Intervening to alleviate word-finding difficulties in children: Case series data and a neurocomputational foundation

Best, W. 1, Fedor, A. 2, Hughes, L. 1, Kapikian, A. 3, Masterson, J. 3, Roncoli, S. 3, Fern-Pollak, L. 3 & Thomas, M. S. C. 2

1 Division of Psychology & Language Sciences, University College London, UK
2 Department of Psychological Sciences, Birkbeck College London, UK
3 Department of Psychology and Human Development, Institute of Education, UK

Running head: Interventions for word-finding difficulties

Address for correspondence:

Prof. Wendy Best
Department of Developmental Science
2 Wakefield Street
London
WC1N 2PF
Email: w.best@ucl.ac.uk
Tel.: +44 (0)20 7679 4257 (Ex: 24257)
Fax: +44 (0)20 7713 0861
Abstract

In this study, we utilised intervention as a key method to investigate causal theories of word-finding difficulties. Two convergent methodologies were used. First, in two case studies of children with WFD, we evaluated the relationship between the specific pattern of difficulties and the most effective behavioural intervention. Second, in order to focus on developmental mechanisms, we constructed an artificial neural network model of productive vocabulary acquisition, seeking to capture patterns of typical vocabulary development, the qualitative patterns of our cases studies, and the effectiveness of intervention techniques that targeted different components of the model.

Keywords: Intervention, word-finding difficulties, connectionist modelling, phonology, semantics, naming.
**Introduction**

Up to 7% of children have specific language needs and around 25% of children attending language support services have word-finding difficulties (WFD; Dockrell et al., 1998). Difficulty finding words can influence children’s relationships, self-esteem and education. The precise nature of the deficit underlying difficulties with word retrieval is controversial (e.g., Best, 2005; Bishop, 1997; Friedmann et al., 2013; Kail et al., 1984; Messer & Dockrell, 2006). The deficit has been attributed to impairments in the storage of word meaning: for instance, these children may also have problems distinguishing between similar semantic neighbours of a superordinate category, or they may produce impoverished word definitions (Dockrell et al., 2003; McGregor et al., 2002). However, children may experience difficulties in retrieving word forms even when they are fully aware of a word’s meaning. This has led to a proposal that word-finding difficulties may be caused by problems in phonological processing, that is, in the retrieval or assembly of the component sounds of a word (e.g., Constable et al., 1997). The true picture may be more complicated, with multiple types of processing difficulty responsible and different children experiencing different sources for their word finding problem (Best, 2005, Faust, Dimitrovsky & Davidi, 1997). A similar account has, indeed, emerged in the case of adult aphasia (cf. Nickels, 2002).

This complex backdrop presents a challenge to researchers: what methods should they employ to advance their understanding of heterogeneous
causes, and how should they construct links between their findings and the implications for intervention? In this paper, we present an approach employing three convergent methods. The first was to study deficit in the context of development. The second was to draw on connectionist computational modelling to deepen our understanding of atypical mechanisms. The third was to investigate the effectiveness of intervention, in a case series study, as a window onto underlying causes. We show that these strands can interact, by using our computational model to simulate both the behavioural deficits and the effectiveness of interventions on our case studies.

Beginning with our first method, it is now accepted that our understanding of developmental deficits may be improved by setting the impairments in the context of how these skills emerge in typically developing children (Karmiloff-Smith, 1998; Thomas et al., 2009). Here, the target impairment is WFD. We considered a set of component skills that contribute to the development of naming, to explore their variability in typical development, and their relation to naming. In this paper, we will focus on the case series study, but the typically developing control group provides a key context to understand when variability is typical or atypical.

Turning to our second method, progress in understanding the cause of developmental deficits is hampered if such deficits are imprecisely characterised. This is true in the case of WFD, although recent attempts at profiling case series of children’s performance across tasks and linking these to levels of deficit have
met with some success (Best, 2005; Friedmann et al., 2013). In group studies, naming deficits have been variously attributed to ‘a general difficulty accessing semantic information’, to ‘a speed of processing deficit’, and to representations that are ‘impoverished’ or ‘less developed’. One valuable tool for forcing greater precision on theory is computational modelling of the developmental process. Modelling also has the virtues of testing the viability of certain theoretical proposals where they can be realised through implementation, and generating predictions about the model’s behaviour in novel situations. In this paper, we present a formal computational model of lexical development that specifies processing mechanisms for the association between word meaning (semantics) and word form (phonology). As in other developmental domains, we were able to simulate the conditions under which normal development could be deflected (Thomas, 2005a, b; Thomas & Karmiloff-Smith, 2002, 2003), in this case to produce different rates of development of naming skills and different patterns of errors. Although implemented models of developmental deficits are becoming more common, in very few cases are these models taken the extra step – to explore the possible effects of intervention.

Our third method sought to generate these intervention data. There are relatively few well-controlled studies investigating therapy for word-finding problems in children. Studies have focused on comparisons between intervention techniques (Wing, 1990, Hyde Wright et al., 1993, McGregor & Leonard, 1989). The results of such studies are generally positive. Overall, they suggest that
therapy can improve word-finding abilities in children. This is the case for both semantic (Ebbels et al., 2012) and phonological approaches (Bragard et al., 2012). In addition, the improvement may be found in children of a wide age range (e.g., Wing, 1990, 6-7 years; Hyde Wright et al., 1993, 8-14 years); it can generalise to untreated words (Hyde Wright, 1993); and can persist (Bragard et al., 2012; McGregor, 1994). Nevertheless, the studies conflict as to the most effective approach. For example, Hyde Wright et al. (1983) and Wing (1990) contrasted semantic and phonological interventions. In the former study, with 8-14 year olds, the semantic techniques appeared to bring about improvements in word finding whilst the phonological techniques did not. In the latter study with younger children (aged 6-7 years) the reverse was found. One reason for this discrepancy may be that different children, for example of different ages, or with different difficulties, respond best to different interventions (e.g., McGregor and Windsor, 1996).

This mirrors the evidence emerging from adults with anomia as part of acquired aphasia. It has been established that both phonological components analysis (Leonard et al., 2008) and semantic features analysis (Coehlo et al, 2000; Boyle & Coelho, 1995) can improve adults’ naming (Van Hees et al. 2013). However, the relationship between the level of deficit and outcomes of intervention is far from straightforward (Lorenz & Ziegler, 2009).

Likewise, the developing body of research into children’s word learning has produced mixed evidence on the role of semantic versus phonological cues in
influencing children’s ability to acquire and retain new words. Gray (2005) found that a group of 24 children with specific language impairment (aged 4;0 - 5;11 years) comprehended more words in a semantic condition and produced more names accurately when given phonological cues. Meanwhile, the typically developing control group performed similarly in both trials. Zens et al. (2009) identified an order effect in their study of 19 children with specific language impairment (aged 6;3 - 8;2 years). Positive treatment effects for producing new words were found for the children who received phonological awareness intervention, followed by semantic intervention. There was no improvement in the comprehension of new words for either group.

Furthermore, practitioners vary in their approach to intervention for children with WFD; some approaches target the areas with which a child shows most clinical need, while others use relative strengths, aiming to bypass areas of difficulty (Best, 2003). Speech and Language Therapists will tailor their treatment plans depending on many factors including the child, the environment and response to previous therapy. In order to shed further light on this debate, we chose to contrast different therapy approaches by using a tightly controlled case series design in which each child was given two interventions, one developing semantic knowledge of target words and the other phonological aspects. We explore the effects on naming with reference to our model of lexical development. We also use wider outcome measures including collecting the children’s views on intervention and outcome.
In the following sections, we first characterise the naming development of two girls with developmental WFD in relation to a sample of 20 typically developing (TD) children from the 100 TD children tested on our core tasks. We then outline the intervention methods used for the two case studies, and indicate the type of intervention that was most effective. In the second half of the paper, we describe the computational model of lexical development, which had three targets: to capture the qualitative developmental patterns of the TD children, to capture under various conditions of processing deficit the pattern exhibited by the two case studies, and finally to simulate the effectiveness of different types of intervention. We then discuss how our three methods are mutually informative about the underlying causes of WFD.
Empirical study

Method

Criteria for inclusion and nature of the sample.

In this study we investigate naming and related skills of two girls with WFD from a larger ongoing study (Best et al., 2013). The presentation of two cases allows in-depth exploration of their profiles in relation to TD children and of the outcome of intervention. Furthermore, the use of two cases, alongside TD children, provided a manageable set of data for the computational modelling. In due course the findings from the wider intervention will become available, but presentation of the larger study will inevitably mean less detail is available on individual cases.

Twenty typically developing (TD) children were selected from a sample of 100 children participating in the larger study of Best et al. (2013). These 20 children were selected to form an age matched comparison group for the two children with WFD in the present study. The ages of the 20 TD children ranged from 7;01 to 8;00 (mean = 90.75 months, SD = 3.86). They were attending schools in London and the surrounding area, with similar catchment to the schools of the two children with WFD. Background assessments of non-verbal ability and receptive vocabulary were carried out. Nonverbal ability was assessed using the Pattern Construction subtest of the British Ability Scales Second Edition (BAS-II, Elliot, Smith & McCullouch, 1996). Receptive vocabulary was assessed with the British Picture Vocabulary Scale Third Edition (BPVS-III, Dunn, Dunn &
The TD sample achieved a mean standard score of 56.95 (SD = 10.43) for Pattern Construction, and 105.35 (SD = 12.03) for the BPVS.

The two girls with WFD were referred to the study by the SENCOs at their schools. Both girls came from English speaking families. Initial assessment indicated that they met the criteria for inclusion in the study: (a) they demonstrated a discrepancy between comprehension and naming on the Test of Word Finding Second Edition (TWF-2, German, 2000: both girls had a word finding quotient of 60, <1 percentile); (b) on non-verbal assessment they performed above the required level (Pattern Construction subtest from the BAS-II, Elliot, Smith & McCullouch, 1996: Amy 21st percentile, Magda 24th percentile); (c) neither had a diagnosis of dyspraxia, ASD, ADHD or global developmental delay.

