention, reading, language and arithmetic. This statistical approach has the advantage of controlling for measurement error (see McCoach, Black, and O'Connell (2007) for an overview).

Results

Learning on the SRT tasks.

Results for two children were missing for SRT1 due to computer malfunction and one for SRT2 due to absence. Inaccurate trials, RTs for the first bigram in each block, and those over 5000ms were removed. In addition, a moving criterion based on sample size (Selst & Jolicoeur, 1994) was used to remove outlying RTs to ensure that results were not biased by the presence of an unequal number of predictable and unpredictable RTs across the task (trials removed: SRT1 probable sequence = 7896 (8.2% of probable trials); SRT1 improbable sequence = 1374 (12.5% of improbable trials); SRT2 probable sequence = 7247 (7.4% of probable trials); SRT2 improbable sequence = 1509 (13.6% of improbable trials)). The remaining response times for each task were analysed using mixed effects models in which both items and participants were treated as crossed random effects (see Baayen, Davidson, & Bates, 2008) to take account of variability across participants and items. Sequence type (probable or improbable), epoch and the interaction between the two were entered as fixed effects. Mean RTs for each sequence type (probable vs. improbable)

13

and epoch are shown in Figure 1. It is clear from Figure 1, and from the error rates, that our procedures for the SRT task were successful in maintaining children's attention. Error rates are fairly low; furthermore, RTs on the task decrease across epochs and there is evidence of procedural learning (faster responses to probable than improbable sequences) on both test occasions (though learning effects are stronger on the first test). These effects are in line in line with earlier studies using similar tasks. The variability of RTs across epochs (as indexed by the confidence intervals) are also stable. These patterns in the data confirm that there was appreciable procedural learning on the SRT task and indicate that our protocol succeeded in maintaining children's attention to the task.

< Insert Figure 1 here >

For both tasks RTs were slower on improbable trials (SRT1: marginal mean difference = 54 ms [95% CI 43, 65], $z = 9.42$, $p < .001$, SRT2: marginal mean difference = 53 ms [95% CI 44, 62], $z = 11.96$, $p < .001$). RTs became faster across epochs (SRT1 Epoch 1 - 5: unstandardized slope = -141 ms/epoch [95% CI -146, -136], $z = -54.76$, $p < .001$; SRT2 Epoch 1 - 5: unstandardized slope = -21 ms/epoch [95% CI -24, -17], $z = -10.43$, $p < .001$). Crucially, the interaction between sequence type and epoch was also significant on all epochs of SRT1 and the last three epochs of SRT2. These interactions confirm what can be seen in Figure 1: the difference in reaction time between the probable and improbable trials increases from earlier to later epochs of both tasks. This difference in RT provides evidence that the children are learning the structure of the probable trials.

Procedural learning scores for each child were derived from performance on the two SRT tasks. These scores indicated the presence of procedural learning in the majority of participants (108 children had positive scores on the first task and 110 on the second task). Split-half reliability for the procedural learning scores was calculated by sequentially numbering the trials for each

sequence and calculating a proportional difference score measure for odd and evenly numbered trials separately. Test-retest reliability for the task was assessed by correlating children's difference scores on the first and second time of taking the task. Split half reliability for the procedural learning score was moderate (SRT1 $r = .51$; SRT2 $r = .62$) though test-retest reliability was much lower ($r = .26$). Means, standard deviations and reliabilities for all measures are given in Table 1 and correlations between all measures in Table 2. Attention to the task varied between children, but was generally good overall. Attention ratings were relatively stable across epochs 1-5 on both attempts: SRT1 epoch 1 – 5, Mean (SD): 7.40 (1.33), 7.14 (1.45), 7.13 (1.56), 7.28 (1.51), 7.31 (1.57); SRT2 epoch 1 – 5, Mean (SD): 7.65 (1.25), 7.22 (1.66), 7.09 (1.67), 7.18 (1.67), 7.00 (1.71)).

< Insert Table 1 here >

< Insert Table 2 here >

Predicting individual differences in attainment from measures of procedural and declarative memory.

The principal objective of this study was to assess the relationships between our measures of basic cognitive skills (procedural learning [SRT]), attention during the SRT task, and declarative memory (free recall) on the one hand and our measures of attainment (Reading [TOWRE word reading, TOWRE nonword reading]), Arithmetic, and Grammatical skills (Test of Reception of Grammar) on the other. For this purpose, we estimated the latent variable path model shown in Figure 2. Modelling was conducted in Mplus 7.31 (Muthén & Muthén, 1998-2016) with the very small amount of missing data being handled with Full Information Maximum Likelihood estimation. Latent variables here allow us to estimate the relationship between constructs after excluding measurement error. This is particularly critical for tasks measuring procedural learning,

where reliability is low. The latent variable for procedural learning was defined by the shared variance between the two overall proportional mean difference scores for each child, one for each SRT task. The latent variable for attention similarly reflected the shared variance between each child's attention score on the two SRT tasks.

We began with a saturated model in which each of the predictors (procedural learning in the SRT task, attention during the SRT task, and declarative memory [verbal free recall]) predicted each of the measures of attainment (reading, grammar, and arithmetic). A single measure was used to represent the latent variable of grammar and we used the reliability of this measure to estimate measurement error (Hayduk & Litvay, 2012).

