Supplemental Online Material

Understanding dyslexia through personalized large-scale computational models

Conrad Perry[1,*], Marco Zorzi[2,3], & Johannes C. Ziegler[4]

[1] Faculty of Health, Arts and Design, Swinburne University of Technology, Hawthorn, Australia

[2] Department of General Psychology and Padova Neuroscience Center, University of Padova, Padova, Italy

[3] Fondazione Ospedale San Camillo IRRCS, Venice-Lido, Italy

[4] Aix-Marseille University, Centre National de la Recherche Scientifique, Laboratoire de Psychologie Cognitive, Marseille, France

* Correspondence: conradperry@gmail.com

Materials and Methods

Behavioral data

All children (N = 622) who had full data on the critical component tasks (phonological awareness, orthographic choice, vocabulary) and who had scores on regular, irregular, and nonword reading were selected from a larger database (N = 1189) that was generously provided by Bruce Pennington, Richard Olson, and Robin Peterson. The database included all of the children documented in Peterson et al. (2013). Note that we did not use any other selection criteria than having complete data on all critical measures. Further information about the testing procedures and diagnostic criteria can be found in the original study. The critical component and reading tasks used were the following:

Phonological processing. This was assessed with a phoneme deletion test. The phoneme deletion test consisted of six practice and 40 test trials presented in two blocks and required subjects to repeat a nonword, then remove a specific phoneme (when done correctly, a real word resulted—e.g., 'Say 'prot'. Now say 'prot' without the '/r/'). Note that the database included two other tasks that we could have used to parameterize phonological processing: phonological choice

(participants chose which of three nonwords sounds like a word) and Pig Latin (participants strip the first phoneme of a word, pronounce the word without the phoneme, and then use a second syllable with the onset of the first syllable plus the vowel /eɪ/). We did not use the phonological choice task because it taps whole-word phonological knowledge and it requires reading the nonwords aloud (prior to the phonological decision). Both phoneme deletion and Pig Latin tasks provide a purer measure of phonological processing, but the latter is more complex (it requires a bigger number of phonological operations) and it is far less commonly used than phoneme deletion when examining the development of reading and reading disorders (e.g., Landerl et al., 2013; Ziegler et al., 2010).

Vocabulary. Vocabulary knowledge was measured with the Vocabulary subtest from the Wechsler Intelligence Scale for Children—Revised.

Orthographic processing. This was assessed with an orthographic choice test (Olson, Forsberg, Wise, & Rack, 1994). The orthographic choice test included 80 real word/pseudohomophone pairs (e.g., easy–eazy, fue–few, salmon–sammon) presented in two blocks and required participants to select the real word. Note that the database included another task that we could have used to parametrize orthographic processing, that is, homophone choice (participants decide which of two possible homophones corresponds to a statement which details the meaning of only one of the homophones). However, homophone choice examines whether participants know which spelling corresponds to a given meaning, whereas orthographic choice only requires visual word recognition. Therefore, orthographic choice offers a purer measure of orthographic processing than homophone choice. Moreover, homophone choice relies on word meaning and is thus likely to have more overlap with vocabulary measures.

Reading aloud measures. Nonword reading was assessed with a nonword reading test (Olson et al., 1994). The nonword reading test was presented in two blocks and included 85 items of varying difficulty levels (e.g., strale, lobsel). Regular and irregular word reading was assessed with the set of words used by Castles and Coltheart (Castles and Coltheart, 1993) that included 30 irregularly spelled words (e.g., island, choir) and 30 regular words of varying difficulty.

Program Availability

A fully executable version of the model that runs under the Windows operating system as well as the data generated in this paper can be found at the following websites:

https://sites.google.com/site/conradperryshome/

http://ccnl.psy.unipd.it/CDP.html

Simulation Methods

A brief description of the Connectionist Dual-Process Model

The core architecture of the models is taken from the Connectionist Dual-Process (CDP) model (Perry, Ziegler, & Zorzi, 2007, 2010), as implemented in its latest version CDP++.parser (Perry, Ziegler, & Zorzi, 2013) which is depicted in Supplementary Figure 1. There are two relatively separate processing pathways ("routes") in the model. One is a lexical route that includes the orthographic and phonological word forms. The other is a sublexical route that computes the phonology of words without knowledge of the whole-word form. Both of these routes share letter features and letter representations, as well as output nodes for phonemes and stress.

