Running head: PHONOLOGICAL SIMILARITY AND REHEARSAL IN CHILDREN

Re-evaluating key evidence for the development of rehearsal: phonological similarity effects in children are subject to proportional scaling artefacts

Christopher Jarrold & Rebecca Citroën

University of Bristol

Address correspondence to:

Chris Jarrold

School of Experimental Psychology

University of Bristol

12a Priory Road

Bristol

BS8 1TU

UK

Electronic mail: C.Jarrold@bristol.ac.uk

Telephone: +44 (0)117 928 8450

Facsimile: +44 (0)117 928 8588

Abstract

The size of an individual's phonological similarity effect for visually presented material is assumed to reflect their ability to recode, and by implication rehearse, information in verbal short-term memory. Many studies have shown that under these conditions the size of this effect interacts with age, tending to be non-significant in children younger than 7 years, and leading to the conclusion that children of this age do not rehearse. In the current study the size of the phonological similarity effect was assessed in a total of 116 children aged between 5 and 9 years, manipulating the modality of both encoding and retrieval of the memoranda. Although we replicated the interaction between age and the size of the phonological similarity effect with visual presentation and verbal recall of material, this interaction was also present in other conditions that do not require recoding. In addition, the data from this 'classic' condition were simulated by a model that assumed that the size of the similarity effect is i) proportional to an individual's recall of dissimilar items, and ii) constrained by a functional floor to recall of similar items. These findings undermine the evidence for a qualitative change in recoding and rehearsal at 7 years, and question the extent to which rehearsal is necessary to explain the development of verbal short-term memory performance.

Keywords: phonological similarity effect, phonological recoding, rehearsal, development of verbal short-term memory

Re-evaluating key evidence for the development of rehearsal: phonological similarity effects in children are subject to proportional scaling artefacts

Verbal short term memory, the process thought to underpin an individual's recall of words in correct serial order, plays an important role in cognitive development for at least two reasons. First, it supports working memory for verbal material (Alloway, Gathercole, & Pickering, 2006; Bayliss, Jarrold, Gunn, & Baddeley, 2003, Kane et al., 2004), and working memory is known to be a strong correlate of children's educational attainment in areas such as reading and mathematics (Bayliss, Jarrold, Baddeley, Gunn, & Leigh, 2005; Leather & Henry, 1994; Lépine, Barrouillet, & Camos, 2005; Swanson, 2008). Second, verbal short-term memory function is even more closely related to aspects of new word learning in children (Gathercole, Briscoe, Thorn, Tiffany, & the ALSPAC Study Team, 2008; Mosse & Jarrold, 2010), with many arguing that verbal short-term memory capacity is a causal factor in constraining children's vocabulary acquisition (Baddeley, Gathercole, & Papagno, 1998; Jarrold, Baddeley, Hewes, Leeke, & Phillips, 2004; Leclercq & Majerus, 2010). Consequently, understanding exactly what drives the development of verbal short-term memory performance in children is an important challenge for cognitive and educational psychology.

One possible source of developmental change that has been the subject of considerable previous investigation is subvocal rehearsal (see Dempster, 1981). Baddeley's (1986) influential model of working memory assumes that fading memory representations in a verbal short-term memory store can be maintained by being rehearsed. Other models similarly invoke rehearsal-like processes to explain at least a component of children's performance on verbal short-term memory tasks (Camos & Barrouillet, 2011; Cowan & Alloway, 2009). The most dominant view of the developmental course of rehearsal use in children is that there is a qualitative change in the use of this strategy at around 7 years of age (Baddeley et al., 1998; Gathercole, 1998; Gathercole, Adams, & Hitch, 1994; Henry, 1991; Hitch, Woodin, & Baker, 1989). For example, Henry (1991, p. 49) argues that children "do not seem to rehearse before the age of seven years".

This assertion followed initially from Flavell, Beach, and Chinsky's (1966) observation that spontaneous vocalization or lip movements during the delay of a delayed recall task, that might signal the rehearsal of the just-presented information, are rare among 5-year-olds but common in 7-year-olds. However, the most widely cited evidence for a qualitative change in the use of rehearsal around 7 years of age comes from studies that have assessed the phonological similarity effect for visually presented material. The phonological similarity effect reflects superior recall of phonologically dissimilar stimuli than of phonologically similar items (Conrad & Hull, 1964). When material is presented visually (i.e., pictures of the objects whose names are to be remembered) then the presence of this effect shows that the participant is recoding the visual stimulus into a verbal code. As this process of recoding involves subvocally naming the stimulus to oneself, many have argued that it is synonymous with rehearsal, and therefore provides evidence of an individual's ability to spontaneously rehearse (Gathercole, 1998; Howard & Franklin, 1990; Vallar & Papagno, 1995). Crucially, studies that have examined the size of the phonological similarity effect for visually presented material in children of different ages have shown that the absolute magnitude of this effect changes with age, being clear in older children but tending not to be significant in children aged less than 7 years (Halliday, Hitch, Lennon, & Pettipher, 1990; Hayes & Schulze, 1977; Hitch, Halliday, Schaafstal, & Heffernan, 1991; Hitch, Woodin, & Baker, 1989; Palmer, 2000a; see also Ford & Silber, 1994).