Non-significant relationships (paths) between the predictors and measures of attainment were dropped iteratively from the saturated model and changes in χ^2 were used to check that there was no significant loss of model fit as a result of omitting each of the non-significant paths. The final simplified model retaining the significant paths is shown in Figure 2, along with their path weights; the parameter estimates shown are the standardized coefficients and correlations. The model gives a good fit to the data, $\chi^2(79) = 94.526$, $p = .112$; RMSEA = .042 [90% CI 0.000 0.071]; CFI = .98; TLI = .98 (for an explanation of fit indices see Kenny, 2015).

< Insert Figure 2 here >

One notable feature of this model is that procedural learning on the serial reaction time task is quite strongly correlated ($r = .56$) with ratings of children's attention while performing the task. The other critical features of this model are the patterns of prediction from the SRT task, attention on the SRT task and declarative learning to measures of attainment (reading, grammatical and arithmetic skills). Procedural learning as measured by the SRT task is not a significant predictor of reading, grammatical or arithmetic skills. In contrast, attention during the SRT task is a predictor of

16

all these outcomes, and in addition declarative memory is a predictor of both reading and grammatical skills.

The relatively strong relationship between rated attention during the SRT task and procedural learning on this task is a novel finding. Attention ratings were allocated during testing, prior to deriving the procedural learning scores from the tasks, so ratings could not have been biased by any knowledge of the extent of each child's degree of procedural learning. Test retest reliability for the attention ratings across time points was good ($r = .75$). Our results show that procedural learning in the SRT task is not itself a good predictor of variations in reading, grammatical or arithmetic skills. However, the extent to which children can maintain attention during the SRT task is a relatively powerful predictor of these skills. This pattern is clarified by inspecting the pattern of correlations between the latent variables in the model (see Table 3). The most powerful correlate of all three measures of attainment is rated attention during the SRT task. In comparison the procedural learning score on the SRT task is a weaker correlate of all outcomes. Finally, declarative memory is a moderate correlate of grammatical skills and a somewhat weaker correlate of reading skills.

It should be noted that all analyses presented here assess linear relationships between variables. Some might argue that the procedural deficit hypothesis claims specifically that it is *disorders* of reading or language that are related to procedural learning problems. That view implies non-linear effects, such that poor procedural learning is found in only those with the poorest reading and language skills, with variations in reading and language being unrelated to variations in procedural learning in those with average to superior reading and language skills. We assessed this possibility by examining scatter plots. Effects were linear, as is illustrated in Figure 3, which shows the relationship between SRT factor scores and reading factor scores.

17

< Insert Table 3 here >

< Insert Figure 3 here >

Discussion

This study explored the extent to which procedural learning (as assessed by an SRT task), verbal declarative learning and ratings of sustained attention during the SRT task predicted variations in measures of attainment (reading, grammatical and arithmetic skills). Our findings are clear and somewhat surprising. Procedural learning (as assessed by an SRT task) was strongly related ($r = .56$) to ratings of children's attention on the task, which in turn, was predictive of variations in attainment. Conversely, once such attentional effects were controlled, there was no statistically significant relationship between procedural learning and measures of attainment.

The empirical relationships found here between learning on the SRT task, attention, and attainment are clear. A more complex issue is how best to interpret these relationships. Our favoured interpretation is that it is the ability to attend to the SRT task that is an important predictor of attainment and not procedural learning on the task per se. In this view, previous findings from extreme group designs that have found deficits on the SRT task in children with dyslexia (e.g. Jiménez-Fernández, Vaquero, Jiménez & Defior, 2011) or developmental language disorder (Lum, Gelgic, & Conti-Ramsden, 2010) may simply reflect comorbid attentional difficulties in these clinical groups. In other words, attentional difficulties may give rise to impaired procedural learning. This contrasts with the procedural deficit hypothesis, which posits that impairment in a general procedural learning mechanism is a causal risk factor for both developmental dyslexia and developmental language disorder.

The current findings come from a concurrent correlational study and clearly do not necessitate the causal interpretation we favour. It is possible, for example, that poor attention on the

18

SRT task is secondary to problems learning the task. For the moment, we would argue that the strong correlation between attention during the SRT task and attainment (in reading, language and arithmetic) is a novel and surprising result. Further longitudinal and intervention studies are needed to understand the causal mechanisms responsible for this association and the implications of our findings for the procedural deficit hypothesis.

Attention as a predictor of attainment

A few studies have examined the impact of attentional variability when assessing the possible role of procedural learning as a predictor of reading or language disorders in children. Sengottuvel and Rao (2013b; 2014) measured attention in children with developmental language disorder and typically developing children performing an adapted SRT task. However, they found similar levels of attention in both groups, and did not go on to examine whether a relationship between attentional scores, procedural learning and language ability existed. Van Witteloostuijn et al., 2019 included a measure of auditory sustained attention, but found no difference in SRT performance between children with dyslexia and typically developing peers. Controlling for attention did not change their results.

A study by Staels and Van den Broek (2017) used a deterministic SRT task in groups of dyslexic and typically developing children. Although measures of procedural learning on this deterministic task were equivalent across groups, they found that a slight deficit in the initial learning of the sequence in the dyslexic group disappeared after controlling for ADHD ratings. This finding is particularly interesting as it is the first extreme group design to suggest that procedural learning performance in children with dyslexia may be differentially impaired by attentional problems.

