MIND, BRAIN, AND EDUCATION

Domain-General Learning and Memory Substrates of Reading Acquisition

F. Sayako Earle[1], Stephanie N. Del Tufo[2], Tanya M. Evans[3], Jarrad A. G. Lum[4], Laurie E. Cutting[2], and Michael T. Ullman[5]

ABSTRACT— Prior research has demonstrated that linguistic skills and knowledge contribute to successful reading acquisition. In contrast, little is known about the influence of domain-general learning abilities on reading. To investigate associations between general memory functions and reading during the early stages of learning to read, performance measures of word-level reading and of declarative and procedural learning were obtained in a cohort of 140 children, annually during their first 4 years of school. We hypothesized that differences in learning task performance would relate to reading ability in the early years, when children are first learning to read. We employed a series of linear mixed effects models to test the relationships between learning abilities and reading across time. Declarative learning performance predicted reading ability in first grade, while procedural learning performance predicted reading ability in second grade. Our findings suggest that reading acquisition may depend in part on general capacities for learning.

Converging evidence over decades of research on reading acquisition has highlighted a variety of skills in young children that are highly predictive of later reading ability. These include pre-literacy competencies such as letter recognition and production, letter-name knowledge, and letter-sound correspondences (see Ehri, 2005, for review). Previous research suggests that these competencies are highly intertwined with awareness of the sound constituents of words (Byrne & Fielding-Barnsley, 1989). This awareness is often indexed by phonological processing ability, which in itself is also a significant early indicator of reading trajectory (Liberman, Shankweiler, Fischer, & Carter, 1974; see Goswami & Bryant, 1990, for review). All of these skills reflect capacities within the linguistic domain, and thus are subject to influence by the acquisition of reading itself (Hogan, Catts, & Little, 2005; Perfetti, Beck, Bell, & Hughes, 1987). Comparatively, we know very little about how non-linguistic domain-general mechanisms contribute to how we learn to read.

Declarative and procedural memory are critical domain-general learning systems that have been associated with language skills (Hamrick, Lum, & Ullman, 2018), and are implicated in the learning of language (Ullman, 2004). The declarative memory system is associated with the hippocampus and its surrounding structures, while the procedural memory system relies on the striatum (Poldrack & Packard, 2003). These neuroanatomical differences underlie the functional distinctions between the two memory systems, broadly summarized as the rapid learning of arbitrary items and their associations (declarative memory), and the incremental automatization of skills and habits (procedural memory). In language, the learning of vocabulary and grammatical rules has been proposed to be differentially supported according to the functional distinctions between declarative and procedural memory (Ullman, 2004).

Importantly, these differential roles also appear to be well-suited for acquiring the different knowledge and skills involved in learning how to read. While there is no shortage of unique proposals regarding the developmental trajectory of literacy and its supporting skills (see Berninger, 1995, for

---

[1]Communication Sciences and Disorders, University of Delaware,
[2]Peabody College of Education and Human Development, Vanderbilt University
[3]Curry School of Education, University of Virginia
[4]School of Psychology, Deakin University
[5]Department of Neuroscience, Georgetown University

Address correspondence to F. Sayako Earle, Communication Sciences and Disorders, University of Delaware, 100 Discovery Boulevard, Newark, DE, 19713; e-mail: fsearle@udel.edu and Laurie E. Cutting, Peabody College of Education and Human Development, Vanderbilt University, 110 Magnolia Circle Nashville, TN 37203; e-mail: laurie.cutting@Vanderbilt.edu

Data presented in this article will be made available upon request.

review), there exists a general consensus that learning to read occurs in overlapping and continuous stages. The first step in learning to read involves the establishment of associations between letters, sounds, pronunciations, and word meanings (Marsh, Friedman, Welch, & Desberg, 1981), which together can be referred to as orthographic mapping (Ehri, 2005). These mappings must also become increasingly automatized through repetition, facilitated by frequently occurring words and letter strings (Reitsma, 1983). This increased efficiency seems to allow the reallocation of cognitive resources, such as attention and working memory, to the higher-order task of reading comprehension (see Logan, 1997, for review).

These stages of reading—orthographic mapping and skill automatization—are consistent with the respective functional roles of the declarative and procedural learning systems. Indeed, in adults learning to read an artificial script, declarative memory has been implicated in forming arbitrary letter-sound associations, and procedural memory in automatization during practice (Bitan & Karni, 2004). Thus, it stands to reason that these systems play similar roles in children when they are first acquiring the ability to read. Moreover, we might predict that such associations will differ across time. For example, dynamical perspectives consider the reading performance to reflect interactions between the state of knowledge by a particular individual and the contextual challenge at hand (e.g., Wijnants, Hasselman, Cox, Bosman, & Van Orden, 2012), eventually leading to a stable system. Thus, from this perspective performance on the same reading task may be differently associated with different learning capacities at the various states of knowledge throughout reading development.

A similar link between the acquisition of reading and domain-general learning systems has been proposed previously, albeit indirectly, in the developmental dyslexia literature. Specifically, one hypothesis regarding the etiology of dyslexia posits that a deficit in implicit learning underlies difficulties in the automatization of decoding skills (Nicolson & Fawcett, 2011). Tests of an implicit learning deficit in dyslexia have yielded mixed results (e.g., Howard Jr, Howard, Japikse, & Eden, 2006; Lum, Ullman, & Conti-Ramsden, 2013; Roodenrys & Dunn, 2008). This inconsistency in findings is unsurprising, given the differences across studies regarding the wide range of task measures of implicit learning, the various neuroanatomical substrates involved in those tasks, and the developmental timing of when implicit learning measures were obtained relative to reading ability.

Interestingly, despite this question regarding a potential implicit learning etiology in developmental dyslexia, there is little empirical evidence suggesting that one or more types of implicit learning support learning how to read. Thus, even in cases wherein individuals with dyslexia are observed with implicit learning difficulties, it is unclear if the problems with reading can be attributed to that impairment. Therefore, an important goal would be to reveal the relationship between memory capacities and reading ability in general. In order to avoid the potential for various interpretations surrounding the functional distinction in memory as being either "implicit" or "explicit," here we focus on declarative and procedural memory. Procedural memory that relies on the striatum appears to be compromised in dyslexia (Lum et al., 2013), and therefore this type of implicit learning may be relevant in learning how to read.