Although it is generally assumed that these data show that phonological recoding, and by implication rehearsal, undergoes qualitative development in children, there is evidence to suggest that the absolute size of the phonological similarity effect for visual material might instead change more gradually with age (see Ford & Silber, 1994; Palmer 2000b). Indeed, studies using this procedure with relatively large samples have shown small but significant phonological similarity effects in children aged 6 years and below (Al-Namlah, Fernyhough, & Meins, 2006; Tam, Jarrold, Baddeley, & Sabatos-DeVito, 2010). This finding raises the possibility that the phonological similarity effect for visually presented material is only smaller in younger individuals because of scaling or range effects (Jarrold & Tam, 2010; Tam et al., 2010, see Beaman, Neath, & Surprenant, 2008). Specifically, if the phonological similarity effect is a proportional effect, reflecting a constant proportional cost of similarity on performance, then the absolute size of the effect will be smaller among individuals who show a lower level of 'baseline' (in this case dissimilar) recall. In other words, the absolute size of the phonological similarity effect will scale with level of baseline performance, making it harder to detect in young children.

Of course, any developmental interaction can be 'explained away' in terms of proportional costs (see Cerella, 1985 and Verhaeghen, 2000 for examples from the aging literature), but there are additional empirical data and theoretical arguments in favour of a proportional scaling interpretation in this case. When pictorial stimuli have been named at the point of presentation, removing the need for recoding by the participant, children younger than seven still tend not to show significant phonological similarity effects (Conrad, 1971; Hayes & Rosner, 1975; Palmer, 2000a). This suggests that scaling issues, and not a failure to recode, limit the absolute size of the effect in younger individuals. Similarly, Hulme (1984, 1987) and Henry (1991) showed that even with verbal presentation of material, young children show a significantly smaller phonological similarity effect than do older participants.[1]

Crucially, significant age differences in the absolute size of children's phonological similarity effect for visually presented material tend to become non-significant when proportionalized against the length of the presentation lists given to different ages (Ford & Silber, 1994) or against level of dissimilar performance (Tam et al., 2010). These findings echo what is seen in adult studies. Beaman et al. (2008) have shown that, in adult participants, the size of the phonological similarity effect is proportional to an individual's level of dissimilar recall (see also Logie, Della Sala, Laiacona, Chambers, & Wynn, 1996). Beaman et al. explain this finding in terms of the Feature Model of verbal short-term memory (Nairne, 1990), which assumes that any feature of an item in memory can be overwritten by the occurrence of that feature in an item that is subsequently encoded. As the likelihood of a given feature re-occurring increases linearly with the number of items in memory, interference will be proportional to an individual's verbal short-term memory capacity. Even accounts of verbal short-term memory that do not rely on interference as the *primary* cause of forgetting assume that the phonological similarity effect reflects interference between items (Baddeley, 2012; Page & Norris, 1998), and most interference accounts would share the Feature Model's assumption that forgetting is proportional to the number of items held in short-term memory (cf. Brown & Hulme, 1995). There are therefore clear theoretical reasons for expecting the degree of forgetting caused by inter-item similarity to be proportional to a child's short-term memory capacity, which in turn would be expected to increase with age. However, to date this prediction has not been directly tested.

The current study therefore aimed to provide a thorough and comprehensive assessment of the size of the phonological similarity effect in children of different ages under different presentation and recall conditions. These manipulations allowed us to assess whether previous evidence of a developmental change in the absolute size of this effect with visual presentation amounts to evidence for a qualitative change in recoding, and by implication, rehearsal. To that end, children aged between 5 and 9 years of age were asked to recall phonologically dissimilar and similar material. However, in addition to simply employing visual presentation and requiring verbal recall, as has been the case in studies that have previously attempted to examine phonological recoding, we included the three other conditions that result from systematically crossing the mode of presentation (verbal or visual) with the mode of recall (verbal or visual). Clearly some of these conditions, e.g., verbal presentation and verbal recall, do not require recoding. Consequently a comparison of the size of the phonological similarity effect across the conditions of the experiment, and across ages, allows a proper test of whether phonological recoding undergoes qualitative developmental change. Equally, such a comparison can also determine whether the evidence previously thought to suggest a qualitative change in recoding instead reflects a proportional cost to phonological similarity that remains invariant across ages.