**Design.**

The children with WFD were assessed on a set of background tasks in order to obtain a picture of their language and wider cognitive processing and both these children and the TD children were assessed on four core tasks and on a choice reaction time measure. After the background assessments the authors who are Speech and Language Therapists (WB and LH) determined whether or not children were suitable for inclusion in the intervention arm of the study. Children were excluded if they did not exhibit WFD in connected speech or if they were having specialist help at school to the extent that further withdrawal
from the classroom would not be appropriate. The two girls were deemed appropriate for intervention and therefore participated in pre-therapy assessments including naming 100 experimental items on three occasions during one half term. Randomisation took place at this point (see Intervention section). Each child then either waited for therapy or began immediately. They participated in both therapies in a crossover design with a wash out phase between interventions and were followed up to investigate maintenance of any effects. Each phase of the therapy (therapy 1, wash out, therapy 2 and follow-up) lasted for half a term - 6 weeks. The design is illustrated in Figure 1. After each phase of the study the children were reassessed on naming all items. This assessment was carried out by a research associate working at a different institution who remained blind to the allocation of children to conditions.

(Insert Figure 1 about here)

Assessment.

Background tasks.

The two girls with WFD were assessed on the following background language tasks: the Word Discrimination subtest of the Test of Auditory Processing Skills Third Edition (TAPS-3, Martin & Brownell, 2005) to obtain a measure of their ability to discriminate sounds within words; the British Picture Vocabulary Scale Third Edition (Dunn, Dunn & Styles, 1997) for a measure of
receptive vocabulary; four subtests from the Clinical Evaluation of Language Fundamentals Fourth Edition (CELF-4, Semel et al., 2003) for which we provide results for Concepts and Directions to give a measure of language comprehension, and the overall Core Language Score; the Test for Reception of Grammar (TROG, Bishop, 1989) which assesses understanding of different grammatical structures; and the fluency sub-tests of the Phonological Abilities Battery (Frederickson, Frith & Reason, 1997) which require word generation on the basis of semantic category and initial sound.

Four core tasks.

Rationale.

The TD children and children with WFD were assessed on four core tasks which were used to build the model of the TD children’s performance. Full details of these and the reaction time tasks are provided in Appendix A. They include confrontation naming as a direct measure of children’s ability to retrieve and produce words in response to a picture, and word picture verification as a measure of the children’s knowledge of the word (to score correct they need to both accept the target name and on a separate occasion reject the name of a close semantic co-ordinate). The word-picture verification task was devised specifically for the study. We also devised a picture judgement task using a sub-set of the naming items designed as a developmental analogue of the widely used Pyramids and Palm Trees test (Howard & Patterson, 1992). Importantly, no language is used in the task, the children are making judgements based on the
semantic relationship between the pictured items. Reaction times for this task were also collected as there were only 20 items. Finally the Children's Test of Nonword Repetition (Gathercole & Baddeley, 1996) was employed to explore the children’s phonological abilities when lexical processing is not required. Repetition is a sensitive task as both phonological input and output processing need to be adequate for correct production of the forms. The use of the picture judgement task and the non-word repetition task enabled us to consider the children’s ability to deal with meaning in the absence of form and the reverse, i.e. form in the absence of meaning.

Description.

Naming: The materials from the study of Funnell, Hughes and Woodcock (2006) were used. They consist of 72 black and white line drawings of objects. The picture naming task was programmed using the experimental software DMDX (Forster and Forster, 2003) running on a laptop computer with a 15.4 inch screen. Naming responses were recorded using an external microphone connected to the laptop. CheckVocal software (Protopapas, 2007) was used to obtain naming latencies. Accuracy of the naming responses was also recorded. Items were presented in one session divided into three blocks of 24 items each. Children were asked to provide a single word for each picture. The tester moved to the next item as soon as the child named the picture. Naming responses were recorded at the time of testing and checked later from the recording. Errors were classified according to the scheme shown in the Results section.
Word-picture verification task (WPVT): A single word comprehension task was developed using the pictures from the naming task. It involved presenting one picture at a time on the computer together with a pre-recorded spoken word. On one occasion the picture was presented with the matching verbal label and on the other occasion, the picture was presented with a semantically related word. The child was asked to decide whether the spoken word corresponded to the picture. Seventy-two object names were selected that were semantically related to the objects depicted in the Funnell et al. (2006) pictures. The 72 items in each testing session were split into three blocks of 24 items, with a rest between blocks. For each item, a combined accuracy score was calculated such that a score of 1 was given only if the child both accepted the correct name and rejected the semantically related word (as in Brees & Hillis, 2004). Thus, WPVT accuracy scores reflect the number correct out of a maximum of 72.

Picture judgement task (PJs): In this task three pictures depicting objects from the Funnell et al. (2006) and Druks and Masterson (2000) picture sets were presented - a target together with two pictures underneath. One of the two pictures presented in the lower part of the screen had an associative semantic relationship to the target, and the alternative item came from the same semantic category (e.g., tie presented with associate shirt and distractor shorts). The child was asked to choose which of the two items in the lower part of the screen fitted best with the item at the top. Scores consist of number correct (out of a maximum 20) and the median of key press times for correct trials.
Non-word repetition (CNRep): The widely used Children’s Test of Nonword Repetition (Gathercole & Baddeley, 1996) was used to investigate children’s repetition of unfamiliar forms. The test consists of 40 nonwords of increasing length and complexity. We report standard scores and percentage correct.

In addition to the four core tasks, all the children were assessed on simple and choice reaction time tasks adapted from Powell, Staintorp, Stuart, Garwood and Quinlan (2007). In the choice reaction time task reported here as a measure of general processing speed, the child was asked to identify which of two stimuli appeared on the computer screen, and to press the appropriate response key as quickly as possible. The targets were two dinosaur pictures (one green and one orange). Children were asked to press the left Ctrl button as soon as the green dinosaur appeared, or the right Ctrl button if the orange dinosaur appeared. Green and orange stickers were placed on the corresponding buttons. A mouse press initiated the practice block of six items, with half containing the orange and half the green dinosaur. A black fixation cross appeared in the middle of the screen for 500 msecs followed by the target stimuli. Lag times varied in randomised order, as did appearance of either the orange or green dinosaur. The lag times were 300, 600 or 900 msecs. There were two blocks of 18 trials. We recorded the median of each child’s key press times for correct trials.

**Intervention.**
Intervention was carried out with the two children diagnosed with WFD. Within the crossover design each child participated in two treatment conditions, acting as their own control. One intervention focused on semantic processing and the other on phonological processing of the target words. Therapy protocols were devised taking account of (a) techniques used widely with children with WFD (word-webs, ‘take a word for a walk’, e.g. Commtap online resource) which explore semantic and phonological attributes of a word, and (b) approaches used successfully with adults with anomia as part of their aphasia (semantic and phonological feature analysis, e.g. Coehlo et al, 2000; Leonard et al., 2008; Boyle & Coelho, 1995). Therapy took place once a week for approximately 30 minutes, with each intervention block consisting of six sessions.

Four sets of 25 items were matched for baseline picture naming accuracy on the following psycholinguistic variables: age of acquisition, log frequency, imageability, and visual complexity. Twenty five experimental items were treated each half term, along with a further 6-12 non-experimental words, which were selected by the children, teachers and/or carers. Thus each child had different sets of experimental items (within the 100) and of personally chosen items. At the start of each session, prior to therapy, children were asked to try and name pictures representing all of the above items, as well the control items (see below). Four different sequences of presentation were used alternately to control for order effects. Children were invited to press a comedy buzzer to pass on items that they were not able to name. This was to reduce frustration at being
asked to repeatedly name items without feedback – especially naming control words, which were not treated.

The therapy blocks were designed to be as similar as possible, albeit one focusing on semantic attributes of the words and the other on phonological attributes. Template semantic and phonological word webs are provided in Appendix B, which also provides an overview of the therapy protocol. In the first phase of therapy (which typically covered sessions one and two), the therapist introduced the appropriate word web and supported the child to ‘think around the word’ together. The therapist used a series of prompt questions, derived from phonological components or semantic feature analysis, to encourage the child to generate features about an item (for example a category in the semantic therapy, or number of syllables in the phonological therapy). If the child was unable to produce a target feature within 5 seconds, or gave vague or inappropriate information, the therapist provided a ‘forced choice’, e.g. ‘Is it an animal or a vegetable?’, or ‘Does it have 2 or 3 syllables/beats?’. If the child was still unable to produce a feature within 5 seconds, the therapist gave the appropriate spoken information and wrote this on the chart. As sessions progressed the word-webs were used in games, with a barrier placed between therapist and child, designed to encourage communicative use of the target items. Throughout therapy emphasis was placed on meta-linguistic skills, encouraging the child to consider ‘what helps you when you can’t find the
word?’, and in the barrier games, ‘what is the main thing about the word that would help me guess?’.

Therapy items were treated in a continuous, cyclical order. Words named correctly at the start of the session were not targeted on that day. Length of therapy sessions remained constant throughout, regardless of how many items were covered. The therapist completed a record of the number of times a child produced each item within a therapy session, as well as their response to each therapy task. If a child offered spontaneous information, which was not directly targeted in therapy, e.g. gestures, drawing the features or writing the word, this was neither inhibited nor encouraged. All sessions were video-recorded.

Children were allocated at random to: (a) start immediately or wait to start therapy (Amy began immediately, Magda was assigned to delay and therefore performed a second pre-therapy baseline); and (b) begin with either the phonological, or semantic therapy condition (both girls were assigned to the phonological before semantic intervention). The 4 matched sets of experimental items were randomly allocated to 4 conditions: unseen control, naming control, phonological therapy and semantic therapy. Thus the items in the treated sets differed across the two girls, but in each case, sets were matched for the three pre-therapy naming assessments. Randomisation was undertaken at another UK University and project investigators were not involved in the process.