To our knowledge, only one study has related attentional functioning to procedural learning and reading ability in a sample unselected for ability. Waber et al. (2003) examined SRT performance in 422 children with a wide range of reading ability. Parental ratings of attention predicted error rates on the SRT task, but not response times.

Our results might be seen as lending support to the findings of Waber et al. (2003) although our within-task measure of attention is quite different to their global attention ratings. Critically, attention ratings had good reliability, and were allocated during testing, prior to any derivation of procedural learning scores from the tasks, so ratings could not have been biased by any knowledge of the extent of each child's degree of procedural learning. Furthermore, in our study attention during the SRT task accounts entirely for the weak relationships between performance on the SRT task and measures of attainment (reading, grammar and arithmetic). The SRT task is the most widely used measure in studies of the procedural deficit hypothesis. Our findings, therefore, raise the possibility that previous findings of procedural learning impairments in children with reading and language disorders may reflect attentional difficulties rather than problems in procedural learning per se.

Our findings in a large sample of children unselected for ability are important for theories dealing with the cognitive causes of language and reading disorders. If performance on the serial reaction time task taps a causal risk factor for language or reading difficulties, we should expect to find a correlation between performance on the task and language or reading skills in unselected groups. By analogy, there is strong evidence that deficits in phoneme awareness are a causal risk factor for the problems in learning to read seen in dyslexia. In a meta-analysis of this relationship, Melby-Lervåg, Lyster and Hulme (2012) reported a group deficit of $d = 1.37$ in extreme group studies comparing children with dyslexia to age-matched controls and a strong correlation between reading and phoneme awareness in studies of unselected samples ($r = .57$). Such converging

20

evidence from extreme group studies and correlational studies is just what we would expect if a hypothesized deficit is an important cause of a disorder, unless the relationship is non-linear and only operates in a sub-group of children with the most severe difficulties. Our results do not give any reason to suggest this might be the case (the relationship between SRT performance and reading is no different at the tail of the distribution of scores than elsewhere; see Figure 3).

The reliability of the SRT task

One aim of the current study was to develop an SRT task with adequate reliability that is suitable for use with children. A major methodological limitation of virtually all previous studies in this area is that the reliability of measures of procedural are not reported. In a previous study (West et al., 2017) we found that the SRT task, as well as a range of other commonly used measures of procedural learning, had very poor reliabilities. The lengthened SRT task that we developed and used here had a moderate split-half reliability (SRT1 $r = .51$; SRT2 $r = .62$) but much lower test-retest reliability ($r = .26$). The reliability of the current SRT task is therefore clearly lower than we would hope for. This might be seen as a limitation of the current study, however, the current study and our previous one (West et al., 2017) are to the best of our knowledge the only studies to report the reliability of measures of procedural learning. The current study also used a latent variable, based on two parallel versions of the SRT task, to correct for measurement error. The correlations between our procedural learning (SRT) latent variable and attainment are not influenced by measurement error and it is notable that they are uniformly low (rs .21 to .31), and lower than the corresponding correlations between the latent variables for attention and attainment (rs .37 to .51). One possible implication of our results is that procedural learning tasks may be inherently unreliable. This reinforces the suggestion made by Ostergaard (1998) that the reliability of procedural learning measures is inherently low due to the myriad influences operating on performance in such tasks (perceptual, motoric and memory processes). His Information

21

Availability model makes clear that when the contribution procedural learning makes to overall performance on a task is small, the reliability of the task for measuring procedural learning will be correspondingly low. In the case of the SRT task, such as the one used here, procedural learning is likely to be only a small influence on task performance, because the task relies to a large extent on a wide range of perceptual, motoric and other cognitive processes. These arguments suggest that developing reliable tests of procedural learning may be extremely difficult.

Limitations

It is important to acknowledge that the current study has a number of limitations that should be addressed in future studies. One limitation is that we used just a single measure of procedural learning: the nonverbal SRT task. In a previous study (West, et al., 2017) we used a wide range of both verbal and nonverbal measures of procedural learning, however, all of these measures showed low reliabilities, and none correlated with literacy, language or arithmetic. In the current study we focused on the SRT task for two reasons: first because it is by far the most widely used measure in studies of the procedural deficit hypothesis, and second because a pilot study suggested that the SRT task could be increased in reliability by making it longer as we did here. Our attempt to increase the reliability of the SRT task was at least partly successful; the current measure defined a latent variable which showed a substantial correlation with our measure of attention on the task. It is notable, however, that this measure still correlated weakly with measures of attainment (and less well than with attention on the task); it is important to note, however, that the low correlations found here between a latent variable for SRT performance and latent variables measuring attainment, cannot be attributed to problems of reliability, since latent variables exclude measurement error. An important aim for future studies should be to include a wider range of procedural memory measures, while at the same time seeking to increase their reliability as we did here for the SRT task. It is possible (but by no means certain) that measures of verbal procedural

memory would correlate more highly with measures of attainment in language and reading, than the nonverbal SRT task used here.

Another limitation of the current study is that it includes just a single measure of attention on the SRT task based on ratings by a single experimenter. It would be useful for future studies to supplement this measure with independent tests of sustained attention (e.g. Attention Sustained subtest from the Leiter International Performance Scale – revised; Roid & Millar, 1997). It should be noted that the test-retest reliability of our measure of sustained attention on the SRT task was good ($r = .75$) and the latent variable for attention in our model, constructed from the shared variance at both time points, was a good predictor of reading, language and arithmetic. Nevertheless, it would be preferable for future studies to get ratings of attention on such tasks from more than a single rater, in order to ascertain their level of inter-rater reliability.