Method

Participants

One hundred and sixteen children (53 girls, 63 boys) were recruited from UK school years 1 to 4 of a single school, corresponding to US grade levels Kindergarten to Grade 3 (approximately 5-year-olds to 9-year olds). These individuals were divided equally into 4 groups each of which experienced a separate condition of the experiment in a between-participants design. Seven individuals from grade K, 8 from grade 1, 7 from grade 2, and 7 from grade 3 were allocated to each condition, to give a total of 29 individuals per condition. Allocation was conducted on the basis of individuals' digit recall scores in order to equate the verbal short-term memory skills of the participants in each condition prior to the test proper (see procedure section for details and results section for the adequacy of this matching).

Procedure

Preliminary assessment of digit recall.

In an initial testing session participants' digit recall was assessed by presenting lists of digits of increasing length for immediate spoken recall. The stimuli were the digits 1 to 9 that were pre-recorded and presented auditorily via computer at a rate of 1 per second. There were four trials at each list length, starting at list length 1, and lists were generated in advance so that the same trials were presented to each participant. No digits were repeated within a single list. Participants were instructed that serial recall of the items was required.

If a participant correctly recalled all items in correct order on one or more lists at a given list length then, after four trials at that list length, they moved on to the next list length; otherwise testing ended at that point. Performance was coded as a partial credit score (see Conway et al., 2005) in which the proportion of items on each trial recalled in correct serial position was totalled across all trials given to the participant.

Phonological similarity task.

The structure of this task was similar to that employed in the digit recall task, with four trials at each list length, the same progression/stopping rule, and the same partial credit scoring of performance. However, the task started at list length 2, and each participant completed two blocks of trials, one with phonologically dissimilar and one with phonologically similar memoranda. The order or presentation of these blocks was counterbalanced across participants, and each participant received a separate partial credit score for dissimilar and for similar items. The stimuli employed in the task were taken directly from Tam et al. (2010), which in turn were based on items employed by Conrad (1971). All were monosyllabic concrete nouns with age of acquisition norms of six years and below (Bird, Franklin, & Howard, 2001; Gilhooly & Logie, 1981; Morrison, Chappell, & Ellis, 1997). Items on phonologically dissimilar trials were selected from a pool of the following stimuli: *bell*, *bird*, *chair*, *door*, *fork*, *kite*, *shoe*, *tree*, *watch*, and the phonologically similar items were selected from the following stimulus set: *bag*, *bat*, *cat*, *hat*, *man*, *pan*, *rat*, *tap*, *van*. Prior to the testing proper, each participant was shown pictures of these items and asked to name them, with any incorrect naming (e.g., “mouse” for *rat*) corrected.

The between-participants manipulation of condition on this task involved crossing two encoding modes (verbal vs. visual) and two recall modes (verbal vs. visual) to create four conditions (cf. Murray, 1966). In Condition 1 (visual encoding verbal recall) and Condition 2 (visual encoding visual recall) to-be-remembered items were shown in pictorial form in the centre of the computer screen. These were the same as the images used in the familiarization phase, and appeared one at a time at a rate of 1 item per second. In Condition 3 (verbal encoding verbal recall) and Condition 4 (verbal encoding visual recall) pre-recorded stimuli were presented auditorily, without any visual support, again at a rate of 1 item per second. In Conditions 1 and 3 the participant recalled the presented items at the end of the trial orally. In Conditions 2 and 4 at recall the participant was presented with a screen that showed all nine pictures in a 3 x 3 arrangement, and had to touch the items that they wished to recall; the arrangement of these pictures on the recall screen changed after each trial. In all cases participants were instructed to recall the memoranda in correct serial order.

Results

Age and digit recall

Table 1 summarises the ages and digit span scores of individuals in each condition, split by school grade level. Analysis of variance of the age data with the between participants factors of condition and grade level showed an expected significant main effect of grade level, $F(3, 100) = 572.080$, $p < .001$, $MSE = 0.076$, $\eta_p^2 = .945$, but a non-significant main effect of condition, $F(3, 100) = 1.241$, $p = .299$, $MSE = 0.076$, $\eta_p^2 = .036$. The interaction between factors was close to significant, $F(9, 100) = 1.722$, $p = .094$, $MSE = 0.076$, $\eta_p^2 = .134$, and was explored further in terms of the simple main effect of condition on each grade level separately. The only significant main effect of condition emerged from the analysis of grade 2 children, $F(3, 24) = 3.890$, $p = .021$, $MSE = 0.051$, $\eta_p^2 = .327$, where grade 2 children in Condition 1 were significantly older than grade 2 children in Condition 3, $t(12) = 3.433$, $p = .005$, and Condition 4, $t(12) = 2.969$, $p = .012$ (see Table 1).