In this study, longitudinal measures of reading performance, and task-based measures of learning in the declarative and procedural memory systems, respectively, were obtained from school-age children over their first 4 years of school. Given the primary emphasis on word reading, or "learning to read" (Chall, 1983) in early grades, we focused on word-level reading skills. We assessed declarative memory via a recognition memory task (Hedenius, Ullman, Alm, Jennische, & Persson, 2013), and procedural memory via a serial reaction time (SRT) task (Nissen & Bullemer, 1987). These two task paradigms were carefully selected on the basis of independent evidence linking each task to neural substrates consistent with recruitment of declarative and procedural memory, respectively (see Methods). Therefore, while we acknowledge that no single task can capture the full range of functions of a particular learning and memory system, performance on these tasks likely reflects at least certain key functions of these systems. We hypothesized that declarative and procedural memory may be more relevant for reading during early stages of literacy, when orthographic mapping and skill automatization appear to be taking place.

METHODS

Participants

Participants were recruited from public and private schools from the greater Nashville area. Trained researchers screened 866 children in these schools in order to capture an ecologically representative sample of readers. Families were then called and screened for inclusion criteria prior to being enrolled in the longitudinal portion of the study. All participants were native speakers of American English with normal or corrected-to-normal vision. In addition, all participants reported no history of major psychiatric illness, traumatic brain injury, or epilepsy. Informed child assents and parental consents were obtained according to Vanderbilt University's institutional review board (IRB) guidelines. Participants received compensation for their participation in the study, as per the study's approved IRB protocol.

This study was part of a larger longitudinal study of reading (grades 1–4) at the Education and Brain Sciences Research Lab at Vanderbilt University. At each visit children were given various tasks, including standardized assessments of reading, and the declarative memory (DM) and procedural memory (PM) experimental tasks reported here (recognition memory task, as used by Hedenius et al., 2013; and the SRT task, Nissen & Bullemer, 1987). A total of 140 children were enrolled at Visit 1. See Appendix A for demographic data for participants at Visit 1. The present analyses include data from all participants who were assessed on PM, DM, or both, at any of the four visits. The dataset includes 129 unique participants, consisting of 98 children at Visit 1 (mean age 7.49[*SD* .32]), 90 children at Visit 2 (age 8.46 [.35]), 89 children at Visit 3 (age 9.49 [.33]), and 79 children at Visit 4 (age 10.43 [.33]).See Table A1 for demographic information.

Procedures

Each visit took place following the completion of the school year, such that Visit 1 took place following first grade, Visit 2 following second grade, and so on. All test administration and experimental procedures were conducted in a quiet room one-on-one with a trained experimenter. Nonstandardized (experimental) testing took place on a PC computer with an LCD screen. The timing of stimulus presentation and response recording was controlled via E-Prime (Version 2.0, Psychology Software Tools, Pittsburgh, PA). Participants indicated their responses through a serial response (SR) box for the DM task, and button presses on a number pad for the PM task.

Reading Assessments

At each visit, participants completed the Letter-Word Identification and Word Attack subtests of the Woodcock Johnson Tests of Achievement (WJ-III; Woodcock, McGrew, & Mather, 2001), and the Sight Word Efficiency and Phonemic Decoding Efficiency subtests of the Test of Word Reading Efficiency (TOWRE; Torgesen, Rashotte, & Wagner, 1999). Performance on these assessments served as our measures of word and nonword reading abilities, under timed and untimed conditions. Assessments were administered by trained graduated students who passed a fidelity check at 90% reliability.

Declarative Memory: Recognition Memory Task

DM was assessed using a nonverbal recognition memory task with incidental encoding (e.g., Hedenius et al., 2013). Recognition memory tasks are a widely used measure of declarative memory. Performance on these tasks has been shown to depend on the hippocampus and other medial temporal lobe structures (Squire, Wixted, & Clark, 2007). Furthermore, recognition memory of objects with incidental encoding minimizes the influence of working memory and language involvement in task performance that is often present in other declarative memory tasks, such as word list learning.

Stimulus materials were 128 black-and-white drawings of 64 real and 64 made-up objects. Images of real objects were taken from free websites, purchased collections, and selections from the standardized set of 260 images published by Snodgrass and Vanderwart (1980). The real items used in the stimulus set at each phase (encoding and recognition) were matched for word frequency, number of syllables, and number of phonemes. Images of made-up objects were drawn from stimuli used in previous studies (Cornelissen et al., 2004; Eals & Silverman, 1994; Williams & Tarr, 1997), resized and modified as appropriate.

Participants completed the task in two phases, encoding and recognition, at each of the four visits. During the (incidental) encoding phase, participants were instructed to place an index finger from each hand on the left and right keys (positions 1 and 5) on the SR box. Participants were told that they would see a series of images, some of which were real and some of which were not real. They were asked to indicate if the object presented on the screen was real or made up by pressing "left" or "right" on the SR box, as quickly and as accurately as possible.

Following three practice trials, participants completed 64 trials of the encoding task (32 real/32 made-up), administered in a pseudorandom order such that the same trial type (real or made-up) never exceeded three consecutive trials. At the start of each trial, a fixation cross appeared in the center of the screen for 1,000 ms, followed by the target stimulus for 500 ms. The prompts ("real"/"made-up") remained at the bottom corners of the screen as a reminder for which keys corresponded to which response. Following the stimulus presentation, the fixation cross returned until the participant indicated their response. If a participant responded within the 500 ms of stimulus presentation, the trial still ended at the end of the 500 ms presentation period. This was done in order to control the duration of exposure to visual stimuli during the incidental encoding of items. Otherwise, the trial ended when the participant indicated a response, with a maximum of 4,500 ms allowed per trial (if a response was not detected at the end of 4,500 ms, a buzzer sounded for 200 ms, followed by 800 ms of a fixation cross preceding the next trial item).

Participants were given approximately 10 min after the encoding phase before being administered the recognition phase of the experiment. During the recognition phase, participants were told that, again, they would see a series of images, and that some of these images were seen before in the previous task (old), and some were not (new). Participants were asked to indicate whether or not the presented object was seen in the previous task or not, using the left and right keys on the SR box. Following 6 practice trials, participants completed the 128 trials of this task (64 old/64 new). The trial structure was identical to the encoding phase, with the visual prompt changed to the question "seen before?" at the bottom center of the screen, with "yes" and "no" at the corners of the screen as a reminder for which keys corresponded with which response. Proportions of correct responses in the recognition memory task were transformed, for each subject, to *d'* scores (MacMillan & Creelman, 2004), defined as the *Z*-score of the hit rate minus the *Z*-score of the false alarm rate.

Procedural Memory: SRT Task

PM was assessed using a version of Nissen and Bullemer's (1987) SRT task. This task is widely used as a measure of procedural memory, and moreover, performance on this task has been shown to rely on the basal ganglia, in particular the striatum (Clark, Lum, & Ullman, 2014).