The primary outcome measure for the intervention was confrontation naming of the pictures. We also collected the girls’ views of the intervention by
interview and by their completion of a 5 point pictorial Likert scale with a member of research staff who had not been involved in the intervention. Finally, conversations with the girls were collected on 3 occasions using the guidance in Appendix C: twice prior to the start of intervention (approximately 2 months apart) and once at follow-up, after the girls had been involved in both interventions (approximately 8 months later). The conversations were transcribed and scored using the Profile of Word Errors and Retrieval in Speech (POWERS, Herbert et al., 2013) by team members blind to the date of each conversation. Two conversation variables were calculated for the present study: content words produced per conversation turn and word finding behaviours per content word.

**Results**

**Background assessments.**

The findings from the background testing are shown in Table 1. Both girls performed well on the word discrimination task (TAPS) suggesting adequate processing of speech input. However, Magda showed impaired performance on language comprehension tasks at the single word (BPVS) and sentence level (CELF Concepts and Directions sub-test and TROG). Amy had relatively good language comprehension as demonstrated by her performance on these three tasks. On the PhAB fluency task, Magda performed poorly with relatively worse generation of semantic than alliterative items. Amy performed well on this task,
although she demonstrated the reverse pattern from Magda with better performance on semantic than alliterative fluency.

(Table 1 about here)

**Profiling children with WFD in relation to TD children.**

Table 2 shows the performance of the girls on the 4 core tasks relative to the performance of the TD children. In line with their performance on the test of word finding and their inclusion in the study, both girls were very poor at naming relative to TD children. Magda found this task particularly difficult. Appendix D shows the error classification scheme and errors in each category made by the girls. Both made semantic errors, however, the number of co-ordinate errors made by Amy and Magda was not more than 1.5 SD above the mean of the TD children. It is clear from the table that Magda also differs from the TD children in that she produces mixed errors (words both semantically and phonologically related to the target e.g. *scraper* for *rake*). These are striking because English does not afford many opportunities for such errors. She also produced mixed errors in conversation. These errors indicate both semantic and phonological influence on word finding (Nickels, 1997). Finally, both girls produced phonologically related non-word errors. These are very unusual in the naming attempts of the TD children and tend to be associated with post-lexical phonological production difficulties.
On WPVT, which tested comprehension of the target items, Amy’s accuracy was nearly 1.5 SD below the mean for TD children. Magda, in contrast, performed well below 1.5 SD from the mean score of the TD children. On the picture judgement task (PJs), which does not require lexical processing, both girls scored 16/20 items correct which is 1.5 SD below the mean for the TD group. However, Magda performed particularly slowly on this task despite the fact that her reaction times, like Amy’s, were not in general slow relative to those of the TD children, as demonstrated by their performance on the choice reaction time task.

(Table 2 about here)

In line with this, both girls performed poorly on the non-word repetition task. While this can occur for several reasons, their relatively good performance on the word discrimination subtest of the TAPS strongly suggests the difficulty may stem from retrieving, holding or producing the phonemes, rather than with input processing.

**Conclusions drawn from assessment and clinical impressions.**

The findings from background testing and our core tasks show Amy has relatively good comprehension. Her performance on the tasks involving semantic processing (PJs and WPVT) is around 1.5 SD below the TD mean. In contrast, on tasks requiring phonological output (naming and CNRep) her scores are more
than 3 SD below the TD mean. Thus, her naming problem appears to arise at least in part from difficulties in post-lexical phonological assembly for word-production. Evidence in support of this includes poor repetition of non-words in the context of good auditory discrimination combined with the production of non-word phonological errors in naming.

Magda has word-finding difficulties in the context of language needs spanning comprehension and expression. Her scores on the background tests suggest wider language impairment beyond her WFD. Neither her performance on tasks tapping semantic processing nor that on tasks tapping phonological processing match those of typically developing children. Her profile on these tasks matches well with that on our core tasks. Specifically, she performed very slowly on the PJs task which requires semantic judgements in the absence of linguistic processing, and her accuracy score was more than 2 SD below the TD mean on the WPVT task, where accurate performance requires acceptance of the target name and rejection of a close semantic co-ordinate. Magda also has considerable difficulty with both naming and CNRep, scoring more than 3 SD below the TD mean on both tasks. The pattern across the tasks suggests her word-finding difficulties may have multiple sources, arising from both semantic and phonological output processing problems, perhaps with a particular difficulty in accessing word forms as indicated by the presence of mixed errors (which are rare in the TD sample) and by her pauses before word retrieval in conversation.
It is important to note that while for the purposes of this paper we have focused on the girls' patterns of difficulties in order to model their profiles and intervention outcomes, both girls have considerable communicative strengths. Amy is better able to find words in conversation than in a constrained picture-naming situation and is a very enthusiastic communicator and storyteller. Magda is aware of her language difficulties and communicates very well, for example, by sometimes holding the conversational floor to avoid questions, and saying things in different ways until she gets her message across. She uses gesture well when unable to find words. Despite these strengths, the girls’ everyday communication is influenced by their difficulty in retrieving words, for example they demonstrate word-finding behaviours in connected speech and in conversation.

**Intervention.**

The girls naming over the course of the study is shown in Figure 2.

(Insert Figure 2 about here)

Statistical analysis of single case and case series (SCEDs) is an area of discord and many authors simply employ visual inspection of the data over the course of the study (for a review see Smith, 2012). We consider statistical analysis to be important and employ a ‘belt and braces’ approach, following both Smith-Lock et al. (2012) in using the stringent McNemar non-parametric test
which takes into account items moving from correct to incorrect as well as in the desired direction) and Hickin et al. (2002) in using statistics weighted according to the phase of the study to test specific hypotheses about change.

**McNemars.**

Firstly we tested whether the girls' naming of the items improved with each type of therapy. This was done separately for the treated items (n=25) and for the whole set (n=100), in each case making a comparison between naming just prior to and immediately after each intervention. We used one-tailed tests as we predicted improvement, employing a cut-off of \( p < 0.05 \). Amy showed significant benefit from the phonological therapy but not from the semantic approach (Treated set: Phonological 0.56 -> 0.88, \( p = 0.011 \), sig., Semantic 0.56 -> 0.76, \( p = 0.063 \), trend. Whole set: Phonological 0.54 -> 0.65, \( p = 0.010 \), sig., Semantic 0.65 -> 0.68, \( p = 0.304 \), ns.), although there was a trend to improvement on the set treated with the semantic approach. Magda showed no significant benefit from the phonological intervention but naming of the treated set benefitted significantly from the semantic approach, with a slight trend to improvement on all items (Treated set: Phonological 0.44 -> 0.48, \( p = 0.5 \), ns., Semantic 0.60 -> 0.92, \( p = 0.004 \), sig. Whole set: Phonological; 0.42 -> 0.47, \( p = 0.151 \), ns., Semantic 0.50 -> 0.57, \( p = 0.072 \), trend.).

We also tested for improvement over the course of both interventions together, again using 1-tailed tests, by comparing the final pre-therapy baseline score with naming immediately after the second phase of therapy on all items.
Both girls made significant progress (Amy \(0.54 - 0.68, p = 0.0013\), sig.; Magda \(0.42 - 0.6, p = 0.00204\), sig.). Finally we compared final post-therapy naming performance with follow-up half a term later. In this case we used 2-tailed tests as naming may have continued to improve or dropped off after interventions ended. Neither of the girls showed significant drop-off post-therapy (Amy \(0.68 - 0.62, p = 0.146\), ns.; Magda \(0.57 - 0.60, p = 0.774\), ns.).

**Weighted statistics.**

We weighted the girls’ naming at each assessment to test 4 different hypotheses. The weightings (provided by David Howard in a personal communication) differ for the two girls as Magda was assigned to the ‘wait’ condition prior to starting therapy and thus had 4 pre-therapy baselines whereas Amy had 3 pre-therapy baselines. We used 1-tailed tests throughout; the full weightings are provided in Appendix E.

1. First we looked for an overall trend for improvement over the course of involvement in the study. Both girls’ naming demonstrated this \((n=100, \text{Amy, } t = 4.31, p < 0.001, \text{sig.; Magda, } t = 5.23, p < 0.001, \text{sig.})\), however, this change may simply reflect development and not be due to the interventions.

2. We tested whether there was greater change during the therapy phases of the study than over the other phases (baseline, wash-out and follow-up), for both the whole set and for treated items only. Neither girl showed significantly greater change during intervention on the whole set \((n=100, \text{Amy } t = 1.04, p = 0.151 \text{ ns; Magda } t = 0.76, p = 0.225 \text{ ns})\) while both showed
significantly greater change on the treated items during intervention phases of the study than the remainder (n=50, Amy, $t = 2.56, p = 0.007$ sig.; Magda, $t = 2.05, p = 0.023$ sig.).

3. We also tested whether there was a different effect of the two treatments on all items. There was a significant difference for Amy, but interestingly not for Magda (n=100, Amy $t = 2.40, p = 0.009$, sig.; Magda, $t=0.70, p = 0.244$ ns).

4. Finally we tested whether there was greater change during therapy for the sub-sets of items used in the different interventions. The findings support those from the McNemar tests. For Amy there was greater change during therapy for the set treated with the phonological intervention and this was not the case for the set given semantic therapy (n=25, Phonological $t = 3.29, p =0.002$ sig.; Semantic $t = 0.44, p = 0.333$ ns.). For Magda the reverse was true (n=25, Phonological $t = -0.15, p = 0.559$ ns.; Semantic $t = 3.09, p = 0.003$ sig.)