Future investigations would also benefit from attempting to distinguish between attention as a trait versus a state. Sustained attention as a trait is the ability to maintain a vigilant state (Posner & Petersen, 1990) and is assessed with measures such as continuous performance tasks or ADHD rating scales. Attention as a state, on the other hand, encompasses fluctuating states such as the level of task engagement and focus, which have been shown to facilitate sustained attention (Matthews et al., 2010) and task performance. These two aspects of attention are not synonymous. Our study did not assess attention from a trait perspective. We speculate that ratings of attention as a trait and our ratings of attention on the SRT task would correlate well, but future research should explore the role of both trait and state measures of attention and their relationships with procedural learning and measures of language and reading.

Summary and conclusions

We found that declarative learning was a moderate predictor of attainment in reading, language and arithmetic but was not significantly associated with variations in attention on the SRT task. This finding is in line with many other studies that have found a robust relationship between measures of verbal declarative memory and reading ability (Melby-Lervåg, Lyster, & Hulme, 2012). However, the relationship between procedural learning (as assessed by an SRT task) and attainment was entirely explained by rated attention during the SRT task. Previous research has documented the frequent comorbidity between reading, language and attention difficulties (Carroll, Maughan, Goodman, & Meltzer, 2005; Finneran, Francis, & Leonard, 2009; Laasonen et al., 2014). On the basis of the present findings, we suggest that poor performance on procedural learning tasks by children selected for having reading and language difficulties may well reflect poorer attentional functioning during testing, rather than any separable procedural learning impairment. Furthermore, poor attention on procedural learning tasks is likely not to be an isolated effect and attentional difficulties may have downstream effects on attainment. We recommend that future studies of procedural learning and reading and language disorders in children include measures of attention and use procedural learning tasks specifically designed both to maintain children's attention and to have adequate reliability.

References

Anobile, G., Stievano, P., & Burr, D. C. (2013). Visual sustained attention and numerosity sensitivity correlate with math achievement in children. *Journal of Experimental Child Psychology*, *116*(2), 380-391.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412.

Baker, C. I., Olson, C. R., & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science*, *15*(7), 460-466.

Bishop, D. V. (2003). *Test for reception of grammar: TROG-2 version 2*. Pearson Assessment.

Bishop, D. V., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological Bulletin*, *130*(6), 858.

Brigstocke, S., Moll, K., & Hulme, C. (2016). *TOBANS: Test of Basic Arithmetic and Numeracy Skills*. Oxford: Oxford University Press.

Carroll, J. M., Maughan, B., Goodman, R., & Meltzer, H. (2005). Literacy difficulties and psychiatric disorders: Evidence for comorbidity. *Journal of Child Psychology and Psychiatry*, *46*(5), 524-532.

Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are specific language impairment and dyslexia distinct disorders?. *Journal of Speech, Language, and Hearing Research*.

Christiansen, M. H., Dale, R. A., Ellefson, M. R., & Conway, C. M. (2002). The role of sequential learning in language evolution: Computational and experimental studies. In *Simulating the evolution of language* (pp. 165-187). London: Springer.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, *120*(3), 235.

Cohen, M. (1997). Children's memory scale (CMS). *San Antonio: The Psychological Corporation*.

Desmottes, L., Meulemans, T., & Maillart, C. (2016a). Later learning stages in procedural memory are impaired in children with specific language impairment. *Research in developmental disabilities*, *48*, 53-68.

Desmottes, L., Meulemans, T., & Maillart, C. (2016b). Implicit spoken words and motor sequences learning are impaired in children with specific language impairment. *Journal of the International Neuropsychological Society*, *22*(05), 520-529.

Desmottes, L., Maillart, C., & Meulemans, T. (2017). Memory consolidation in children with specific language impairment: Delayed gains and susceptibility to interference in implicit sequence learning. *Journal of clinical and experimental neuropsychology*, *39*(3), 265-285.

Ebert, K. D., & Kohnert, K. (2011). Sustained attention in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*.

Evans, T. M., & Ullman, M. T. (2016). An extension of the procedural deficit hypothesis from developmental language disorders to mathematical disability. *Frontiers in Psychology*, *7*, 1318.

Finneran, D. A., Francis, A. L., & Leonard, L. B. (2009). Sustained attention in children with specific language impairment (SLI). *Journal of Speech, Language, and Hearing Research*, *52*, 915-929.

Fletcher, J. M. (2009). Dyslexia: The evolution of a scientific concept. *Journal of the International Neuropsychological Society*, *15*(4), 501-508.

Gabriel, A., Maillart, C., Guillaume, M., Stefaniak, N., & Meulemans, T. (2011). Exploration of serial structure procedural learning in children with language impairment. *Journal of the International Neuropsychological Society*, *17*, 336-343.

Hayduk, L. A., & Littvay, L. (2012). Should researchers use single indicators, best indicators, or multiple indicators in structural equation models? *BMC medical research methodology*, *12*, 1.

Hedenius, M., Persson, J., Tremblay, A., Adi-Japha, E., Veríssimo, J., Dye, C. D., ... & Ullman, M. T. (2011). Grammar predicts procedural learning and consolidation deficits in children with specific language impairment. *Research in developmental disabilities*, *32*, 2362-2375.