The basic function of the lexical route is to allow the whole word form of words to be stored and recalled. In the orthographic lexicon, there is a single node for each spelling, and in the phonological lexicon, there is a single node for each phonological word. At the letter level, the orthographic form of words is simply represented as a contiguous set of letters, and at the letter feature level, the visual patterns of the letters are represented. At the phoneme level, the phonological form of words uses a representation that is structured in terms of its speech form, with phonemes being organized into a syllabic template. This template has slots for phonemes that are organized according to an onset-vowel-coda distinction. It allows three phonemes in the onset, one in the vowel, and four in the coda for each of two possible syllables. Stress information is also stored (i.e., whether the word has first or second syllable stress).
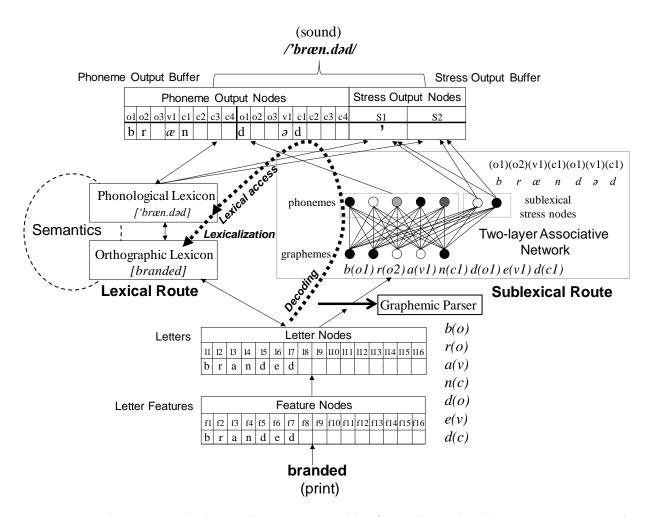
Figure S1. The Connectionist Dual-Process Model of Reading Aloud (CDP++.parser version; Perry, Zielger, & Zorzi, 2013). Note: f = feature, l = letter, S = Stress, o = onset, v = vowel, c = coda. Numbers correspond to the overall slot number within the Feature, Letter, and Stress nodes, or the particular slot within an onset, vowel, or coda grouping for other representations. The thick divisor in the Phoneme Output Buffer represents a syllable boundary. The thick dotted lines represent how self-teaching occurs (i.e., letters→ sublexical decoding→ output nodes→ phonological lexicon→ orthographic lexicon).

Processing dynamics of nodes in the feature and letter level, the orthographic and phonological lexicons, and the phoneme and stress output buffers is based on standard interactive-activation equations (McClelland & Rumelhart, 1981), where all inputs into a given node are first summed and then transformed using a sigmoid function. This includes input from other nodes, and, with the phonological and orthographic lexicons, inhibitory input from a frequency scaling parameter

that is proportional to the log frequency of the word. Injection of noise into these representations (as in our dyslexia simulations) is done at the summing stage (i.e., before the nonlinear transformation).

The basic function of the sublexical route is to generate the phonology of letter strings without lexical information. This is important as it represents one way the model can read words without being previously exposed to their orthographic form. There are a number of steps in this process. The first involves the graphemic parser of the model. The graphemic parser is a simple two-layer network which learns to break letter strings into graphemes and then assign them to a slot in the input layer of the two-layer associative (TLA) network. This layer consists of a syllabically organized template where graphemes are organized according to an onset-vowel-coda structure that is largely homologous with the phoneme organization described above. Since the parser has no knowledge of the lexical form of a word, however, it can potentially parse words in ways that are not similar to how they might be represented lexically. The graphemic representation of the letter string is then propagated through the TLA network, where the activation of phoneme nodes is computed in the standard way by dot product of input and weight patterns followed by a nonlinear (sigmoid) transformation. Finally, these values are propagated to the phoneme output nodes, where they are pooled with activation coming from the lexical route.