Participants were seated in front of the computer and told that a smiley face would be presented in one of the four boxes on the screen, arranged in a diamond pattern (Lum, Gelgic, & Conti-Ramsden, 2010). Using their dominant hand, participants were asked to press, as quickly and as accurately as possible, marked buttons on the number pad (2, 4, 6, 8) corresponding to the location of the smiley face on the screen. After 10 practice trials, participants completed 6 blocks, each of 60 trials. In the first and last blocks, the image was presented in a pseudorandom sequence of locations. In blocks 2–5, the image appeared in a sequence of 10 locations (4–6–2–8–6–4–2–8–4–2), which was repeated 6 times within each of these 4 blocks. Average reaction times (RT) from the SRT task were calculated by block only for those trials on which participants responded accurately. The difference in average RT between the last random (Block 6) and the prior sequence block (Block 5) was employed as our measure of PM, as has been done in prior investigations of PM in children (e.g., Lum et al., 2010).

In order to ensure that the data were treated on commensurate scales across measures in our statistical models, raw reading scores, PM scores from the SRT task, and *d'* scores from the recognition memory task, were scaled using the Proximity to Maximum Scaling (POMS) method (see Moeller, 2015, regarding its application to longitudinal data). This method of transformation was chosen in particular in order to conserve the relationship between individuals within each visit as well as the relationship within individuals across visits.

See Table B1 for means and standard deviations of performance in each task by Visit, and Table C1 for reliability estimates.

RESULTS

We were primarily interested in potential relationships between reading and learning abilities over the first 4 years of formal reading instruction. In order to test these relationships, we fitted a series of linear mixed-effects models that modeled separate intercepts for participant-by-visit variability (VanLeeuwen, Murray, & Urquhart, 1996). For our outcome measure, we used the average of the scaled timed/untimed and word/nonword reading metrics as a composite reading score, since changes in reading scores across visits were comparable across the four measures (Appendix B). The full model specified a three-way interaction term between DM, PM, and Visit, the subsuming two-way interactions, individual predictors, and the by-participant-by-visit intercept as random effects (see recommendations for longitudinal models by Laird & Ware, 1982). We employed a backward-fitting procedure to determine the model of best fit, using a likelihood ratio test (lmtest package; Zeileis & Hothorn, 2002) in R (R Development Core Team; Version 3.3.1, 2016). Models were fitted using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in R. Marginal and Conditional R^2 values for overall models (as described by Nakagawa & Schielzeth, 2012) were calculated using the MuMIn package (Version 1.15.16, Barton, 2016), and proportional effect sizes for individual predictors in the mixed effects models (partial r^2_b values) as defined by Edwards, Muller, Wolfinger, Qaqish, and Schabenberger (2008) were computed using the r2glmm package (Version 0.1.0, Jaeger, 2016) in R.

The accepted model ($AIC = -518.7$, $BIC = -483.8$, Log Likelihood = 269.3, $r^2_m = .191$, $r^2_c = .965$) included a main effect of Visit as well as a main effect of DM, which were qualified by a significant interaction between Visit and DM. There was also an interaction between Visit and PM, but no significant main effect of PM. The direction of the main effect of DM suggests that DM was positively predictive of reading overall. The negative interactions suggest that the effect of both DM and PM on reading ability declined over the four visits. See Table 1 for parameter estimates and significance values. Correlations of task performances by visit are provided in Appendix D (Tables D1-D4).

In order to parse the interactions between Visit and DM, and between Visit and PM, four separate regression models were run, one for each visit, with DM and PM as predictors. At Visit 1 ($F_{2,60} = 6.65$, $p = .002$, $r^2 = .154$), we found that while DM was significantly associated with reading ability ($\beta = .62$, $SE = .19$, $p = .002$, $r^2_\beta = .154$), while PM was not ($\beta = -.10$, $SE = .12$, $p = .409$, $r^2_\beta = .111$). Conversely, at Visit 2 ($F_{2,42} = 3.42$, $p = .042$, $r^2 = .100$), we found that while PM was significantly associated with reading ability ($\beta = .35$, $\mathrm{SE} = .17$, $p = .049$, $r^2_\beta = .090$), DM was not ($\beta = .34$, $\mathrm{SE} = .25$, $p = .178$, $r^2_\beta = .043$). The regression model was not significant at either Visit 3 ($F_{2,70} = 1.64$, $p = .201$, $r^2 = .018$) or Visit 4 ($F_{2,60} = 2.86$, $p = .065$, $r^2 = .058$), reflecting no association between PM and DM with reading ability at either visit (Visit 3: PM, $\beta = .12$, $SE = .17$, $p = .478$, $r^2_\beta = .077$; DM, $\beta = .36$, $SE = .23$, $p = .126$, $r^2_\beta = .033$ and Visit

Table 1
Linear Mixed Effects Model

	Full model				*Accepted model*			
Fixed effects	*Estimate*	SE	p	r^2_b	*Estimate*	SE	p	r^2_b
(Intercept)	0.35	0.25	.151		0.21	0.06	<.001***	
Visit	0.05	0.11	.641	0.001	0.13	0.02	<.001***	0.012
DM	0.02	0.39	.960	< 0.001	0.24	0.08	.003**	0.003
PM	−0.10	0.39	.792	< 0.001	0.10	0.06	.067	0.002
Visit * DM	0.06	0.17	.743	< 0.001	−0.07	0.03	.023*	0.002
Visit * PM	0.07	0.17	.666	< 0.001	−0.04	0.20	.032*	0.002
DM * PM	0.31	0.64	.629	< 0.001	n/a			
Visit * DM * PM	−0.17	0.27	.522	< 0.001	n/a			
Random	Variance	*SD*			Variance	*SD*		
Subject (intercept)	0.02	0.15			0.02	0.15		
Visit	0.00	0.00			0.00	0.00		
Residual	0.00	0.03			0.00	0.03		

Parameter estimates, significance values, and proportional effect sizes (r^2_b) for the full (AIC = −519.5, BIC = −477.5, Log Likelihood = 271.7, r^2_m = .194, r^2_c = .963) and accepted (AIC = −518.7, BIC = −483.8, Log Likelihood = 269.3, r^2_m = .191, r^2_c = .965) models.
*Statistical significance at .05 level.
**Statistical significance at .01 level.
***Statistical significance at .001 level.

4: PM, $\beta = .26$, $SE = .16$, $p = .117$, $r^2_\beta = .047$; DM, $\beta = .49$, $SE = .29$, $p = .097$, $r^2_\beta = .042$). Therefore, the effect of DM on reading ability appears to be driven primarily by its association with reading at Visit 1, whereas PM has an effect on reading ability at Visit 2. See Figure 1 for a graphical depiction of these relationships.