Hedenius, M., Persson, J., Alm, P. A., Ullman, M. T., Howard, J. H., Howard, D. V., & Jennische, M. (2013). Impaired implicit sequence learning in children with developmental dyslexia. *Research in developmental disabilities*, *34*, 3924-3935.

Henderson, L. M., & Warmington, M. (2017). A sequence learning impairment in dyslexia? It depends on the task. *Research in Developmental Disabilities*, *60*, 198-210.

Howard, D. V., Howard Jr, J. H., Japikse, K., DiYanni, C., Thompson, A., & Somberg, R. (2004). Implicit sequence learning: effects of level of structure, adult age, and extended practice. *Psychology and aging*, *19*(1), 79.

Howard, J. H., Howard, D. V., Japikse, K. C., & Eden, G. F. (2006). Dyslexics are impaired on

    implicit higher-order sequence learning, but not on implicit spatial context learning.

    Neuropsychologia, *44*, 1131-1144.

Hsu, H. J., & Bishop, D. V. (2014). Sequence-specific procedural learning deficits in children with

    specific language impairment. *Developmental Science*, *17*, 352-365.

Jiménez-Fernández, G., Vaquero, J. M., Jiménez, L., & Defior, S. (2011). Dyslexic children show

    deficits in implicit sequence learning, but not in explicit sequence learning or contextual

    cueing. *Annals of Dyslexia*, *61*, 85-110.

Kenny, D. A. (2015). Measuring model fit. Retrieved from: http://davidakenny.net/cm/fit.htm.

Kidd, E., & Kirjavainen, M. (2011). Investigating the contribution of procedural and declarative

    memory to the acquisition of past tense morphology: Evidence from Finnish. *Language and*

    Cognitive Processes, *26*(4-6), 794-829.

Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S. A. H., Gustafsson, J. E., & Hulme, C. (2015).

    Sentence repetition is a measure of children's language skills rather than working memory

    limitations. *Developmental Science*, *18*(1), 146-154.

Korpipää, H., Koponen, T., Aro, M., Tolvanen, A., Aunola, K., Poikkeus, A. M., ... & Nurmi, J. E.

    (2017). Covariation between reading and arithmetic skills from Grade 1 to Grade

    7. *Contemporary educational psychology*, *51*, 131-140.

Krishnan, S., Watkins, K. E., & Bishop, D. V. (2016). Neurobiological basis of language learning

    difficulties. *Trends in Cognitive Sciences*, *20*(9), 701-714.

Laasonen, M., Väre, J., Oksanen-Hennah, H., Leppämäki, S., Tani, P., Harno, H., ... & Cleeremans, A.

    (2014). Project DyAdd: Implicit learning in adult dyslexia and ADHD. *Annals of Dyslexia*,

    64, 1-33.

Landerl, K., & Moll, K. (2010). Comorbidity of learning disorders: Prevalence and familial

    transmission. *Journal of Child Psychology and Psychiatry, 51*, 287–294.

Lord, F. M. (1958). The utilization of unreliable difference scores. *Journal of Educational Psychology*,

    49, 150.

Lum, J. A., & Bleses, D. (2012). Declarative and procedural memory in Danish speaking children with

    specific language impairment. *Journal of Communication Disorders*, *45*, 46-58.

Lum, J. A., Gelgic, C., & Conti-Ramsden, G. (2010). Procedural and declarative memory in children

    with and without specific language impairment. *International Journal of Language &*

    Communication Disorders, *45*, 96-107.

Lum, J. A., Ullman, M. T., & Conti-Ramsden, G. (2013). Procedural learning is impaired in dyslexia:

    Evidence from a meta-analysis of serial reaction time studies. *Research in Developmental*

    Disabilities, *34*, 3460-3476.

Matthews, G., Warm, J. S., Reinerman-Jones, L. E., Langheim, L. K., Washburn, D. A., & Tripp, L.

    (2010). Task engagement, cerebral blood flow velocity, and diagnostic monitoring for

    sustained attention. *Journal of Experimental Psychology: Applied*, *16*, 187.

McCoach, D. B., Black, A. C., & O'Connell, A. A. (2007). Errors of inference in structural equation

    modeling. *Psychology in the Schools*, *44*, 461-470.

Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning

    to read: a meta-analytic review. *Psychological Bulletin*, *138*, 322-348.

Muthén, B. O., & Muthén, L. (1998-2016). Mplus user's guide (7th ed). *Los Angeles, CA: Muthén &*

    Muthén.

Nicolson, R. I., & Fawcett, A. J. (2007). Procedural learning difficulties: reuniting the developmental disorders?. *TRENDS in Neurosciences*, *30*, 135-141.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1-32.

Nunnally, J., & Bernstein, I. (1994). *Psychometric theory* (3rd edition). MacGraw-Hill, New York.

Ostergaard, A. L. (1998). The effects on priming of word frequency, number of repetitions, and delay depend on the magnitude of priming. *Memory & Cognition*, *26*, 40-60.

Overall, J. E., & Woodward, J. A. (1975). Unreliability of difference scores: A paradox for measurement of change. *Psychological Bulletin*, *82*, 85.

Perlant, A. S., & Largy, P. (2011). Are implicit learning abilities sensitive to the type of material to be processed? Study on typical readers and children with dyslexia. *Journal of Research in Reading*, *34*, 298-314.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual review of neuroscience*, *13*, 25-42.