**Wider outcomes.**

**Children’s views.**

The girls rated which aspects of the research interventions they perceived as most helpful to them using a 5 point pictorial Likert scale with 5 at the positive end of the scale. The results are summarised in Table 3. The girls were asked how helpful it was to think about the meaning in words (i.e., semantic therapy), versus the sounds in words (i.e., phonological therapy).
When asked ‘What helps you most when you are stuck in finding words?’,

Amy cited a strategy worked on during phonological therapy (the most effective of the two interventions for her). She described this as ‘chunking it out’, i.e. breaking down longer words into shorter, more memorable parts. Meanwhile, Magda’s response: ‘I show someone the action’ cites an idea not directly targeted in therapy, but which formed part of the semantic intervention and which she used spontaneously with some success to help get her message across during conversation. Both children rated their enjoyment of the project as 5 (the maximum score). Magda stated that finding words was ‘a little bit easier’ at the end of the study, while Amy spontaneously used numerical ratings to illustrate her perceived progress: ‘At the beginning it was 1 and now it is 3’.

Conversation.

The findings from exploring word retrieval in the girls’ conversations are provided in Table 4. Amy showed a gradual increase in the number of content words (mainly nouns and verbs) that she produced per conversational turn over the three occasions. For Magda the pre-therapy conversations are fairly stable in this regard with the post therapy conversation showing a dramatic change with around twice as many content words produced per turn. Both girls show a reduction in word finding behaviours per content word over the three
conversations. For Amy this occurs gradually, for Magda the drop occurs between the two pre-therapy conversations.

(Table 4 about here)

**Discussion**

We aimed to allocate children with WFD to one or more sub-categories, based on their profile across background assessments and to relate this to the outcome of intervention. We were successful in the first endeavour and the patterns of difficulties across background assessment and our 4 core tasks form a coherent picture.

While previous studies have explored a link between children’s underlying language profiles and their naming, this has largely been in the context of group studies (e.g. Dockrell & Messer, 2007) which do not enable the level of precision possible with case series. Lahey and Edwards (1999) compared two subgroups of children with SLI and found that those with expressive-only language impairments made a higher proportion of phonological errors, relative to TD controls, while those with both expressive and receptive difficulties produced more semantic substitutions. However, the author’s categorisation of participants was an ‘all-or-nothing’ distinction and did not allow for multiple levels of deficit, as we found with our case studies.

Regarding their own views of involvement and change, when the girls’ rated how helpful each intervention was for them, neither chose the approach that most improved their naming. A possible explanation for this is that both
favored the activities they found easiest, rather than those targeting their core difficulty. When asked what helps them retrieve words Amy suggested a strategy that was part of the phonological intervention. We concurred that this was a useful tool for her, based on her response to therapy tasks. Magda suggested gesture, which she used in conversation prior to therapy and which was part of the semantic approach which aided her word-retrieval. It is encouraging that both children rated their enjoyment of the project at the top of the scale, given the relative severity of their word-finding difficulties and the frequent requirement for them to attempt to name hard-to-reach words. Both accurately reflected that they had made some progress with their naming post-therapy, while acknowledging the persistent nature of their difficulty.

There is very little research attempting to relate different profiles of word finding difficulties to the outcome of intervention and the endeavour is far from straightforward. For example, the outcome in Best’s (2005) intervention study using a cueing aid did not differ across the 5 children who took part, meaning it was not possible to meaningfully relate their naming profiles to the outcome of the therapy. Bragard et al. (2012) attempted to relate four individual children’s therapy outcomes to their linguistic profiles. Participants’ WFDs were characterised as either semantically or phonologically-grounded on the basis of poor performance on picture or spoken judgment tasks. Full assessment results are not reported, but two children with semantically-categorised WFDs also presented with severe phonological and/or morpho-syntactic difficulties. Each
responded better to the phonological intervention, rather than the predicted semantic treatment. There are some methodological concerns with this study (e.g. second pre-therapy baseline data is not provided and treatment sets differed in their pre-therapy scores) making findings difficult to interpret.

The headline findings from analysis of Amy and Magda’s naming over time are that there was greater change during therapy than the other phases of the study, demonstrating an effect of intervention over and above development. On treated items Amy, who had particular difficulty with assessment tasks requiring phonological output, benefitted from the intervention highlighting phonological properties of target words but did not benefit from the semantic intervention. In contrast, Magda, with wider language difficulties including with semantic and phonological output processing, benefitted from semantic intervention, but her naming did not benefit significantly from the phonological intervention. The effects of intervention are largely specific to treated items, although Amy did improve on the set as a whole after the phonological intervention. The effects of therapy maintain for at least half a term. One implication for a theory of therapy could be that therapy resources may be best directed at areas of need for children with WFD. This would fit well with Amy benefitting from the phonological intervention and Magda from the semantic approach. However, given that Magda also had difficulty with phonological output tasks, on this account, we would need to explain why she did not benefit from the phonological intervention. In fact, there was no significant difference in the effect of the two
interventions on the whole set of items, hinting at the possibility that the outcomes of the different approaches are not as distinct as they appear for Magda. We return to this and relate it to our findings from the modelling in the general discussion. Certainly, the relatively clear cut contrast in naming outcome for the two girls provides a target for the modelling of the intervention.

**Computational modelling**

**Background**

Connectionist models are influential in theories of naming, particularly that of Dell (Dell et al., 1997; Dell et al., 2013). This model was handwired into its adult state and designed to account for errors in aphasia following damage. It is therefore not well suited to consider developmental mechanisms. A number of computational models have conceptualised lexical acquisition in terms of learning mappings between representations of semantics and phonology. For example, Plunkett et al. (1992) used a connectionist network to associate localist labels with abstract semantic codes and vice versa, focusing on phenomena such as the vocabulary explosion, the comprehension-production asymmetry, and under- and over-extension errors. However, for WFD, one of the issues at stake is whether the semantic and phonological representations have developed normally, and therefore these representations must be a product of development rather than specified by the modeller. The DevLex model of Li et al. (2004), the DevLex II model of Li, Zhao and MacWhinney (2007), and the early word learning model of
Mayor and Plunkett (2010) acquire representations of semantics and phonology in self-organising maps, before learning associations between the map via Hebbian links to capture lexical acquisition. While the DevLex architecture or that of the Mayor and Plunkett model would have provided an avenue to pursue our aims, by their nature, self-organising maps enforce a two-dimensional feature space on both semantics and phonology. We felt that a richer representation of both semantic and phonological space might be necessary to capture subtle developmental anomalies, and therefore opted to encode these types of information over autoassociators where the internal representational space was a free parameter, allowing representations to develop with (in our case) up to 500 dimensions. Similar to the DevLex architectures, our model then learned associations between semantic and phonological codes, which were themselves at various stages of development.

Previously, connectionist networks have been used successfully to model developmental deficits, such as dyslexia (Harm, McCandliss & Seidenberg, 2003), developmental delay in inflectional morphology (Thomas, 2005a), aphasia (Foygel & Dell, 2000) or impaired word reading (Plaut et al, 1996). These models usually introduce developmental deficits as different parameter settings in the impaired model compared to the typically developing model, such as reduced number of hidden units, reduced connectivity between layers, shallow sigmoid activation function (low temperature), or slow learning rate. To our knowledge, there has been only one computational study that has explored the efficacy of
intervention: Harm et al. (2003) used a connectionist model of reading to explore why certain classes of interventions are more effective than others to alleviate reading impairments in developmental dyslexia. Another non-developmental study sought to show how an adult model of aphasia could guide actual interventions depending on patients’ error patterns (Abel et al., 2007). In other work, we have begun to explore the computational foundations of intervening to improve performance in atypical connectionist learning systems (Fedor et al., 2013) but the modelling of intervention remains in its early stages.

Method

Lexicon.

Words were modelled as randomly paired semantic and phonological representations. The semantic representations were fed into the semantic module and the phonological representations were fed into the phonological module. The model employed a simplified domain with a lexicon of 100 words. Semantic representations were constructed around a prototype structure in line with Plunkett et al. (1992). Five prototypes were randomly generated, each consisting of 57 semantic features, 28 active and 29 inactive. Semantic representations for the lexicon were then generated by randomly activating/inactivating units in these prototypes with a probability of 0.05. The result was 5 prototype classes, with 20 semantic representations each, where the
average Euclidean distance between semantic representations was lower within a prototype class (around 17) than between prototype classes (around 30).

Phonological representations were generated using consonant-vowel templates, where each word was nine phonemes long, and each phoneme was encoded using an articulatory feature based code; there were 42 phonemes, 24 consonants and 18 vowels, based on English (Thomas & Karmiloff-Smith, 2003). Similarly sounding phonemes therefore had similar representations, and the Euclidian distance of words that had more phonemes in common was less than that of words that had fewer phonemes in common.

**Architecture.**

An artificial neural network was constructed to model (a) typical development, (b) atypical development, and (c) intervention. The model consisted of two modules, a semantic module, a phonological module, and two associative layers. The architecture is shown in Figure 3. The semantic and phonological modules each had an input layer, an output layer and a hidden layer. They were used to input and output the semantic and phonological representations of words, respectively. They also included recurrent connections from the output layers to the input layers. The recurrent connections were not employed during training, only during testing. They gave the model the facility of settling into its ‘best guess’ output given an input, with the number of cycles required to reach this settled state serving as a simulation of reaction time. The
associative layers served as pathways to connect the hidden layers of the semantic and phonological modules in each direction of activation flow. The size of the semantic input and output layers was 57 units, the size of the phonological input and output layers was 171 units, and the size of the hidden layers (semantic, phonological and both associative layers) was 500 units, for the ‘normal’ model. Adjacent layers were fully connected (connection density was 1).

(Insert Figure 3 around here)

**Training.**

The model was trained using the backpropagation learning algorithm (Rumelhart, Hinton & Williams, 1986) to perform four tasks, simulating the four core tasks that were used to test the children. During training, these four tasks were interleaved and alternated randomly, until the model reached 100% performance on each. The learning rate was 0.1 and momentum of 0.3. Fifty replications (models with different random initial weights) were run with the ‘normal’ parameters and were used as a baseline to represent typical development. The four tasks were as follows:

**Semantic input – semantic output (SS) task:** this task was used to train the semantic module independently of the phonological module. The semantic representation of words was fed into the semantic input (SI) layer, and the same representation was expected on the semantic output (SO) layer. This task
captures the input-output characteristics of the Picture Judgment task, where children look at and point to pictures, without listening to or producing the phonological form of words.