The model works differently depending on whether a word is in training mode or whether it is being read aloud. When reading a word aloud, a string of letter features is first activated, and the model iterates through the processes described above until activation criteria in the phoneme and stress output buffers are satisfied. In learning mode, the graphemes and phonemes in a word are aligned in the TLA network, and the TLA network is then trained. The training rule used by the TLA network is the delta rule (formally equivalent to the Rescorla-Wagner learning rule; Sutton & Barto, 1981), and since the network only has two layers, this means only linear relationships between graphemes and phonemes can be learnt.

One limitation of the graphemic parsing mechanism is that, in very rare circumstances, a disyllabic word may be parsed into three orthographic syllables. This happened in the present study for the word *colonel* (which was included in the Castles and Coltheart (1993) word set). This word was therefore removed from the lexicon of the model and it was not used to calculate the percentage of correct words in that set. Control simulations where this word was left in and could be learnt via direct instruction produced virtually identical results.

New mechanisms

New learning dynamics and mechanisms were introduced to capture reading acquisition within a realistic learning environment. These include:

1) The learning method described in Ziegler et al. (2014), where the model was first trained on a small set of grapheme-phoneme correspondences (listed in the Appendix) and then words were added to the orthographic lexicon if they were successfully decoded through the decoding network -- that is, when the phonemes derived from letters were able to activate the correct word in the phonological lexicon of the model. When a word was successfully decoded and added to the orthographic lexicon or if it was already in the orthographic lexicon, the decoding network was trained on that word.

2) A novel lexicalization method reflecting the probabilistic nature of lexicalization and memory consolidation, as well as the fact that learning can occur via direct teaching and other methods that do not necessarily need self-generated decoding.

2.1) Lexicalization was made probabilistic. In particular, rather than a word being lexicalized every time it passed the activation threshold in the phonological lexicon of the model via decoding, it was only lexicalized some proportion of the time. This proportion was linked to the orthographic choice parameter: the better a child was at orthographic learning, as estimated by his or her performance on the orthographic choice task, the higher the probability that the word entered the orthographic lexicon. This assumption allows inter-individual differences in orthographic learning to occur that do not depend on decoding.

2.2) Words were given a chance of being lexicalized by direct instruction if they did not reach the threshold for decoding or were not lexicalized after decoding. The probability that a given word would become a candidate for lexicalization via direct learning was simply a function of its frequency (i.e., log [frequency of target word +2] / log [frequency of highest frequency word +2]). In practice, this means that words of a very low frequency have about a 5% chance of being selected for direct learning after not being successfully decoded.

3) A child-specific vocabulary, which in its full version included all words (N = 9663) of the CELEX database (Baayen, Piepenbrock, & van Rijn, 1993) that had an age-of-acquisition rating of 10 years or less (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012) and only one or two

syllables. The word *colonel* was not included (see above). The number of word presentations (i.e., learning events) for a model with the full vocabulary was 57978, which is equivalent to 6 passes (i.e., training epochs) through the full database (9663×6). For models with a smaller vocabulary, the number of word presentations was reduced proportionately by keeping the number of training epochs the same as for the full vocabulary model. On average, there were 6423 words in the MDM's simulations (across all 622 children), and thus the average number of word presentations was 38538. The three alternative models all used the same vocabulary parameter as the MDM, and thus the number of words used for each simulation of each child was very similar to the MDM. The order of presentation of the words was random in the first epoch and the same random order was used in successive epochs.