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: a critical reexamination and new recommendations. *Psychological Methods*, *10*, 178.

Putzke, J. D., Williams, M. A., Adams, W., & Boll, T. J. (1998). Does memory test performance in children become more consistent with age? Cross-sectional comparisons using the WRAML. *Journal of Clinical and Experimental Neuropsychology*, *20*, 835-845.

Rabiner, D., Coie, J. D., & Conduct Problems Prevention Research Group. (2000). Early attention

  problems and children's reading achievement: A longitudinal investigation. *Journal of the*

  American Academy of Child & Adolescent Psychiatry, *39*, 859-867.

Roid, G. H., & Miller, L. J. (1997). Leiter international performance scale-revised (Leiter-R). *Wood*

  Dale, IL: Stoelting.

Schmalz, X., Altoè, G., & Mulatti, C. (2017). Statistical learning and dyslexia: a systematic review.

  Annals of Dyslexia, 67, 1-16.

Schmalz, X., Moll, K., Mulatti, C., & Schulte-Körne, G. (2019). Is statistical learning ability related to

  reading ability, and if so, why?. *Scientific Studies of Reading*, *23*(1), 64-76.

Schvaneveldt, R. W., & Gomez, R. L. (1998). Attention and probabilistic sequence learning.

  Psychological Research, *61*, 175-190.

Shanks, D. R., Green, R. E., & Kolodny, J. A. (1994). A critical examination of the evidence for

  unconscious (implicit) learning. In C. Umilta & M. Moscovitch (Eds.), Attention and

  performance XV (pp. 837-860). Cambridge, MA: MIT Press.

Shanks, D. R., & Johnstone, T. (1999). Evaluating the relationship between explicit and implicit

  knowledge in a sequential reaction time task. *Journal of Experimental Psychology: Learning,*

  Memory, and Cognition, *25*, 1435.

Shanks, D. R., Wilkinson, L., & Channon, S. (2003). Relationship between priming and recognition in

  deterministic and probabilistic sequence learning. *Journal of Experimental Psychology:*

  Learning, Memory, and Cognition, *29*(2), 248.

Selst, M. V., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology*, *47*, 631-650.

Sengottuvel, K., & Rao, P. K. (2013b). Aspects of grammar sensitive to procedural memory deficits in children with specific language impairment. *Research in Developmental Disabilities*, *34*, 3317-3331.

Sengottuvel, K., & Rao, P. K. (2014). Sequence learning pattern in children with specific language impairment. *International Journal on Disability and Human Development*, *13*, 55-62.

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of Learning and Memory*, *82*, 171-177.

Staels, E., & Van den Broeck, W. (2017). A specific implicit sequence learning deficit as an underlying cause of dyslexia? Investigating the role of attention in implicit learning tasks. *Neuropsychology*, *31*, 371.

Stoodley, C. J., Ray, N. J., Jack, A., & Stein, J. F. (2008). Implicit learning in control, dyslexic, and garden-variety poor readers. *Annals of the New York Academy of Sciences*, *1145*, 173-183.

Tomblin, J. B., & Zhang, X. (2006). The dimensionality of language ability in school-age children. *Journal of Speech, Language, and Hearing Research*.

Torgesen, J. K., Wagner, R., & Rashotte, C. (1999). TOWRE–2 Test of Word Reading Efficiency. *Austin, TX: Pro-Ed*.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), B25-B34.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*, 231-270.

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, *41*, 399-433.

Van Daal, V., van der Leij, A., & Adèr, H. (2013). Specificity and overlap in skills underpinning reading and arithmetical fluency. *Reading and Writing*, *26*(6), 1009-1030.

van Witteloostuijn, M., Boersma, P., Wijnen, F., & Rispens, J. (2019). Statistical learning abilities of children with dyslexia across three experimental paradigms. *PloS One*, *14*(8).

von Koss Torkildsen, J., Arciuli, J., & Wie, O. B. (2019). Individual differences in statistical learning predict children's reading ability in a semi-transparent orthography. *Learning and Individual Differences*, *69*, 60-68.

Vukovic, R. K., & Lesaux, N. K. (2013). The relationship between linguistic skills and arithmetic knowledge. *Learning and Individual Differences*, *23*, 87-91.

Waber, D. P., Marcus, D. J., Forbes, P. W., Bellinger, D. C., Weiler, M. D., Sorensen, L. G., & Curran, T. (2003). Motor sequence learning and reading ability: Is poor reading associated with sequencing deficits?. *Journal of Experimental Child Psychology*, *84*, 338-354.

West, G*., Vadillo, M. A., Shanks, D. R., Hulme, C.* (2018). The procedural learning deficit hypothesis of language learning disorders: we see some problems*. Developmental Science,. 2017;00:e12552.* <https://doi.org/10.1111/desc.12552>