DISCUSSION

The current paper presents preliminary evidence that early stages of reading are associated with domain-general learning abilities. Specifically, performance reflecting learning in declarative memory (DM) was associated with word-level reading ability at the end of first grade. In contrast, performance reflecting learning in procedural memory (PM) was associated with word-level reading ability at the end of second grade.

Our findings resonate with well-known theories of reading development (e.g., Ehri, 2005), which posit that as children learn to read (around first grade), they must be explicitly taught orthographic-phonological mappings, and that initial application of this knowledge is slow and labored, presumably because readers are having to remember the phonemes associated with each orthographic pattern. However, proficient readers must learn to automatically apply these phonological-orthographic mappings in order to be able to focus on the meaning of text. This is best illustrated by the classic word superiority effect (letters within words are recognized more readily than isolated letters and nonword letter strings), which is absent in young readers, but present in skilled readers (Joula, Schadler, Chabot, & McCaughey, 1978). In other words, beginning reading requires learning explicit knowledge of correspondences between letter strings and words, and the phonological constituents of words, whereas subsequent reading stages require efficient and automatically applied knowledge of shared orthographic and phonological features across words in order to optimize the efficiency of processing written text.

Under this distinction, we might suppose that initial phonological-orthographic mappings may be supported by the learning function of DM, whereas PM may assist in building on this knowledge to facilitate the automatization of reading. Interestingly, reading theories generally posit that after about 2 years of learning to read, reading should increasingly depend upon a common, flexible set of skills rather than separable strategies. Specifically, the ability to manipulate speech sounds and learning orthography has been found to reciprocally influence each other, such that learning to read reshapes the spoken language system while simultaneously refining reading skill (Perfetti et al., 1987). This prediction is echoed by our observations of distinct associations with DM and PM and reading at Visits 1 and 2, respectively, but not at later Visits. However, is important to acknowledge that the process of learning how to read likely varies within each individual throughout the year, and thus a more detailed time series of reading growth over first to fourth grade, which may perhaps be examined in future studies, could provide deeper insight into this question.

These findings may implicate the importance of timing for the onset of literacy. Evidence suggests that DM learning abilities increase during childhood (DiGiulio, Seidenberg, Oleary, & Raz, 1994). The developmental trajectory of PM is more complex, though PM abilities may even attenuate at

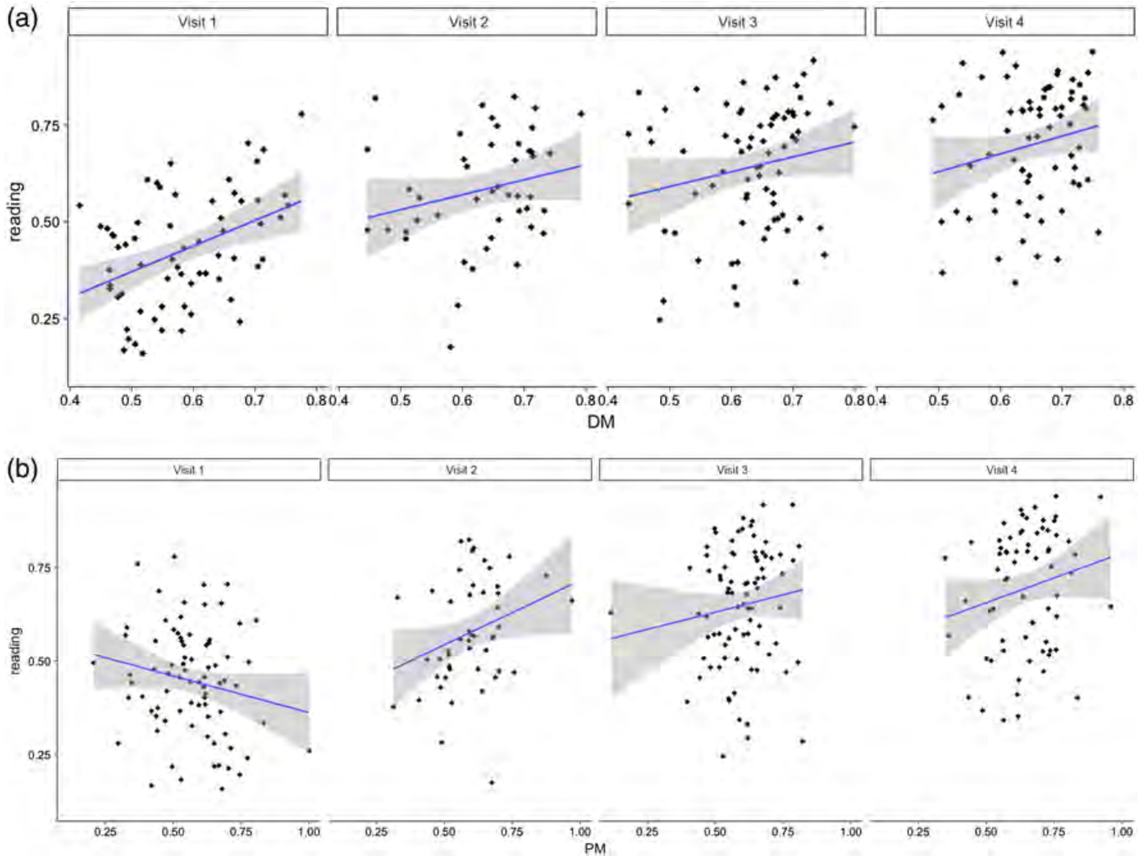

Fig. 1. Interactions between visit and DM and visit and PM on reading ability. DM significantly predicts reading at Visit 1 (1.A), but not at subsequent visits. PM significantly predicts reading at Visit 2 (1.B), but not at other visits. A. The effect of DM on reading by visit B. The effect of PM on reading by visit. Gray bars denote 95% confidence intervals.

the end of childhood (Zwart, Vissers, Kessels, & Maes, 2017). This pattern implies that the division of labor between DM and PM may change over the course of development, raising questions regarding the potential for an optimal window for learning how to read. In particular, while aspects of reading (e.g., orthographic mapping) that are supported by DM may continue to be acquired into adulthood, the subsequent automatization of reading processes through the procedural memory system might become increasingly difficult if the onset of reading is delayed. While this may be a theoretical point rather than of practical consideration in countries with compulsory primary education, this issue is of real material importance for those in developing nations.

The findings may also have implications for struggling readers. Our results suggest that DM abilities may need to sufficiently develop prior to first grade to adequately support the process of learning to read. Furthermore, the successful learning of orthographic mapping may not guarantee a child's ability to acquire fluent, automatic decoding skills if the child struggles with PM in general (Lum et al., 2013; Nicolson & Fawcett, 2011; see Ullman & Pullman, 2015 for discussion).