4) The presence of noise during learning, which implies that the results are non-deterministic. Therefore, all simulations were run 10 times and the average of the results was taken for subsequent analyses. Overall, the simulations required around 240 million word presentations / learning events (i.e., 622 subjects $\times$ 38538 words $\times$ 10 repeats). Despite the systematic use of supercomputing facilities, the computational burden was too large to run the model with a full lexicon (the final simulations reported here took approximately 20,000 hours of computing time). During learning, a reduced "runtime" lexicon was therefore compiled by taking the word/nonword presented to the model and all words that were 1st or 2nd order phonological or orthographic neighbors (Coltheart, 1978). For words differing in length, each letter or phoneme different was counted as one neighbor different (i.e., *dog* and *dogs* were counted as 1 neighbor different). This meant that, for a full vocabulary model, the "runtime" phonological lexicon included on average 71.28 (SD: 98.8) words (even when the orthographic lexicon had no words yet). During model testing (i.e., after the learning phase), the same restriction was used but the "runtime" lexicon also included all words that had the same first letter/phoneme as the word being tested, as well as any words that had the same phoneme as the regularized first grapheme of the word (i.e., the phoneme based on simple spelling-sound translation rules, see (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001)). This was done because it meant that highly irregular words like *whole* /hɒl/ had lexical competitors that had the most common pronunciation of the grapheme used by the word (/w/, e.g., *one*, *word*, *wart*) as well as the phoneme used in the lexical form (i.e., /h/, e.g., *hope*). Thus, during testing, a stimulus could activate on average 796.2 (SD: 412.1) words in the phonological lexicon (for a full vocabulary model).

Parameter settings

To set the parameter values, we used the method in Ziegler et al. (2008), whereby each individual started with the same parameter set and these values were modified based on individual performance in the subcomponent tasks. To get the distributions of parameters for each individual, a high and a low value was chosen based on the child who scored the worst on a particular task and the child who scored the best. All other children were then given a score between these two values based on simple linear interpolation. For example, if the parameter values varied between 0 and 1 and the task scores went from 0 to 5, then the parameter value given for the child that scored 2.5 on a task would simply be .5.

For each child, the individual set of model parameters was determined on the basis of performance in the subcomponent tasks in the following way:

1) The orthographic choice task was used to index the level of noise in the orthographic lexicon. This was computed for each word by taking the parameter of each child (based on his or her performance in the orthographic choice task) and multiplying it with a number sampled from a Z-distribution. This product was then added to the net input of the Interactive Activation equations (Perry et al., 2007) used for each lexical item.

2) The orthographic choice task was also used to index the probability at which a child lexicalizes a word after either successfully decoding it or being successfully given it via direct learning. This was determined exactly the same way as the level of noise in the orthographic lexicon, with the child with the best score having a 100% chance of lexicalization and those with a lower score having a lower chance.

3) The phoneme deletion task was used to set how much noise was generated in the decoding network during learning for each participant. Based on this parameter, for a given word presented to the model, phonemes that would be active in the output were turned off with a certain probability and another phoneme in the same syllabic position was turned on (i.e., if a phoneme was switched off in the first onset position, another phoneme was always turned on in the first onset position). The replacement phoneme was not chosen purely randomly but based on phonetic similarity (e.g., /p/ is more likely to be switched with /b/ than with /m/, see Ziegler et al. (2014)).

4) The vocabulary score was used to set how many words were in the phonological lexicon of each participant. The function used to determine whether a word should be in-or-out of the lexicon

was weighted towards keeping high over low frequency words. This was done by first calculating a value for each word based on its frequency (log [frequency of target word +2] / log [frequency of highest frequency word +2]). For each word, a random number between one and zero was then generated and multiplied by the vocabulary parameter. If the value of this number was less than the value calculated from word frequency, the word was kept in the lexicon; otherwise it was not. On average (i.e., across the 622 individual models), this meant that 66.5% of the words were in the lexicon; the lexicon of the child with the smallest vocabulary contained only 36.1% of all possible words.

The parameters that were manipulated in the MDM and in the alternative models (see below) to simulate individual differences across the children are listed in Table S1. The parameter values for the MDM were found by choosing an initial set by hand and then making minor modifications to them so that they produced similar overall means to the children. All other parameters were identical to those reported in Perry et al. (1) with two exceptions: the Letter-to-Orthography inhibition parameter was set lower (from -1.5 to -0.7, which meant incorrect lexical entries were more likely to get activated) and the lexicon frequency scaling parameter was also set slightly lower (from .15 to .10, which meant that the effect of word frequency on the resting activation levels of word nodes was smaller). The models also used an identical threshold to identify when successful decoding occurred (.15).