**Phonological input – phonological output (PP) task:** this task was used to train the phonological module in isolation, to develop representations of the phonological forms existing in the model. The performance of this module was used to simulate the children’s performance on the non-word repetition task.

**Semantic input – phonological output (SP) task:** To simulate naming, the model was given a semantic representation on the SI layer and required to output the appropriate phonological form on the phonological output (PO) layer. During testing, activation flowed from SI to semantic hidden (SH) to the associative layer (AR) to the phonological hidden (PH) layer and then phonological output (PO). However, during training of this task, the semantic and phonological modules were held constant and only the weights between SH and PH layers were trained. The activation of the PH layer was checked against the activation of the same layer when the input originated from the PI layer, to derive error signals for weight change. The objective was to elicit the same hidden representations irrespective of the origin of the input (semantic or phonological). This task represents the Picture Naming task, where the input is a picture, and the output is the phonological form of the verbal label for that picture.
**Phonological input – semantic output (PS) task:** As the mirrored pair of the SP task, activation propagation goes through the PI – PH – AL – SH - SO layers (see Figure 2) to test performance, with only weights between the PH and SH layers adjusted during training. This task models the word-picture verification task, where children match a spoken word against a picture but are not required to produce the word themselves.

A training epoch consisted of training the whole lexicon with one of the tasks. In the non-intervention cases the four tasks were trained with equal probability, thus in a round of 100 epochs, each task was trained in 25 epochs on average. Development of normal models was followed across 4000 epochs of training, by which point performance was at ceiling. The ‘age’ of the model was defined as the number of epochs divided by four. During testing, the outputs of the model were considered as 1 (active) if activation was higher than 0.9 and 0 (inactive) if activation was lower than 0.1. A response was scored as correct if all units were in the required state.

**Simulating developmental deficits.**

The typical model was compromised in three different ways to induce computational deficits. These disturbances included: (1) decreasing the number of hidden units in various layers, (2) decreasing the number of connections between layers, or (3) using a shallow sigmoid unit activation function for the artificial neurons in various components of the model. Deficits were applied prior
to the onset of training (Thomas & Karmiloff-Smith, 2002), and could be applied across the whole architecture or to specific parts. We examined the effect of these deficits on the developmental trajectories of the model to establish the deficits that best simulated Amy’s and Magda’s performance on the four core tasks.

**Simulating intervention.**

Intervention was simulated as increased training on one of the tasks, in addition to the four-cycle training that represented experience-driven development in everyday situations. Semantic therapy was modelled by increasing training on the SS task (twice as much as usual), and phonological therapy was modelled by increasing training on the PP task (also twice as much as usual), while continuing training on all the other tasks to model normal learning. Intervention started after 500 epochs of training and continued until the model reached 100% performance on each task or until the model reached the age of 1000. The age of these models were calculated according to their non-therapy training epochs; thus, models with intervention received more training on one of the tasks compared to models of the same age without intervention.

The intervention experimental design was not modelled in detail. For example, the model did not experience the two interventions sequentially. This was not deemed appropriate given this is the first attempt at modelling both
deficits and intervention in the field. Instead, both types of intervention were independently tested on a given atypical network and then compared to the original. We measured how much intervention speeded up development: we subtracted the age at which the model reached 90% performance with intervention from the age at which the model reached 90% performance without intervention. Positive scores on this metric represent more effective interventions in speeding up the development of naming.

**Results**

**Typical development.**

Models with typical parameter settings (TD models) usually learnt all four tasks within 3000 training epochs. Figure 4 shows median values averaged over 50 networks. Since the four tasks differed in relative difficulty, the model's rate of acquisition of the four tasks could not be a target of simulation. In addition, the simulated trajectories depict the whole learning process, whereas those of children show only an intermediate portion of this. Because of the way in which the model was trained, associative tasks always depended on the simpler tasks, i.e., performance on the PS task could never be higher than performance on the SS task, and similarly, performance on the SP task could never be higher than performance on the PP task. Our simplified semantic prototype structure made it harder for the network to learn semantic representations (SS task) than the phonological representations (PP task). Development of the semantic
representations therefore limited development on the comprehension task. Therefore, the SS and PS trajectories generally overlapped. On the naming task, where networks produced errors prior to developing ceiling performance, errors were mostly semantic, that is, the name of another item in the same semantic category.

(Insert Figure 4 around here)

**Atypical development.**

Our goal was to model the qualitative difference between Amy’s and Magda’s performance on the four core tasks compared to the range of variation exhibited by the TD children. Both girls were quite close to the TD range on the SS and PS tasks, and much poorer on the PP and SP tasks. Based on pilot work, we narrowed our search to deficits induced by low number of hidden units (H), shallow sigmoid (low temperature, T) and low connection density (C) in the start states of the networks. First, we applied these computational constraints one-by-one in the semantic module (S), the phonological module (P) or the associative layers (A) of the model. Figure 5 compares the performance of TD models and atypical models after 500 epochs of training for the three types of deficit, respectively.

(Insert Figure 5 around here)
None of these parsimonious, single location deficits captured the behavioural patterns produced by Amy and Magda. Deficits in the semantic module usually produced lower performance on the SS task and did not influence the PP task. Conversely, deficits in the phonological module resulted in lower performance in the PP task and did not influence the SS task. Both girls performed more poorly than TD children on both SS and PP tasks, implying that, in terms of the model, they had deficits at multiple locations.

The effect of single semantic or phonological module deficits on the SP and PS tasks, which involved both modules, varied according to the location and the type of the deficit. One limitation of the model was that due to the prototype structure of the semantic representations of the lexicon, the semantic hidden representations were harder to learn than the phonological hidden representations. This resulted in asymmetric performance on those tasks that involved both modules.

Accepting that multiple deficits might be necessary to capture the profiles of the case studies, we next evaluated combinations of deficits. Both girls performed closer to TD in semantic tasks (SS and PS) than in phonological output tasks (PP and SP). This suggests that their deficits were more serious in the phonological module than in the semantic module. Keeping this in mind, we experimented with deficits of different strength in the two modules and identified three deficits that captured Amy’s profile, shown in Figure 6. The modified
parameters for these models can be found in Table 5; the rest of the parameters were set to the same values as in the TD models. Figure 6 represents our fit to Amy’s deficit.

(Insert Figure 6 around here)

To model the difference between Amy and Magda, we induced a further deficit: Magda was worse than Amy in the associative tasks (SP and PS), so we hypothesised that she could have limitations in the links between the semantic and phonological modules, in addition to deficits in the semantic and phonological modules themselves, corresponding to a widespread deficit. We considered three methods of inducing the further deficit. In deficit C, the connection density of the associative layers was reduced to 0.1; in deficit H, the size of the associative layers was reduced to 30 (AR) and 20 (AL) units; and in deficit T, the temperature of the associative layers was reduced to 0.5 (AR) and 0.4 (AL). The performance of these models after 500 training epochs is shown in Figure 7. The performance of these models is the same as the performance of double location deficit models on the SS and PP tasks, but now lower on the SP and PS tasks. Figure 7 represents our fit to Magda’s deficit.

(Insert Figure 7 around here)
Intervention.

Each of our deficit-fit models was trained in three different conditions: no intervention, semantic intervention and phonological intervention. Intervention was modelled by increasing the amount of SS or PP training to be twice as much as in the no-intervention condition. The effectiveness of interventions was measured by comparing the speed of development in the intervention conditions to the no-intervention condition: the age of the model was measured when the naming task reached 90% accuracy. Positive values represent a more effective intervention. We derived this measure for semantic and phonological therapies and then used t-tests with Bonferroni correction to evaluate whether the difference was significantly higher than zero: a significant result indicates that intervention was effective in speeding up development of naming.

In case of the double deficit models (simulated Amy), we observed that phonological therapy significantly speeded up development whichever deficit (connectivity, hidden units, temperature) was applied. However, semantic intervention did not. The result is shown in Figure 8. This reproduces the empirical finding that phonological intervention, but not semantic intervention, was effective for Amy. The patterns for deficit C and T more closely mirror Amy’s benefit from intervention in that the semantic intervention produced some numerical benefit, albeit not statistically significant (see Figure 3).

Interventions had more diverse results in the case of the triple location deficit models (simulated Magda), where the response depended to some extent
on the nature of the computational deficit. In case of deficit C, both intervention types were successful. In the case of deficit H, neither of the interventions was successful. In the case of deficit T, only the phonological intervention speeded up development significantly. The data are shown in Figure 9. Therefore, we could not reproduce the empirical finding that semantic, but not phonological intervention was successful for Magda. The closest the model came was a prediction of a deficit pattern (of restricted connectivity between phonology and semantics, as well as deficits within these modules) where both semantic and phonological interventions would be effective.

(Insert Figures 8 and 9 around here)

**Discussion**

In the computational modelling that we carried out, the model of lexical acquisition was much simplified, because it was being asked to attend to so many goals: to capture typical development at the group level, atypical development at the individual level, and differential results of intervention at the individual level. It is therefore much beyond what has been attempted with this type of computational model before, and where modelling is progressed in this way, the first such models must be inevitably simplified, with resulting shortcomings. The next step is clearly to employ a more realistic corpus and to target quantitative rather than qualitative fits to the data. Nevertheless, we have
shown a method that was able, for one of the two case studies at least, to both capture conditions of atypical development and response to intervention (with the results on the second more mixed). This approach provides a powerful way to consider atypical development as an alteration of the computational constraints of typical development, and to validate the nature of these alterations by modelling response to interventions. The ultimate value of modelling is that, once you have a simulation of an atypical system, a much wider range of possible interventions can be evaluated in silico before suggestions can be offered to research clinicians for trialling with children. Interventions may be tailored to match the characteristics of the deficits. This outcome may lie some way down the road, but we believe the modelling we have presented offers promising first steps.