It is important to acknowledge certain limitations regarding the current study. For example, the children in this study performed on average less well at all the tasks during earlier than later visits (see Appendix B), and thus a potential concern is that our results are reflecting variability in the children's ability to perform tasks in general. However, concerns regarding the validity of the task measures are mitigated by the observation that the DM/PM dependent measures differ in their respective relationships to reading over time. If these measures were simply an index for the ability to perform the tasks, we should observe similar relationships with reading across tasks and across early visits, a pattern that was not observed. Finally, our classical task-based measures of declarative and procedural memory are just that: indexes for functional performance. A future direction of this research is to examine neuroanatomical structures associated with declarative and procedural memory directly, in order to substantiate these relationships between reading and the learning and memory systems.

In conclusion, the current study presents behavioral evidence for the contribution of domain-general learning and memory systems to the acquisition of reading. The results suggest the possibility that domain-general learning and memory may influence learning how to read. The replication and further examination (e.g., via neuroimaging) of these findings may offer the potential for identifying localized weaknesses in learning and memory that may contribute to particular difficulties with reading, and thus may lead to tailored means of remediation. Such questions offer a potentially rich set of future investigations on the domain-general mechanisms supporting reading acquisition.

Acknowledgments—This study was supported by NICHD R21 HD087088, R01 HD067253, R01 HD044073 and U54 HD083211 to L.C., and R21 HD087088 and NSF 1439290 to M.T.U. The RedCap database system was funded by NIH/NCATS UL1 TR000445 awarded to the Vanderbilt Institute for Clinical and Translational Research (VISTR). We thank Laura A. Barquero, Lanier P. Sachs, Angela Sefcik, Abigail E. Randall, and Hannah Reynolds for assistance with the collection of data, as well as Julie Delheimer, Ariel Leonard, Lanier P. Sachs, and Angela Sefcik for assistance with recruitment.

CONFLICT OF INTEREST STATEMENT

The authors have no interest or relationship, financial or otherwise that might be perceived as influencing an author's objectivity to declare as it pertains to this work.

REFERENCES

Barton, K. (2016). *MuMIn: Multi-model inference*. R package version 1.15.6.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1–7.

Berninger, V. W. (Ed.) (1995) *The varieties of orthographic knowledge: II: Relationships to phonology, Reading, and writing*. (Vol. *11*). Berlin: Springer Science & Business Media. https://doi.org/10.1007/978-94-011-0385-5_1

Bitan, T., & Karni, A. (2004). Procedural and declarative knowledge of word recognition and letter decoding in reading an artificial script. *Cognitive Brain Research*, *19*(3), 229–243. https://doi.org/10.1016/j.cogbrainres.2004.01.001

Byrne, B., & Fielding-Barnsley, R. (1989). Phonemic awareness and letter knowledge in the child's acquisition of the alphabetic principle. *Journal of Educational Psychology*, *81*(3), 313–321.

Chall, J. S. (1983) *Learning to read: The great debate*. New York, NY: McGraw-Hill.

Clark, G. M., Lum, J. A. G., & Ullman, M. T. (2014). A meta-analysis and meta-regression of serial reaction time task performance in Parkinson's disease. *Neuropsychology*, *28*(6), 945–958.

Cornelissen, K., Laine, M., Renvall, K., Saarinen, T., Martin, N., & Salmelin, R. (2004). Learning new names for new objects: Cortical effects as measured by magnetoencephalography. *Brain and Language*, *89*(3), 617–622. https://doi.org/10.1016/j.bandl.2003.12.007

DiGiulio, D. V., Seidenberg, M., Oleary, D. S., & Raz, N. (1994). Procedural and declarative memory: A developmental study. *Brain and Cognition*, *25*(1), 79–91. https://doi.org/10.1006/brcg.1994.1024

Eals, M., & Silverman, I. (1994). The hunter-gatherer theory of spatial sex differences: Proximate factors mediating the female advantage in recall of object arrays. *Ethology and Sociobiology*, *15*(2), 95–105. https://doi.org/10.1016/0162-3095(94)90020-5

Edwards, L. J., Muller, K. E., Wolfinger, R. D., Qaqish, B. F., & Schabenberger, O. (2008). An R2 statistic for fixed effects in the linear mixed model. *Statistics in Medicine*, *27*(29), 6137–6157. https://doi.org/10.1002/sim.3429

Ehri, L. C. (2005). Development of sight word Reading: Phases and findings. In M. J. Snowling, & C. Hulme (Eds.), *Blackwell handbooks of developmental psychology. The science of reading: A handbook*. (pp. 135–154). Malden: Blackwell Publishing. https://doi.org/10.1002/9780470757642.ch8

Goswami, U., & Bryant, P. (1990) *Essays in developmental psychology series. Phonological skills and learning to read*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hamrick, P., Lum, J. A., & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(7), 1487–1492. https://doi.org/10.1073/pnas.1713975115

Hedenius, M., Ullman, M. T., Alm, P., Jennische, M., & Persson, J. (2013). Enhanced recognition memory after incidental encoding in children with developmental dyslexia. *PLoS One*, *8*(5), e63998. https://doi.org/10.1371/journal.pone.0063998

Hogan, T. P., Catts, H. W., & Little, T. D. (2005). The relationship between phonological awareness and reading: Implications for the assessment of phonological awareness. *Language, Speech, and Hearing Services in Schools*, *36*(4), 285–293. https://doi.org/10.1044/0161-1461(2005/029)

Hollingshead, A. B. (1975). *Four factor index of social status* (pp. 21–52). New Haven. CT: Yale University.

Howard, J. H., Jr., 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*(7), 1131–1144.

Jaeger, B. (2016). *r2glmm: Computes R squared for mixed (multilevel) models (LMMs and GLMMs)*. R package version 0.1.0.

Joula, J. F., Schadler, M., Chabot, R. J., & McCaughey, M. W. (1978). The development of visual information processing skills related to reading. *Journal of Experimental Child Psychology*, *25*, 459–476.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*, 963–974. https://doi.org/10.2307/2529876

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*, *18*(2), 201–212. https://doi.org/10.1016/0022-0965(74)90101-5

Logan, G. D. (1997). Automaticity and reading: Perspectives from the instance theory of automatization. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, *13*(2), 123–146. https://doi.org/10.1080/1057356970130203

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*(1), 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*(10), 3460–3476. https://doi.org/10.1016/j.ridd.2013.07.017

MacMillan, N. A., & Creelman, C. D. (2004) *Detection theory: A user's guide*. London, England: Psychology Press.

Marsh, G., Friedman, M., Welch, V., & Desberg, P. (1981). A cognitive-developmental theory of reading acquisition. *Reading Research: Advances in Theory and Practice*, *3*, 199–221.