Table S1. Parameters that varied across the models

Parameter	**Model**			
	Multi-deficit	Global Noise	Phonological Deficit	Visual Deficit
Letter Noise		0 - 0.008		
Letter Switching				0 - .15
Orthographic Noise	0 - .16	0 - 0.008		
Phonological Noise		0 - 0.008		
Phoneme Noise		0 – 0.008		
Phoneme Switching	0 - .78		0.0 - .92	
Lexicalization Threshold	.01 – 1	.7	.6	.55
Vocabulary	0 - .80	0 - .80	0 - .80	0 - .80

During learning, the following parameter changes were made to all models so that items in the phonological lexicon could be activated comparatively more easily: Phonological Lexicon to Phoneme Buffer inhibition: 0; Phonological Lexicon to Phoneme Buffer excitation: 0; Phonological Lexicon lateral inhibition: -.03; Phoneme Buffer to Phonological Lexicon Inhibition: -.02.

Alternative Models

The multi-deficit model was compared with three simpler models. Two implemented single-deficit hypotheses (a phonological deficit and a visual deficit model), and the third a hypothesis examining the effect of using the same distribution of noise across all representations (a global noise model). These differed in how and where noise was applied, and all used the same vocabulary parameterization as the MDM. All three of the models used a fixed probability of lexicalization, and an attempt was made to try to find a parameter set that caused the models to show a pattern as close as possible to the overall means as the actual data. This was done by starting with the MDM parameter values and setting all of the parameters not associated with the specific model to zero. The parameters left were then set to a point that produced results as similar as possible to the overall means. These parameters were found using a hand search where the parameter that was modified for each model was changed in conjunction with the lexicalization threshold parameter. Vocabulary score was also used in all alternative models to set the size of the individual phonological lexicon (as for the MDM model). The specific processing assumptions of the models were:

1. Phonological deficit model. The critical parameter was the probability of a phoneme being switched during learning, which was derived from the phoneme deletion scores of each child (in the same way as in the multi-deficit model).

2. Visual deficit model. Letters at the letter level were switched with adjacent letters with a probability that was determined from the orthographic choice scores of each child. Switching was assessed for each letter in each position, starting from the first letter and excluding the last one. If switching occurred, the letter was switched with the letter to the right of it.

3. Global noise model. Noise was added to each processing level (letter level, orthographic lexicon, phonological lexicon, and phoneme output buffer) whenever the model was run. The

amount of noise was determined by a parameter that was set for each individual based on the mean performance he or she had on the regular, irregular, and nonwords.

Model Evaluation and Model Comparisons

As mentioned above, the predicted reading scores for each child were computed by averaging 10 simulation runs with the model (due to the non-deterministic nature of the learning process). The model scores were then compared with the actual child scores (i.e., individual dots in Fig. 5, S2 and S3). Note that there was no fitting procedure to minimize the prediction error on the distribution of reading scores (this would be computationally unfeasible given the stochastic nature of the model). The parameters described above (e.g., probability of lexicalization, size of phonological lexicon, etc. see Table 1) were simply set to vary within a range that allowed the model to produce mean scores similar to the mean across all participants (see also above). This implies that the model predictions on the full distribution of reading scores are not tied to the dataset and are not influenced by specific cases (i.e., overfitting is not possible). The same procedure was used for the alternative models, and, as can be seen in the results in the main text (Fig. 4) and Table S2, the mean results are very similar to the actual results found for all but the global noise model.

Table S2. Mean overall percentage correct scores for the three word types and summed squared error differences between the model scores and the human data

Dataset	Percentage Correct			Summed squared error (SSE)			
	Regular	Irregular	Nonword	Regular	Irregular	Nonword	Total SSE
Human Data	88.66	72.75	67.09				
MDM	88.82	71.79	64.79	0.03	0.93	5.32	6.27
Visual Def	86.01	70.02	66.66	7.00	7.45	0.19	14.63
Phon Def	89.16	71.27	65.25	0.26	2.18	3.41	5.84
Global Noise	86.60	54.69	77.66	4.24	326.15	111.71	442.11