Table 1

Performance on attainment and memory measures

	N	Mean	*SD*	Range	Reliability
Age in months	112	92.11	5.71	83 - 105	-
TROG-2 (Total correct)	112	71.58	6.89	45 - 80	.88 [sm]
TOWRE-2 Words	111	61.12	11.44	11 - 79	.90 [rm]
TOWRE-2 Nonwords	110	36.75	10.71	0 - 57	.90 [rm]
Arithmetic composite	111	53.72	25.62	1 - 147	.97 [r]
Addition	111	17.62	8.14	1 - 45	.92 [r]
Addition plus carry	112	8.62	5.31	0 – 30	.89 [r]
Subtraction	112	12.14	5.76	0 - 36	.88 [r]
Subtraction plus carry	112	6.02	3.88	0 - 22	.85 [r]
Multiplication	112	9.43	5.82	0 - 22	.93 [r]
Word Lists (WL)					.84 [sm]
Learning	112	32.44	6.03	14 - 45	
Delay	111	6.09	1.94	1 - 10	
Consolidation	111	4.29	1.78	0 - 9	
SRT1 RT Difference Score	110	.13	.07	-.04 - .38	.51 [s] /.26 [r]
SRT2 RT Difference Score	111	.14	.08	-.06 - .36	.62 [s] /.26 [r]
SRT1 Attention Rating	112	7.25	1.29	2.8 - 9	.75 [r]
SRT2 Attention Rating	111	7.23	1.45	2.6 - 9	.75 [r]

[s] = split-half reliability; [r] = test-retest reliability calculated from task; [sm / rm] = split half or test-retest reliability estimate respectively from standardisation manual

Table 2

Correlation matrix for all measures

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Age (months)																
2. TROG-2	.31**															
3. Towre Words	.18	.45*														
4. Towre Nonwords	.19*	.35*	.83**													
5. Arithmetic composite	.44**	.42**	.50**	.55**												
6. Addition	.42**	.40**	.50**	.52**	.92**											
7. Addition with carry	.33**	.32**	.38**	.43**	.89**	.78**										
8. Subtraction	.36**	.38**	.42**	.48**	.93**	.78**	.81**									
9. Subtraction with carry	.25**	.36**	.50**	.52**	.85**	.69**	.73**	.78**								
10. Multiplication	.50**	.37**	.40**	.46**	.83**	.68**	.62**	.72**	.63**							
11. WL Learning	.07	.47**	.37**	.32**	.26**	.19*	.19*	.23*	.29**	.29**						
12. WL Delay	-.07	.34**	.24*	.19	.15	.12	.06	.13	.17	.22*	.76**					
13. WL Consolidation	.03	.30**	.24**	.21*	.19*	.14	.10	.17	.22*	.25**	.63**	.68**				
14. SRT1	.09	.05	.07	.05	-.01	.04	-.03	-.05	.01	-.05	-.05	.02	-.03			
15. SRT2	.15	.32**	.31**	.22*	.25**	.21	.23*	.27**	.28**	.15	.20*	.13	.10	.26**		
16. Attention SRT1	.17	.39**	.37**	.33**	.29**	.26**	.31**	.23*	.32**	.19*	.18	.10	.04	.18	.19	
17. Attention SRT2	.17	.37**	.44**	.37**	.29**	.27**	.33**	.28**	.31**	.10	.16	.05	.12	.26**	.36**	.75**

*p < .05; **p < .01

Table 3

Correlation matrix for the latent variables

|	**1**	**2**	**3**	**4**	**5**
1. Procedural Learning (SRT)					
2. Declarative Learning (CMS - WL)	.22				
3. Attention	.56	.19			
4. Reading (TOWRE)	.31	.29	.51		
5. Grammar (TROG-2)	.30	.44	.46	.45	
6. Arithmetic (TOBANS)	.21	.07	.37	.54	.39
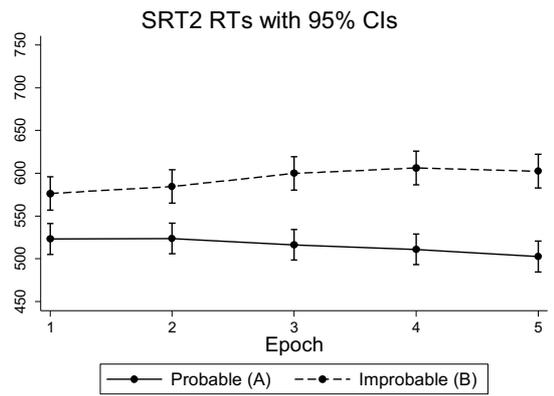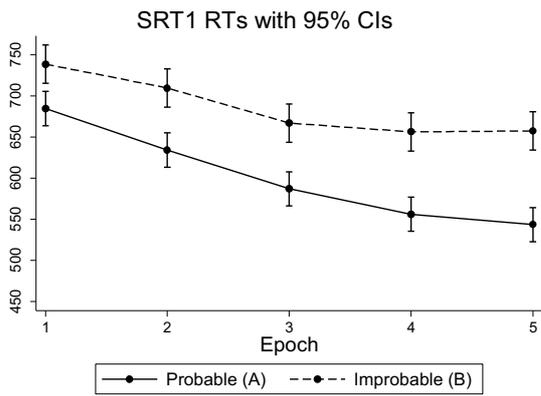
Figure 1. Response times in milliseconds per sequence and epoch for both SRT tasks, with 95% CIs
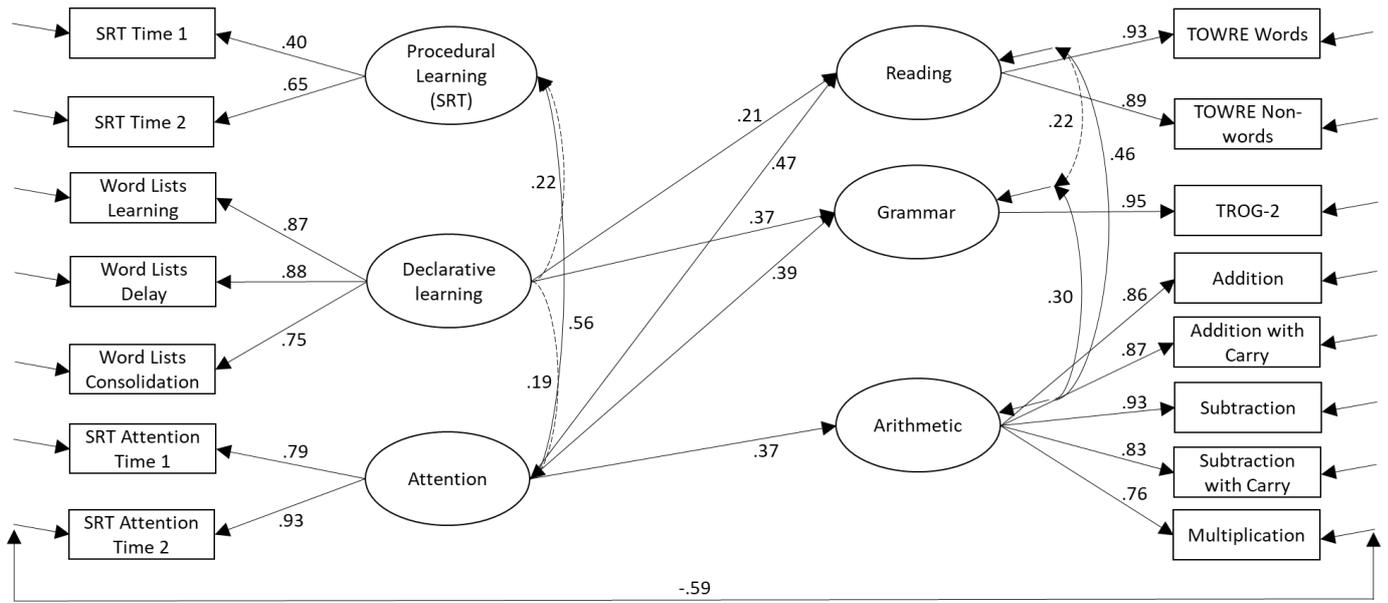
$(\chi^2 (79) = 94.526, p = . 112; RMSEA = . 042[90\% CI . 0.000\ 0.071]; CFI = .98; TLI = .98)$