Moeller, J. (2015). A word on standardization in longitudinal studies: don't. *Frontiers in Psychology*, *6*(1389). https://doi.org/10.3389/fpsyg.2015.01389

Nakagawa, S., & Schielzeth, H. (2012). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, in press, *4*, 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nicolson, R. I., & Fawcett, A. J. (2011). Dyslexia, dysgraphia, procedural learning and the cerebellum. *Cortex*, *47*(1), 117–127. https://doi.org/10.1016/j.cortex.2009.08.016

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. https://doi.org/10.1016/0010-0285(87)90002-8

Perfetti, C. A., Beck, I., Bell, L. C., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly*, *33*(3), 283–319.

Poldrack, R. A., & Packard, M. G. (2003). Competition among multiple memory systems: Converging evidence from animal and human brain studies. *Neuropsychologia*, *41*(3), 245–251. https://doi.org/10.1016/S0028-3932(02)00157-4

R Development Core Team (2016) *R: A language and environment for statistical computing, version 3.3*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Reitsma, P. (1983). Printed word learning in beginning readers. *Journal of Experimental Child Psychology*, *36*(2), 321–339. https://doi.org/10.1016/0022-0965(83)90036-X

Roodenrys, S., & Dunn, N. (2008). Unimpaired implicit learning in children with developmental dyslexia. *Dyslexia*, *14*(1), 1–15.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Squire, L. R., Wixted, J. T., & Clark, R. E. (2007). Recognition memory and the medial temporal lobe: A new perspective. *Nature Reviews Neuroscience*, *8*(11), 872–883.

Torgesen, J. K., Rashotte, C. A., & Wagner, R. K. (1999) *TOWRE: Test of word reading efficiency*. Austin, TX: Pro-ed.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1–2), 231–270. https://doi.org/10.1016/j.cognition.2003.10.008

Ullman, M. T., & Pullman, M. Y. (2015). A compensatory role for declarative memory in neurodevelopmental disorders. *Neuroscience & Biobehavioral Reviews*, *51*, 205–222. https://doi.org/10.1016/j.neubiorev.2015.01.008

VanLeeuwen, D. M., Murray, L. W., & Urquhart, N. S. (1996). A mixed model with both fixed and random trend components across time. *Journal of Agricultural, Biological, and Environmental Statistics*, *1*, 435–453. https://doi.org/10.2307/1400438

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999) *Comprehensive test of phonological processing: CTOPP*. Austin, TX: Pro-Ed.

Wijnants, M. L., Hasselman, F., Cox, R. F. A., Bosman, A. M. T., & Van Orden, G. (2012). An interaction-dominant perspective on reading fluency and dyslexia. *Annals of Dyslexia*, *62*(2), 100–119.

Williams, P., & Tarr, M. J. (1997). Structural processing and implicit memory for possible and impossible figures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(6), 1344–1361. https://doi.org/10.1037/0278-7393.23.6.1344

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001) *Woodcock-Johnson III*. Itasca, IL: Riverside.

Zeileis, A., & Hothorn, T. (2002). Diagnostic checking in regression relationships. *R News*, *2*(3), 7–10. Retrieved from http://CRAN.R-project.org/doc/Rnews/

Zwart, F. S., Vissers, C. T. W., Kessels, R. P., & Maes, J. H. (2017). Procedural learning across the lifespan: A systematic review with implications for atypical development. *Journal of Neuropsychology*, *13*, 149–182. https://doi.org/10.1111/jnp.12139

APPENDIX A

Table A1
Demographic Breakdown of Participants at Visit 1

	M *(SD) or* N
Age, *M* (*SD*)	7.49 (.32)
Socioeconomic status, *M* (*SD*)	3.7 (1.06)
Sex	
Male	65
Female	75
Race/ethnicity	
White	85
Black/African American	38
Native American	1
Asian	1
More than one race	14
No response given	0
Hispanic/Latino	
Hispanic or Latino	8
Not Hispanic or Latino	129
No response given	3
Handedness	
Right	95
Left	11
Ambidextrous	29
No response given	5

Note. The numeric value assigned to socioeconomic status is based on Hollingshead's (1975) four-factor socioeconomic status. Handedness was determined by the Edinburgh handedness Inventory (Oldfield, 1970).

APPENDIX B

CHANGES IN READING AND EXPERIMENTAL TASK PERFORMANCES OVER TIME

Changes in reading scores over time

To examine reading during development, we conducted a $4 \times 2 \times 2$ repeated measures analysis of variance (ANOVA; Type-III sum of squares) on reading scores with four levels of Visit, two levels of Word (word/nonword), and two levels of Speed (timed/untimed) as the within-subjects factors on the scaled reading scores. Mauchly's test of multivariate normality indicated that the assumption of sphericity for Visit was violated (Mauchly's $W = .849$, $p = .038$), and thus the Greenhouse–Geisser correction was applied for the effect of Visit. All other contrasts showed equal variance across condition. The ANOVA revealed a significant main effect of Visit ($F_{2.72,198.68} = 458.85$, $p < .001$, $\eta^2 = .863$), a significant main effect of Word ($F_{1,73} = 170.78$, $p < .001$, $\eta^2 = .701$), and a significant main effect of Speed ($F_{1,73} = 399.29$, $p < .001$, $\eta^2 = .845$). We also found a significant interaction between Word and Speed ($F_{1,73} = 31.74$, $p < .001$, $\eta^2 = .303$). All other interactions were not statistically significant (Visit * Time: $F_{3,219} = 1.93$, $p = .125$, $\eta^2 = .026$; Visit * Word: $F_{3,219} = 1.12$, $p = .343$, $\eta^2 = .015$; Visit * Time * Word: $F_{3,219} = 1.47$, $p = .223$, $\eta^2 = .020$). The lack of any significant interactions involving Visit indicated that the changes in scores over time were comparable across Word and Speed.

In order to examine the sources of the main effects, we conducted a series of Bonferroni corrected paired samples t-tests ($\alpha = .05$). We found the main effect of Visit to be driven by significant increases in scores at each Visit (Visit 1 < Visit 2, $t_{73} = 20.34$, $p < .001$, $d = -2.365$; Visit 2 < Visit 3, $t_{73} = -13.44$, $p < .001$, $d = -1.563$; Visit 3 < Visit 4, $t_{73} = -10.99$, $p < .001$, $d = -1.278$). Unsurprisingly, this demonstrates that children are becoming better readers over time.