Note: Phon = Phonological; Def = Deficit

Despite the small differences in overall means, inspection of Figure 5, Figure S2, and Figure S3 shows that all of the single deficit models produce distributions of data that differ considerably to those found in the actual data. These cannot be fixed by any simple modifications to the parameters used. In particular, with the phonological deficit model, the distribution of irregular word scores is too tight compared to the actual data. This is because the phoneme switching parameter affects nonwords more than irregular words. Thus, if this parameter is increased to try to widen the distribution of irregular word scores, nonword performance drops below the overall mean results. There is also a less obvious difference with the nonwords, where the model produces a more sigmoidal function than the MDM when the correct function should look linear. This also cannot be simply fixed, because alternative values of the phoneme switching parameter decrease the fit to the overall means, whereas with the MDM, nonwords are also affected by orthographic noise and this makes the distribution of simulated scores more similar to the distribution observed in the human data.

With the visual deficit model, the distribution is too restricted with all groups of words. This cannot be fixed by increasing the range of noise, because this causes the performance of the model to drop too low. With the global noise model, there appeared to be no set of parameters that can could be chosen to get the model to display a pattern of means similar to the means of the children. The reason for this is that injecting the same level of noise in all representations causes a much larger detriment to irregular word performance than nonword performance, a pattern also observed by Nickels, Biedermann, Coltheart, Saunders, and Tree (2008) in simulations with the Dual-Route cascaded model of Coltheart, Rastle, Perry, Langdon, and Ziegler (2001). With nonwords, noise increases the competition between alternative phonemes, but this does not necessarily cause poor performance – for example, both /zu:d/ or /zʊd/ are reasonable pronunciations of *zood*. Alternatively, with irregular words, increased competition from (incorrect) phonemes may prevent the correct phoneme from becoming the most active (due to lateral inhibition), thereby leading to a word error.

In terms of model comparisons, we provide r^2 as well as BIC scores. The latter were computed as: BIC = n + n ln (2π) + n ln(RSS/n) + (ln n) (p + 1), where n is sample size, p is the number of free parameters, and RRS is the residual sum of squares (i.e., sum of the squared prediction errors). Note that this formula is often used without the two initial terms (though here it is identical to the one used in the base package of R software). The MDM was considered to have 4 free parameters

(phoneme switching, orthographic noise, lexicalization threshold, vocabulary; see Table S1). The three alternative models were considered to have 2 free parameters based on vocabulary and one model-specific parameter.

References

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). The CELEX Lexical Database (CD-ROM). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition*, 47(2), 149-180.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), Strategies of information processing (pp. 151-216). London: Academic Press.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108(1)*, 204-256.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavioral Resesearc Methods*, 44(4), 978-990.

Landerl, K., Ramus, F., Moll, K., Lyytinen, H., Leppanen, P. H., Lohvansuu, K., . . . Schulte-Korne, G. (2013). Predictors of developmental dyslexia in European orthographies with varying complexity. *Journal of Child Psychology and Psychiatry*, 54(6), 686-694.

McClelland, J. L., & Rumelhart, D. E. (1981). An Interactive Activation model of context effects in letter perception: 1. An account of basic findings. *Psychological Review*, 88(5), 375-407.

Nickels, L., Biedermann, B., Coltheart, M., Saunders, S., & Tree, J. J. (2008). Computational modelling of phonological dyslexia: how does the DRC model fare? *Cognitive Neuropsychology*, 25(2), 165-193.

Olson, R. K., Forsberg, H., Wise, B., & Rack, J. (1994). Measurement of word recognition, orthographic, and phonological skills Frames of reference for the assessment of learning disabilities: New views on measurement issues (pp. 243-277). Baltimore, MD, US: Paul H Brookes Publishing.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. *Psychological Review*, 114(2), 273-315.

Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology*, 61(2), 106-151.

Perry, C., Ziegler, J. C., & Zorzi, M. (2013). A Computational and Empirical Investigation of Graphemes in Reading. *Cognitive Science*, 37(5), 800-828.

Peterson, R. L., Pennington, B. F., & Olson, R. K. (2013). Subtypes of developmental dyslexia: Testing the predictions of the dual-route and connectionist frameworks. *Cognition*, 126(1), 20-38.