Figure 2. Path model of relationships between constructs. Single headed arrows represent regression paths or factor loadings. Twin headed arrows represent correlations between variables. Standardized regression coefficients and factor loadings are shown.
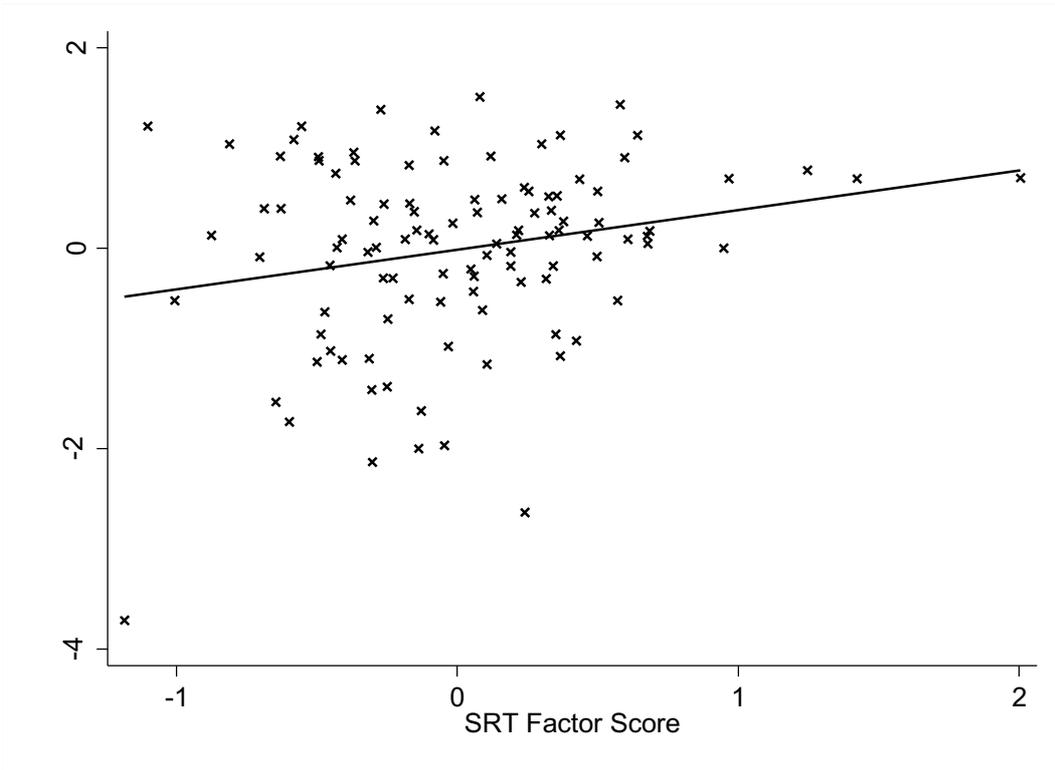
Figure 3. Scatter plot showing the relationship between SRT and reading factor scores