Table B1
Average Reading Scores by Grade Level

		Visit 1	*Visit 2*	*Visit 3*	*Visit 4*
	N	*98*	*90*	*89*	*79*
	Age	7.49 (.32)	8.46 (.35)	9.44 (.33)	10.43 (.33)
Untimed	Word	38.90 (7.35)	13.68 (6.03)	46.49 (14.67)	18.45 (10.91)
	Nonword	47.06 (8.37)	18.2 (6.65)	58.62 (14.05)	26.34 (12.91)
Timed	Word	51.79 (7.84)	19.93 (6.88)	65.38 (12.69)	30.30 (13.88)
	Nonword	54.65 (7.60)	21.29 (6.70)	70.36 (11.05)	34.36 (14.09)

Note. Means and standard deviations of reading scores at each grade level.

The main effect of Speed was driven by higher performance on untimed over timed reading ($t_{73} = 35.58$, $p < .001$, $d = 4.136$). This indicates that children performed better when reading at their natural pace (untimed) than when encouraged to read more quickly (timed). The main effect of Word was driven by higher performance on words over nonwords ($t_{73} = 29.68$, $p < .001$, $d = 3.450$), indicating that real words are easier for children to read than nonwords (pseudowords), even when both are phonotactically plausible.

In order to examine the nature of the interaction between Word and Speed, we conducted paired comparisons across differences between word/nonword and timed/untimed measures, collapsed across Visits. This revealed the source of the interaction as a greater effect of word in the timed over untimed condition ($t_{73} = 5.63$, $p < .001$, $d = .655$), and a greater effect of time in the word over nonword condition ($t_{73} = 19.74$, $p < .001$, $d = 2.279$). Thus, whether or not trials are speeded has a greater effect on words that are familiar. See Figure B1 for a summary of the developmental trajectories of the untimed (a) and timed (b) scores. See Table B1 for a means and standard deviations of the reading scores at each grade.

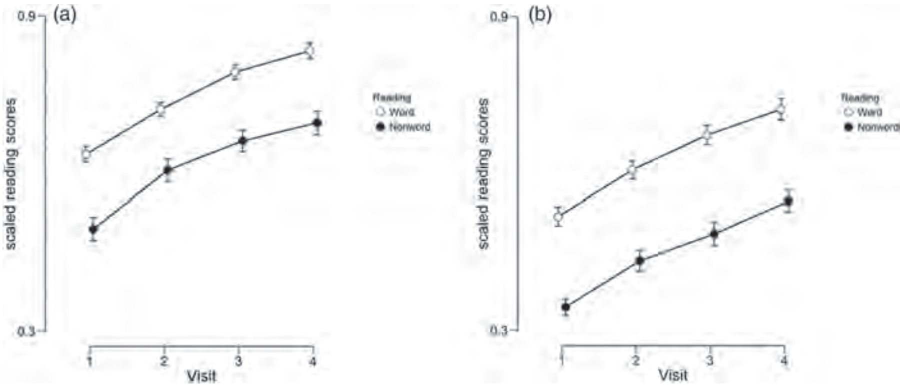

Fig. B1. Developmental trajectory of reading ability. *Note*: Mean proximity to maximum scaling (POMS; Little, 2013) scaled reading ability plotted by Visit, by Word and Nonword. Untimed reading performance are plotted in (a), and Timed performances in (b). Error bars denote 95% confidence interval.

Changes in declarative memory scores over time

Proportions accuracy on the object recognition task were transformed to d' scores (MacMillan & Creelman, 2004) for real and made-up objects separately. Prior recognition memory studies of children have reported a significant advantage for real objects as compared to made-up objects (e.g., Hedenius et al., 2013). Thus, separating d' scores by Object Type allowed us to determine if the developmental trajectory of object recognition differed across real and made up items.

To that end, a 4 × 2 repeated measures ANOVA (Type-III sum of squares) was conducted on object recognition performance by Visit (Visit1–Visit4), with two levels of Object Type (real and made-up objects). As was the case in reading scores, Mauchly's test indicated that the assumption of sphericity for Visit was violated (Mauchly's $W = .369$, $p = .018$), and thus the Greenhouse–Geisser correction was applied for the effect of Visit. All other contrasts met assumptions of multivariate normality. The ANOVA revealed a significant main effect of Visit ($F_{1.79,26.87} = 3.43$, $p = .050$, $\eta^2 = .188$), a significant main effect of Object Type ($F_{1,15} = 76.20$, $p < .001$, $\eta^2 = .836$), but no interaction between Visit and Object Type ($F_{3,45} = 1.56$, $p = .216$, $\eta^2 = .093$).

In order to examine the sources of the main effects, we conducted a series of paired samples t-tests (Bonferroni correction applied). We found the main effect of Visit to be driven primarily by an increase in scores between Visits 1 and 4 ($t = -3.415$, $p = .011$, $d = -.854$), with no significant differences observed between Visits 1 and 2, Visits 2 and 3, and Visits 3 and 4 ($t_{43} = -2.613$, $p = .082$, $d = -.653$; $t_{43} = .924$, $p > .999$, $d = .231$; $t_{43} = -1.52$, $p = .831$, $d = -.380$, respectively). Thus, declarative memory scores show evidence of developmental change over time.

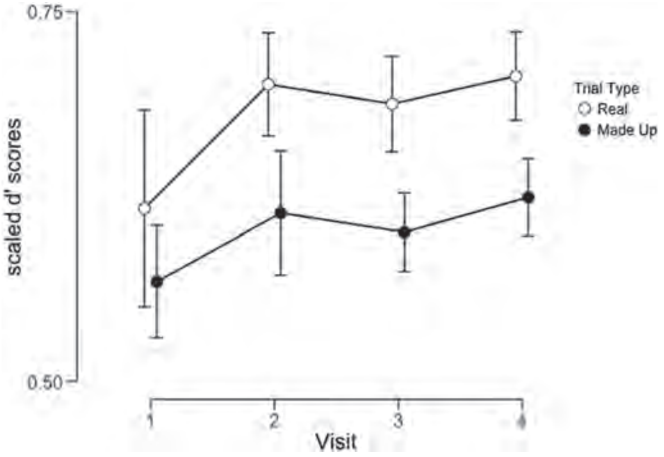

Fig. B2. Developmental trajectory of object recognition scores. *Note.* Mean POMS scaled object recognition performance across Visit by, and across, Trial Types. Error bars denote 95% CIs of the mean. Means and standard deviation of real and made-up trials are as follows: Visit 1: 1.12 (1.09), .55 (.73); Visit 2: 1.52 (.98), .87 (.72); Visit 3: 1.55 (.95), .90 (.69); Visit 4: 1.85 (.76), 1.02 (.60).