**General discussion**

Our project aims for a circle of development in which theory informs intervention and the outcomes of assessment and intervention inform theory. The theories in question may cover multiple domains, including how lexical retrieval develops in typical children, the nature of deficits in children with WFD and how change occurs with intervention both in the children and in our model of development. Here we focus particularly on how the findings from intervention with two children with WFD inform our understanding of development of lexical
retrieval, in this case instantiated in the parameters of our model. Given the limited research in this field our aims are ambitious and our conclusions necessarily cautious. The general discussion highlights strengths of our study and outlines areas for future development.

**Assessment**

We assessed 100 TD children on the same four core tasks, data from 20 children closest in age to the two girls with WFD are used here. This approach means that the relationship between performance on the different tasks can be compared with confidence, as the same TD children were tested rather than different samples for each task. There is considerable variability within the TD sample and this is important as it reflects the children’s different points of their developmental trajectories. The inclusion of our two bespoke tasks, PJs (picture judgement) and WPVT(word picture verification), provided new and appropriate measures of aspects of processing relevant to lexical retrieval that have not been measured sensitively in the past. They are particularly useful as WFDs frequently occur in the presence of wider language difficulties including expressive language. In contrast with other tasks, such as providing definitions for items the children are unable to name, neither task requires spoken output, meaning the tasks can be employed to find areas of relative strength (such as Amy’s typical response times for PJs). They enable us to understand that WFD may have multiple causes within a single child and this understanding was supported by
the findings from our model which was unable to match the girls’ performance by disrupting just one module. The match between the behavioural findings and the model’s performance is not perfect but many aspects of the girls’ performance are captured by multiple deficits in the model.

**Intervention**

We carried out a tightly experimentally controlled intervention study which demonstrated an effect of therapy over development on Amy and Magda’s word finding as measured by picture naming. The intervention was based on techniques commonly used with children and with adults with anomia as part of their aphasia (Boyle & Coelho 1995, Leonard et al, 2008). There are several clinical resources which employ related approaches with children (Comtap, word whizzer), but this is the first study to our knowledge to test the use of word-webs, focusing separately on semantic and phonological features, experimentally with children. The modelling also employed semantic and phonological interventions and obtained differing outcomes.

We were able to model Amy’s profile and the outcome of therapy successfully. The modelling demonstrated that deficits of different strengths in the semantic and phonological modules reflected her patterns of performance relative to the TD children. With this double deficit, whichever of the three deficits we applied, the phonological intervention was effective and the semantic intervention was not. The best fit may be altering C (connectivity) and T
(temperature), as in these cases there was some numerical (non-significant) benefit from the semantic intervention, as was the case behaviourally.

While our modelling of Magda’s deficit as affecting the semantic and phonological modules as well as the links between these fit her behavioural profile, the modelling of intervention for this triple deficit was somewhat less successful. The closest fit was deficit C (connections). In this case both the semantic and phonological interventions should be effective. Behaviourally, the success of the semantic approach matches the modelling. While Magda did not benefit from the phonological approach, there was no significant difference in effect of the two treatments on all items. While this particular point can be seen as a ‘straw-clutching’ attempt at post-hoc matching of our modelling and behavioural outcomes, we hope the endeavour is understandable in the context of an approach in its infancy.

Finally, it is interesting to note that throughout modelling of intervention the phonological approach produces numerically superior outcomes. While this may simply reflect the model, particularly the semantic hidden representations being harder to learn than the phonological hidden representations, it may also be tested against behavioural findings in future cases.

Our findings from the behavioural study could be taken to imply that a focus on areas of need rather than strength maybe more useful in clinical practice. However, this does not fit with the finding that Magda did not show significant benefit from the phonological intervention, despite her difficulties at this level. It may be that had the interventions been in the reverse order for her,
she would have been able to benefit from both. The word learning research of Zens et al (2009) suggests that there can be order effects, but in that case it was phonological then semantic intervention (the order Magda experienced) that was effective. An argument could be made for the reverse, especially given the multiple levels of difficulty influencing Magda’s word retrieval, perhaps she would have benefitted from the phonological approach had it followed an improvement in her semantic abilities. One might also predict that an approach targeting both semantics and phonology simultaneously would be effective. We are very cautious about making firm clinical recommendations on the basis of the findings from the two girls, and will be able to incorporate findings from other cases in the wider study in due course. In the meantime clinicians will continue to be eclectic in their choice of approaches and take into account the progress of individual children in response to different techniques. What is clear from the study, and fits with related research, is that both semantic (Ebbels et al., 2012) and phonological (Wing, 1990) approaches can be used successfully in helping children with WFD.

One key aspect of the approach is to encourage the children to use strategies that they can employ when learning new words in the future. The demands for new word learning and retrieval are considerable with children in the UK. Once a child enters school, s/he is exposed to about 10,000+ new words each year and adds approximately 3,000 of these words per year to their productive vocabulary (Nagy & Anderson, 1984; Nagy & Herman, 1987). While the effects were clear for treated items, the evidence for wider generalisation to
untreated items and to conversation remain unclear. Interestingly both girls found word finding easier after the study, but, in line with the persisting nature of language difficulties, were aware of strategies that helped them. Furthermore there were changes in conversation, particularly the increase in Magda’s use of content words after therapy that are in line with wider changes but which can not be attributed unequivocally to the interventions rather than development. While the experimentalist might suggest conversations repeated at each point naming is reassessed, this would demand resources beyond the scope of the project and furthermore would be likely to result in the girls being very bored by these conversation topics!

Strengths of the intervention study include the guidance of experienced therapists in intervention selection and development, random assignment of the order of therapy and of sets to conditions, assessment by a researcher blind to assignment, pre-therapy baseline matching specific to each child’s naming, relatively large item sets and the inclusion of naming controls and personally chosen items to supplement the experimental set.

The cross-over design comparing semantic and phonological intervention does not allow exploration of the effects of order of intervention. The outcome of a combined semantic and phonological intervention is a priority for future research. This is an issue for the behavioural study and also for our modelling which did not, at this stage, incorporate either combined or sequential interventions.
There are several aspects of the intervention which warrant exploration in the future including which are the ‘active ingredients’ of the multi-component intervention, the effect of therapist feedback of different types, duration of intervention, and intensity. While these and other aspects are of interest the current study affected change without demanding too much of the girls as could be the case with a more intense intervention of longer duration.

In respect of the modelling, it is important to question whether there truly are phases of phonological or semantic development that are independent of each other (as the model assumes), such that deficits in non-word repetition and picture discrimination tasks must necessarily imply multiple locations of deficits.

Using the model to fit case studies of deficits is only interesting in as much as it is possible. In the case of our modelling we don’t simply fit the model to the data, because there is always a developmental process between the deficit and the atypical behavioural profile. Nevertheless, models with a lot of free parameters give themselves a lot of scope to data fit. The most powerful validation of our model, and therefore implications for theories of lexical development, is whether we get the same response to intervention as observed in the case studies. The answer from this new approach spanning very different domains of experimentation is a tentative, yes.
Acknowledgements

This research was supported by ESRC grant RES-062-23-2721. We are greatly indebted to the children who took part so willingly in the study and to their parents, teachers and schools and to Speech & Language Therapist Vivien Gibson who carried out some of the therapy. We acknowledge the important contributions of our Wider Advisory Group (including an Educational Psychologist and a young person with speech and language difficulties) on the design of the behavioural study, and of our Clinical Advisory Group (including Dr Susan Ebbels, Kathleen Cavin & Sarah Simpson) for shaping the therapies used and for more general advice. Dr Mike Coleman constructed the RT version of the Picture Judgement Task. The randomisation and the weighted statistics were kindly provided by Professor David Howard, Newcastle University. UCL MSc student Elisabeth Salt contributed to the analysis of the conversation data.
Appendix A: Full details of experimental tasks

**Picture naming:** The materials from the study of Funnell, Hughes and Woodcock (2006) were used. They consist of 72 black and white line drawings of objects from four categories, with 18 items in each category. Two categories (animals and fruits/vegetables) represent living things and two (implements and vehicles) represent artefacts. The picture naming task was programmed using the experimental software DMDX (Forster and Forster, 2003) running on a laptop computer with a 15.4 inch screen. Naming responses were recorded using an external microphone connected to the laptop.

Items were presented in one session divided into three blocks of 24 items each. The child was asked to provide a single word for each picture. The tester moved to the next item as soon as the child named the picture. Naming responses were recorded at the time of testing and checked later from the recording. Four fixed randomized orders were rotated across children during testing. No more than two objects from the same category appeared in succession, as in the naming study of Funnell et al. (2006). Each trial began with the presentation of a fixation cross in the centre of the screen for 500 msecs. Then the picture appeared and stayed on the screen for a maximum of 5000 msecs in the case of the TD children and 10000 msecs in the case of children with word finding difficulty. Three items, not used in the main testing session, were presented for practice. Feedback on accuracy was given during the practice trials but not during the main task.
**Word-Picture Verification Task:** A word-picture verification task was developed using the pictures from the naming task. It involved presenting one picture at a time on the computer together with a pre-recorded spoken word. On one occasion the picture was presented with the matching word and on another the picture was presented with a semantically related word. The child was asked to decide if the spoken word corresponded to the picture or not. Seventy-two object names were selected that were semantically related to the objects depicted in the Funnell et al. pictures. Related object names that were phonologically close to the target picture name or that started with the same phoneme (e.g., picture of a butterfly, semantically related word “bee”) were replaced.

The names of the pictures and the semantically related words were recorded by an adult English speaker. The audio files were edited to add 500 msecs of silence at the beginning and end of each file. The task was run on a laptop computer with a 15.4 inch screen and was programmed using the experimental software DMDX (Forster & Forster, 2003). There were two testing sessions with individual target pictures appearing once in each session. In one session the picture appeared with its name, and in the other with the semantically related word. The 72 items in each testing session were split in three blocks of 24 items each with a rest pause between blocks. The child was asked to press the left Ctrl button on the keyboard for NO responses and the right Ctrl button for YES responses. Buttons were highlighted with stickers.
Responses were scored correct only if the child accepted the correct name and rejected the semantically related word. Three practice trials were presented with stimuli that were not included in the main testing session. Feedback was given after practice trials but not during the main session. Four fixed random orders of stimuli were rotated across participants. Each trial began with the presentation of a fixation cross in the centre of the screen for 500 msec. The picture preceded the audio file by 16.62 msecs.