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., . . . Blomert, L. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science*, 21(4), 551–559.

Ziegler, J. C., Castel, C., Pech-Georgel, C., George, F., Alario, F. X., & Perry, C. (2008). Developmental dyslexia and the dual route model of reading: Simulating individual differences and subtypes. *Cognition, 107*, 151–178.

Ziegler, J. C., Perry, C., & Zorzi, M. (2014). Modelling reading development through phonological decoding and self-teaching: implications for dyslexia. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369 (1634), 20120397.

Appendix. Grapheme-phoneme correspondences used for initial explicit teaching

Grapheme	Phoneme		Grapheme	Phoneme
A	{		Nn	n
Ae	1		O	Q
Ai	1		Oa	5
Au	9		Oe	5
Augh	$		Oi	4
Ay	1		Oo	u
B	b		Ou	6
c	k		ow	6
Ch	J		Oy	4
Ck	k		P	p
D	d		Ph	f
E	E		Pp	p
Ea	i		R	r
Ee	i		Rr	r
Ei	1		S	s
Eigh	1		sh	S
Eu	u		ss	s
Ew	u		t	t
Ey	1		tch	J
F	f		th	T
Ff	f		tsch	J
G	g		tt	t
Gn	n		u	V
H	h		ue	u
I	I		ui	u
Ie	2		uy	2
J	_		v	v
K	k		w	w
Kn	n		wh	w
L	l		wr	r
M	m		y	2
N	n		z	z
Ng	N			

Note: Phonemes are in the format of the CELEX database
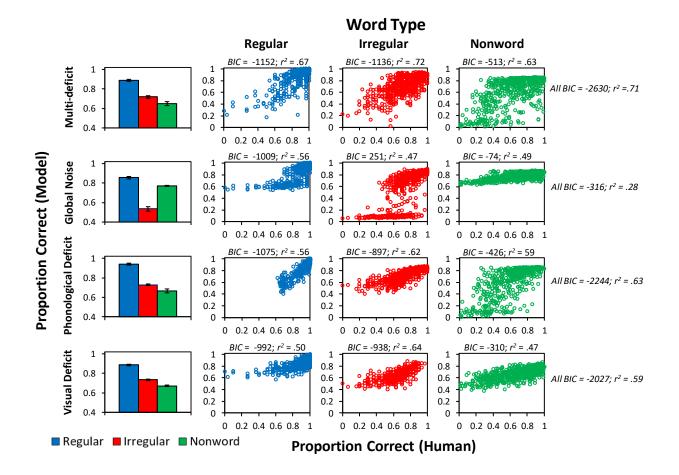
Supplementary Figures

Fig. S2. Predicted versus actual reading performance for all children (mean scores in the leftmost column) with the multi-deficit, global noise, phonological deficit, and visual deficit model. BIC = Bayesian Information Criterion.
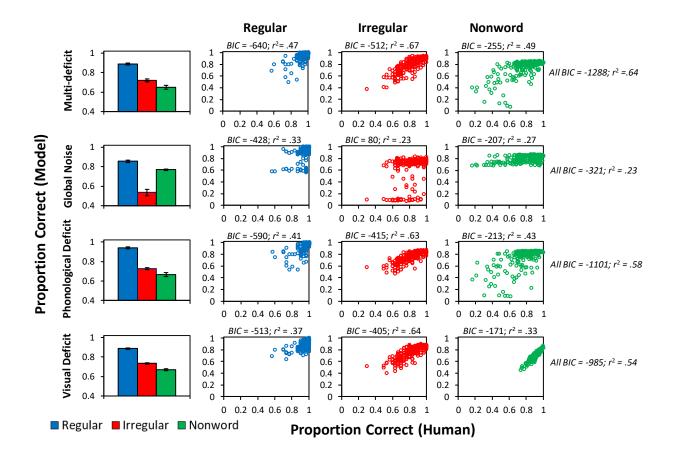
Fig. S3. Predicted versus actual reading performance for the normally developing children (mean scores in the leftmost column) with the multi-deficit, global noise, phonological deficit, and visual deficit model. BIC = Bayesian Information Criterion.