As expected, the main effect of Object Type was driven by higher performance on real than made-up objects ($t_{14} = 9.85$, $p < .001$, $d = 2.461$). However, we did not find a significant interaction between Visit and Object Type. While the lack of interaction may be due to parallel developmental trajectories of real and made-up objects, it is also possible that no interaction was observed due to the high collinearity between real and made-up objects across Visits ($N = 322$, Pearson's $R = .75$, $p < .001$). Either way, we collapsed scores across Trial Types into an average object recognition score per participant per Visit, prior to inclusion in the primary analyses. See Figure B2 for a graphical summary of the developmental trajectory of scaled object recognition performance by Object Type.

Changes in procedural memory (PM) over time

First, we determined whether or not our sample demonstrated evidence of procedural learning, by conducting a 2 × 4 repeated measures ANOVA (Type-III sum of squares) on the raw average reaction times, with two levels of Trial Type (Final Random Block/Final Sequence Block), and four levels of Visit, as within-subjects factors. Mauchly's test confirmed that the data met assumptions of sphericity. The ANOVA revealed a significant main effect of Visit ($F_{3,54} = 65.66$, $p < .001$, $\eta^2 = .785$), and a significant interaction between Trial Type and Visit ($F_{3,54} = 2.82$, $p = .048$, $\eta^2 = .135$). There was no main effect of Trial Type ($F_{1,18} = 2.82$, $p = .108$, $\eta^2 = .137$). We examined the source of the main effect and interaction by conducting a series of paired samples t-tests (Bonferroni correction applied). We found that RTs significantly decreased at each visit (Visit 1 > Visit 2, $t_{36} = 10.90$, $p < .001$, $d = 2.501$; Visit 2 > Visit 3, $t_{36} = 3.58$, $p = .006$, $d = .822$; Visit 3 > Visit 4, $t_{36} = 4.50$, $p < .001$, $d = 1.032$). In examining the Trial by Visit interaction, we included all data available for the contrast of interest. In doing this, we found that the Trial Type by Visit interaction to be driven by a lack of significant differences between Final Random and Final Sequence trials at Visits 1 and 2 ($t_{79} = -.82$, $p = .413$, $d = -.092$; $t_{47} = 1.33$, $p = .190$, $d = .192$, respectively), with a trending difference at Visit 3 ($t_{82} = 1.95$, $p = .055$, $d = .214$), and a significantly reduced RT for Final Sequence versus Final Random trials at Visit 4 ($t_{67} = 5.05$, $p < .001$, $d = .613$).

These results suggest that, while a steady decrease in RT over the four visits indicates improvement of general skills contributing to increased motor speed, our specific measure of PM appears to emerge gradually over the course of the Visits on average. However, we note that the lack of a

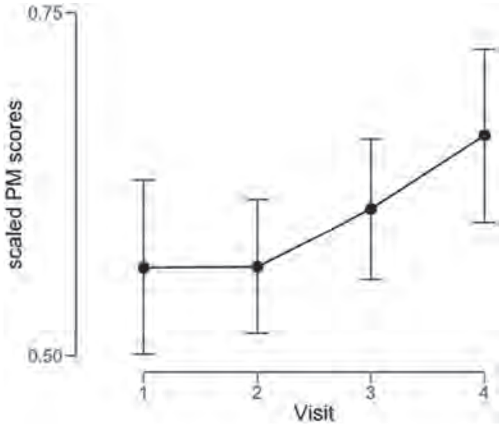

Fig. B3. Developmental trajectory of SRT scores. *Note*. Reaction time (RT) by Visit for the operationalized measure of procedural memory (PM), that is, the difference between random and sequence trials (mean scaled RT in random trials—mean scaled RT in sequence trials), scaled using the POMS method. Error bars denote 95% CIs of the mean. Means and standard deviations for scaled PM scores are as follows: Visit 1: .56 (.15); Visit 2: .59 (.12); Visit 3: .60 (.11); Visit 4: .65 (.12).

statistically significant separation between reaction times for sequence and random trials at the earlier visits does not negate the predictive value of task performance in our individual differences approach: the wide variability in PM scores suggest that some students did show evidence of learning even if the group, on average, did not. See Figure B3 for a graphical summary of operationalized procedural learning scores across the four visits.

APPENDIX C

CORRELATION MATRIX BETWEEN TASK PERFORMANCES BY VISIT

Table C1
Reliability Estimates

	Visit 1	*Visit 2*	*Visit 3*	*Visit 4*
PM	0.942	0.855	0.909	0.925
DM	0.841	0.833	0.814	0.79
Reading	0.893	0.923	0.921	0.909

Note. Reliability estimates (Chronbach's α) is provided by construct by Visit for procedural memory (PM: reaction time for random—standard trials), declarative memory (DM: d' for real and made up trials), and reading (word and nonword reading across timed and untimed conditions). All reliability measures are good to excellent, except for DM at visit 4 which is considered acceptable.

APPENDIX D

Table D1
Correlation of Measures at Visit 1

		Untimed		*Timed*	
	DM	*Word*	*Nonword*	*Word*	*Nonword*
PM	−0.2	−0.1	−0.03	−0.14	−0.13
DM		0.45*	0.38*	0.33*	0.42*
Word (untimed)			0.82*	0.84*	0.82*
Nonword (untimed)				0.67*	0.87*
Word (timed)					0.76*

*Significance at .05 level (uncorrected).

Table D2
Correlation of Measures at Visit 2

		Untimed		*Timed*	
	DM	*Word*	*Nonword*	*Word*	*Nonword*
PM	0.13	0.3*	0.26	0.3*	0.23
DM		0.1	0.19	0.16	0.16
Word (untimed)			0.88*	0.86*	0.86*
Nonword (untimed)				0.76*	0.9*
Word (timed)					0.82*

*Significance at .05 level (uncorrected).

Table D3
Correlation of Measures at Visit 3

		Untimed		*Timed*	
	DM	*Word*	*Nonword*	*Word*	*Nonword*
PM	0.14	0.1	0.15	0.08	0.11
DM		0.25*	0.16	0.19	0.16
Word (untimed)			0.92*	0.82*	0.86*
Nonword (untimed)				0.78*	0.88*
Word (timed)					0.79*

*Significance at .05 level (uncorrected).

Table D4
Correlation of Measures at Visit 4

		Untimed		*Timed*	
	DM	*Word*	*Nonword*	*Word*	*Nonword*
PM	0.04	0.2	0.13	0.24	0.2
DM		0.23	0.3	0.05	0.13
Word (untimed)			0.89*	0.71*	0.9*
Nonword (untimed)				0.66*	0.91*
Word (timed)					0.8*

Note. Correlations are expressed in Pearson's r. While the measures of word-level reading are collapsed across the four measures in our analyses, we report here the correlations between each reading measure and our measures of DM and PM separately for those who are interested in potential differences in associations between timed/untimed, word/nonword measures.
*Significance at .05 level (uncorrected).