**Picture judgement task of associative semantics (PJs):** In this task three pictures depicting objects were presented - a target together with two pictures underneath. One of the two pictures presented in the lower part of the screen had an associative semantic relationship to the target, the second came from the same semantic category as the first. Sixty-nine pictures depicting items from the Funnell et al. (2006) and Druks and Masterson (2000) picture sets were selected from the Shutterstock website. The task was administered using a laptop computer with screen size 15.4 inches and it was programmed using Visual Basic software. There were three practice trials using items that did not appear in the main session and twenty trials in the main task. A fixation point appeared at the start of each trial. The child was asked to choose which of the two items in the lower part of the screen fitted best with the item at the top. If it was the item on the left the child was asked to press the Z button, for the item on the right, the M button. The two buttons were designated with stickers. Feedback on accuracy was given during the practice trials but not in the main task.
Simple and Choice Reaction Time: Computerized tasks of simple and choice reaction time were adapted from Powell, Staintorp, Stuart, Garwood and Quinlan (2007) and programmed on a laptop computer with a 15.4 inch screen using the DMDX software (Forster and Forster, 2003). The simple reaction time task measured the time taken to make a key press response following the appearance of a target on the screen. Six different coloured drawings of monster characters were the target stimuli. The six pictures and instructions appeared on the welcome screen. The instructions were read aloud to ensure that the child understood the task. There were six items for practice followed by two blocks of 18 trials each. Each trial started with the presentation of a fixation point (a black cross) in the centre of a white screen, followed by a lag and then the appearance of the target stimulus. The duration of the lag varied, to discourage anticipatory responses, and was either, 300, 600 or 900 msecs. The lag times were randomised across trials and presentation of the six target stimuli was also randomised across trials. The target stimuli remained on the screen for 1500 msecs.

In the choice reaction time task the child was asked to decide which of two stimuli appeared on the computer screen, and to press the appropriate response key as quickly as possible. The targets were two dinosaur pictures (one green and one orange) from the Shutterstock pictures. Children were asked to press the left Ctrl button as soon as the green dinosaur appeared, or the right Ctrl button if the orange dinosaur appeared. Green and orange stickers were placed
on the two buttons. As for the simple reaction time task, instructions that appeared on the welcome screen were read aloud by the tester. A mouse press initiated the practice block of six items, with half containing the orange and half the green dinosaur. A black fixation cross appeared in the middle of the white screen for 500 msecs followed by the target stimuli. Lag times varied in randomised order, as did appearance of either the orange or green dinosaur. The lag times were 300, 600 or 900 msecs. The target stimulus remained on the screen for 1500 msecs. There were two blocks of 18 trials each in the main test session.

- Sessions occur once a week for 45 minutes (approximately 10 minutes assessment, 5 minutes activity whiles therapist selected unamed items and 30 minutes intervention). Six sessions per intervention block.
- 25 items from the experimental set treated each half term, plus a further 6 – 10 words selected by the children, carers and teachers.
- Therapy items treated in a continuous, cyclical order. Words correctly named at the start of a session will not be treated on that day.
- Record sheets used, including tick charts for monitoring participants’ production attempts and overall response to therapy.

Therapy, first 2/3 sessions:
- Task introduction
- Generation of features, using prompt questions linked to word webs:

  ![Word Web Diagram]

  **Example prompt questions:**
  - What sound does it start with?
  - What other words start with the same sound?
  - Can you break the word into any smaller words or sounds that will help you remember the

- If unable to generate features a choice of features is provided (e.g. it has 2 or 3 beats (Phonological - syllables), it has stripes or spots (Semantic – appearance).
- All features are considered in the same order, starting with the hexagon at the top right and proceeding clockwise. Once all features have been generated or chosen they are reviewed ar the child is asked to say the word. If unable, it is provided by the therapist and the child is encouraged to say the word.

As sessions continue and according to child’s ability:
- Develop metacognitive awareness: encourage child to reflect on what aspects of word webs are most helpful to them.

Therapy, sessions 4/5/6:

![Word Web Diagram]
• Barrier games: Position a screen between the therapist and child. Using completed word webs, take turns to describe and guess items covered in previous sessions.
• Review of most useful strategies learnt during therapy, create card to help child remember what helps them when they cannot retrieve a word.
Guidelines for conversation.

**Conversation measure:**
Start by saying ‘let’s talk a little’...
*Begin with a topic that is personal to the child, based on their own interests/experience – as reported by the child themselves and/or their parent/teacher.*

**Standard short set of questions for all children:**
Can you tell me:
- about your bedroom?
- What TV programmes do you like to watch?
  - Follow-up probes:
    - “Tell me about that one, I haven't seen it.”
    - “What happened on the last one you watched?”
    - “Do you ever watch (insert current programs likely to be of interest)?”
- what you are good at?
- what you would like to be better at?

**Hierarchy of cues to help support/scaffold conversation:**
Mmn
Uhuh
Tell me more.
Just do your best / you’re doing great
I’d like to hear more about that/ Tell me what you can.
That sounds interesting
What else?

**Non-verbal prompts:**
Smiles and eye contact.
Nods of affirmation and agreement.

Appendix D Error categorisation and girls’ errors
<table>
<thead>
<tr>
<th>Error type</th>
<th>Error subtype</th>
<th>Description</th>
<th>Amy's errors</th>
<th>Magda's errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>semantic</td>
<td>coordinate</td>
<td>within same semantic category</td>
<td>tapir named cow</td>
<td>coconut named lettuce</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>coconut named pineapple</td>
<td>donkey named horse</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>jet ski named speedboat</td>
<td>vulture named duck</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>tomato named apple</td>
<td>parachute named air balloon</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>lemon named pear</td>
<td>ladle named spoon</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>carrot named pepper</td>
<td>sledge named boat</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>cheetah named leopard</td>
<td>pelican named duck</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>milk float named bus</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>torch named light</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>donkey named horse</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>garlic named plum</td>
<td></td>
</tr>
<tr>
<td></td>
<td>superordinate</td>
<td>semantic category to which target belongs</td>
<td>ostrich named bird</td>
<td>barge named boat</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>barge named boat</td>
<td>windsurf named boat</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>submarine named boat</td>
<td>tandem named bike</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>windsurf named boat</td>
<td>yacht named boat</td>
</tr>
<tr>
<td></td>
<td>functional</td>
<td>functional attributes/use of target</td>
<td>trowel named digger</td>
<td></td>
</tr>
<tr>
<td></td>
<td>circumlocution</td>
<td>multiword descriptive response</td>
<td>grater named something that you grate cheese on</td>
<td></td>
</tr>
<tr>
<td></td>
<td>visual attributes</td>
<td>similar physical features</td>
<td>ladle named big spoon</td>
<td>scorpion named crab</td>
</tr>
<tr>
<td>phonological</td>
<td>nonwords</td>
<td>nonword that shares at least 50% phonemes with target</td>
<td>squirrel named grIrɘl</td>
<td>caravan named carara</td>
</tr>
<tr>
<td></td>
<td>formal</td>
<td>real word that shares at least 50% phonemes with target, but not semantically related</td>
<td>binoculars named mInQku:l3:</td>
<td>aeroplane named eslesaiPnIn</td>
</tr>
<tr>
<td>mixed</td>
<td>semantic and phonological</td>
<td>semantically and phonologically-related</td>
<td></td>
<td>tractor named truck</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>saw named sword</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>motorbike named bike</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>tank named truck</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>rake named scrape</td>
<td></td>
</tr>
</tbody>
</table>
Appendix E Weightings for statistical comparisons between phases of the study

1. Weightings of naming assessments for Amy

<table>
<thead>
<tr>
<th>Naming Assessment</th>
<th>Pre 1</th>
<th>Pre 2</th>
<th>Pre 3</th>
<th>Post Phon</th>
<th>Post wash-out</th>
<th>Post Sem</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trend</td>
<td>-6</td>
<td>-4</td>
<td>-2</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Treatment vs no treatment</td>
<td>9</td>
<td>-2</td>
<td>-13</td>
<td>4</td>
<td>-7</td>
<td>10</td>
<td>-1</td>
</tr>
<tr>
<td>Treatment A vs treatment B</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>-13</td>
<td>-11</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Phon trt</td>
<td>2.24</td>
<td>-4.47</td>
<td>-11.18</td>
<td>13.42</td>
<td>6.71</td>
<td>0</td>
<td>-6.71</td>
</tr>
<tr>
<td>Sem trt</td>
<td>7</td>
<td>2</td>
<td>-3</td>
<td>-8</td>
<td>-13</td>
<td>10</td>
<td>5</td>
</tr>
</tbody>
</table>

2. Weightings of naming assessments for Magda

<table>
<thead>
<tr>
<th>Naming Assessment</th>
<th>Pre 1</th>
<th>Pre 2</th>
<th>Pre 3</th>
<th>Pre 4</th>
<th>Post Phon</th>
<th>Post wash-out</th>
<th>Post Sem</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trend</td>
<td>-7</td>
<td>-5</td>
<td>-3</td>
<td>-1</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Treatment vs no treatment</td>
<td>5</td>
<td>1</td>
<td>-3</td>
<td>-7</td>
<td>1</td>
<td>-3</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Treatment A vs treatment B</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>-13</td>
<td>-11</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Phon trt</td>
<td>7</td>
<td>-1</td>
<td>-9</td>
<td>-17</td>
<td>17</td>
<td>9</td>
<td>1</td>
<td>-7</td>
</tr>
</tbody>
</table>
References


disorders from a neurocomputational perspective. Manuscript in preparation.


Smith Lock et al. (in press? IJLC) Effective Intervention for Expressive Grammar in Children with Specific Language Impairment


## Tables

**Table 1**: Background assessments.

<table>
<thead>
<tr>
<th>Test</th>
<th>Amy</th>
<th>Magda</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test of Auditory Processing Skills Third Edition: Word Discrimination scaled score (percentile)</td>
<td>10 (50)</td>
<td>9 (37)</td>
</tr>
<tr>
<td>British Picture Vocabulary Scale Third Edition: standard score (percentile)</td>
<td>80 (9)</td>
<td>71 (3)</td>
</tr>
<tr>
<td>Clinical Evaluation of Language Fundamentals Fourth Edition: Concepts &amp; Directions scaled score</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>Clinical Evaluation of Language Fundamentals Fourth Edition: Core Language standard score (percentile)</td>
<td>81 (10)</td>
<td>60 (0.4)</td>
</tr>
<tr>
<td>Test for Reception Of Grammar: percentile</td>
<td>25-50</td>
<td>10-25</td>
</tr>
<tr>
<td>Phonological Abilities Battery: Fluency Test Alliteration standard score (percentile)</td>
<td>95 (37)</td>
<td>87 (20)</td>
</tr>
<tr>
<td>Phonological Abilities Battery: Fluency Test Semantic standard score (percentile)</td>
<td>111 (77)</td>
<td>77 (6)</td>
</tr>
</tbody>
</table>
Table 2: Performance of Amy and Magda on 4 core tasks and a measure of general processing speed (Choice RT), relative to 20 age-matched typically-developing children. Cells are highlighted where case studies differ by more than 1.5 SD from the TD mean. Values in the final column show 1.5 SD below the TD mean for accuracy and standard scores, and 1.5 SD above the TD mean for reaction times and errors.

<table>
<thead>
<tr>
<th>Task</th>
<th>Amy</th>
<th>Magda</th>
<th>TD Mean</th>
<th>TD SD</th>
<th>1.5 SD from TD mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naming (accuracy/72)</td>
<td>21</td>
<td>14</td>
<td>40.40</td>
<td>6.31</td>
<td>30.93</td>
</tr>
<tr>
<td>WPVT (accuracy/72)</td>
<td>48</td>
<td>42</td>
<td>55.15</td>
<td>5.35</td>
<td>47.12</td>
</tr>
<tr>
<td>PJs (accuracy/20)</td>
<td>16</td>
<td>16</td>
<td>18.65</td>
<td>1.50</td>
<td>16.41</td>
</tr>
<tr>
<td>PJs (RT, msecs)</td>
<td>2855</td>
<td>4290</td>
<td>2886</td>
<td>575</td>
<td>3748</td>
</tr>
<tr>
<td>CN Rep (standard score)</td>
<td>51</td>
<td>52</td>
<td>93.68*</td>
<td>13.40</td>
<td>73.58</td>
</tr>
<tr>
<td>CN Rep (percent correct)</td>
<td>22.5</td>
<td>25</td>
<td>66.84*</td>
<td>13.04</td>
<td>47.28</td>
</tr>
<tr>
<td>Semantic (co-ordinate)</td>
<td>11</td>
<td>7</td>
<td>8.50</td>
<td>3.73</td>
<td>14.10</td>
</tr>
<tr>
<td>Mixed (semantic and phonological)</td>
<td>0</td>
<td>5</td>
<td>0.25</td>
<td>0.55</td>
<td>1.08</td>
</tr>
<tr>
<td>Formal (phon. real word)</td>
<td>0</td>
<td>0</td>
<td>0.10</td>
<td>0.31</td>
<td>0.56</td>
</tr>
<tr>
<td>Phonological (non-word)</td>
<td>2</td>
<td>2</td>
<td>0.15</td>
<td>0.37</td>
<td>0.70</td>
</tr>
</tbody>
</table>

**Choice RT task (msecs)**

<table>
<thead>
<tr>
<th>Amy</th>
<th>Magda</th>
<th>TD Mean</th>
<th>TD SD</th>
<th>1.5 SD from TD mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>589</td>
<td>525</td>
<td>588</td>
<td>140</td>
<td>797</td>
</tr>
</tbody>
</table>

*TD data for the CNRep task are for 19 children rather than 20 as for the other tasks.*
Table 3: Children’s views of the intervention and outcome for them.

<table>
<thead>
<tr>
<th></th>
<th>Amy</th>
<th>Magda</th>
</tr>
</thead>
<tbody>
<tr>
<td>How much did you enjoy taking part in WORD?</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>How helpful was it to think about the MEANING of words?</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>How helpful was it to think about the SOUNDS in words?</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>What helps you most when you are stuck?</td>
<td>Chunking out; doing the actions; sometimes spelling.</td>
<td>I show someone the action... Tell a teacher or friend.</td>
</tr>
<tr>
<td>Do you think finding words is easier now?</td>
<td>At the beginning 1 and now it is 3.</td>
<td>A little bit easier</td>
</tr>
</tbody>
</table>
Table 4: Conversation, counts using POWERS (Herbert et al., 2013).

<table>
<thead>
<tr>
<th></th>
<th>Amy</th>
<th></th>
<th>Magda</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-therapy</td>
<td>Post-therapy</td>
<td>Pre-therapy</td>
<td>Post-therapy</td>
</tr>
<tr>
<td>C1</td>
<td>C2</td>
<td>C3</td>
<td>C1</td>
<td>C2</td>
</tr>
<tr>
<td>Content words per turn</td>
<td>3.52</td>
<td>7.56</td>
<td>13.19</td>
<td>5.80</td>
</tr>
<tr>
<td>Errors per content word</td>
<td>0.26</td>
<td>0.19</td>
<td>0.13</td>
<td>0.46</td>
</tr>
</tbody>
</table>
Table 5: Parameter settings in the three double deficit models.

<table>
<thead>
<tr>
<th>Deficit</th>
<th>Deficit at semantic module</th>
<th>Deficit at phonological module</th>
</tr>
</thead>
</table>
| C at S+P | Connection density of SISH = 0.7  
Connection density of SHSO = 0.7 | Connection density of PIPH = 0.3  
Connection density of PHPO = 0.3 |
| H at S+P | Size of SH = 250 | Size of PH = 250 |
| T at S+P | Temperature of SH = 0.92 | Temperature of PH = 0.60 |
Figure captions

Figure 1. Design of the intervention study. A1 to A8 represent assessments. Randomisation (R) takes place after three pre-therapy baseline assessments. The baseline assessments were carried out over the half term immediately prior to randomisation. The comparison between the two random assignments (progress at A4) will form part of the wider intervention study which includes more children. Each phase of the study represented by a square (wait, therapy, wash-out and follow-up) lasted for 6 weeks (half a school term) and the assessment following each phase was carried out as soon as possible thereafter (i.e. on a later day in the final week of half term, in the seventh week of a longer half term or, less usually, during the school holiday).

Amy was randomly assigned to starting intervention immediately as illustrated in the top line of the figure. Magda was randomly allocated to the delayed condition. She was therefore assessed again prior to the start of the first therapy (A4) and eight times in total, as illustrated in the second line of the figure. Both Amy and Magda were randomly allocated to the phonological condition for therapy 1 and thus semantic for therapy 2.

Figure 2. Naming over the course of the study. The girls’ picture naming on the 4 experimental sets of 25 items at each assessment. Note Magda has 4 pre-therapy baselines as she was randomly assigned to the ‘wait’ condition.
Figure 3. The architecture of the model. Blue bars represent layers of units, red arrows represent learning weights between these units. Recurrent weights represented by yellow arrows were not trained. Abbreviated layer names stand for: SI - semantic input, SH - semantic hidden, SO - semantic output, PI - phonological input, PH - phonological hidden, PO - phonological output, AR - associative layer to the right (from the semantic to the phonological module), AL - associative layer to the left (from the phonological to the semantic module).

Figure 4. Developmental trajectories of the four core tasks across 4000 training epochs. Trajectories were calculated as medians from 50 TD models.

Figure 5. Comparison of TD and single location deficit models after 500 training epochs. The boxes represent the TD range (median +/- 1.5 standard deviations) calculated from 50 simulations. The separate data points represent different locations of the deficit calculated as the average of 10 atypical simulations: S - semantic module, P - phonological module, A - associative layers. Deficits were (A) lower connection density, (B) lower number of hidden units or (C) lower temperature of the sigmoid transfer function.

Figure 6. Comparison of TD and double location deficit models after 500 training epochs. The boxes represent the TD range (median +/- 1.5 standard deviations) calculated from 50 simulations. The separate data points represent different
types of the deficit calculated as the average of 10 atypical simulations: C - connection density, H - number of hidden neurons, T - temperature of the sigmoid transfer function. Deficits affected the semantic and the phonological modules to different extent.

Figure 7. Comparison of TD and double location deficit models after 500 training epochs. The boxes represent the TD range (median +/- 1.5 standard deviations) calculated from 50 simulations. The separate data points represent different types of the deficit calculated as the average of 10 atypical simulations: C - connection density, H - number of hidden neurons, T - temperature of the sigmoid transfer function. Deficits affected the semantic module, the phonological module and the associative layers to different extent.

Figure 8. Mean and SD of age difference of models when they reached 90% performance with and without intervention in the three double deficit group of models. Asterisks indicate effects that were significantly different from zero after a Bonferroni correction for multiple comparisons.

Figure 9. Mean and SD of age difference of models when they reached 90% performance with and without intervention in the three double deficit group of models. Models in deficit group H never reached 90% so in the case of this group age was measured when the model reached 65% performance instead.
Asterisks indicate effects that were significantly different from zero after a Bonferroni correction for multiple comparisons.
Figures

Figure 1
Figure 3
Figure 4.
Figure 5.

(A) Connection density deficit

(B) Hidden unit deficit

(C) Temperature deficit
Figure 6.
Figure 7.
Figure 8.
Figure 9.