The corresponding analysis of digit span scores showed a significant main effect of grade level, $F(3, 100) = 304.333$, $p < .001$, $MSE = 13.343$, $\eta_p^2 = .406$, a clearly non-significant main effect of condition, $F(3, 100) = 0.067$, $p = .977$, $MSE = 13.343$, $\eta_p^2 = .002$, and a clearly non-significant interaction, $F(9, 100) = 0.065$, $p > .999$, $MSE = 13.343$, $\eta_p^2 = .006$. In other words, despite some age differences across conditions among grade 2 individuals, participants in each condition were very closely matched for their verbal short-term memory performance on this common task at all grade levels. As a result, any between-condition differences in recall of the stimuli in the phonological similarity task reported below can be directly attributed to the experimental manipulations that distinguish those conditions.

The phonological similarity task

Table 2 shows partial credit scores for each grade level in each condition on the dissimilar and similar recall trials of the phonological similarity task. In contrast to the above analyses, which simply served to show that individuals in each condition were matched for age and digit span, the data in Table 2 were analysed by splitting condition into two factors, encoding mode (visual vs. verbal) and recall mode (visual vs. verbal). Once again the effect of grade level was examined. However, because of concerns about the lack of power that would be associated with a four-factor analysis, a three-factor

ANOVA was instead employed with the dependent variable of the absolute size of the similarity effect shown by each individual (dissimilar – similar partial credit scores).[2]

This analysis revealed a main effect of encoding mode that was close to significant, $F(1, 100) = 3.304$, $p = .072$, $MSE = 8.094$, $\eta_p^2 = .032$. The absolute size of the similarity effect was numerically smaller for the visual encoding conditions (Conditions 1 and 2, $M = 3.174$, $SD = 2.085$) than it was for the verbal encoding conditions (Conditions 3 and 4, $M = 3.702$, $SD = 1.613$). The main effect of recall mode was significant, $F(1, 100) = 9.039$, $p = .003$, $MSE = 22.144$, $\eta_p^2 = .083$. In this case the similarity effect was reliably smaller for the verbal recall conditions (Conditions 1 and 3, $M = 3.010$, $SD = 1.870$) than it was for the visual recall conditions (Conditions 2 and 4, $M = 3.866$, $SD = 1.795$). The corollary of these two results is that the absolute size of the phonological similarity effect was smaller in Condition 1 than in any of the other three conditions (see Table 3). The two-way interaction between encoding mode and recall mode was not significant, $F(1, 100) = 1.770$, $p = .186$, $MSE = 4.336$, $\eta_p^2 = .017$.

The main effect of grade level was also significant, $F(3, 100) = 14.130$, $p < .001$, $MSE = 34.616$, $\eta_p^2 = .298$. The absolute size of the similarity effect increased steadily with grade level; grade K, $M = 2.018$, $SD = 1.630$; grade 1, $M = 3.229$, $SD = 1.422$; grade 2, $M = 3.900$ $SD = 1.536$; grade 3, $M = 4.634$, $SD = 1.946$. None of the interactions in the analysis were significant, highest F and lowest p = 1.291 and .282 respectively. Figure 1a plots the size of the similarity effect, by grade level, for each condition.

This initial analysis of the data in Table 2 therefore replicates the classic finding that the absolute size of the phonological similarity effect increases with age between 5 and 9 years. However, this increase is not limited to the traditional condition involving visual presentation and verbal recall (Condition 1), but rather extends across the other conditions of the experiment where recoding demands are presumably reduced, or non-existent (e.g., Condition 3, verbal recoding, verbal recall). Indeed, Condition 1 shows the smallest phonological similarity effect numerically. One potential explanation for this is that Condition 1 is more difficult than the other conditions. This can be seen from the dissimilar scores shown in Table 2; as the individuals in each condition are extremely well matched for digit span, there is no reason to expect their recall of dissimilar items to differ unless the encoding and recall manipulations affect task difficulty, yet these scores are numerically lower in Condition 1. An analysis of these dissimilar scores was carried out, again with the factors of encoding mode, recall mode, and grade level. This showed significant main effects of encoding mode, $F(1, 100) = 26.546$, $p < .001$, $MSE = 3.190$, $\eta_p^2 = .210$, and of recall mode, $F(1, 100) = 52.638$ $p < .001$, $MSE = 3.190$, $\eta_p^2 = .345$, with no significant interaction between these factors, $F(1, 100) < 1$. No interactions with grade level approached significance, all *F*s < 1. Table 3 summarises the dissimilar recall performance in each condition, and indicates that the main effect of encoding mode reflects a benefit to verbal ($M = 11.62$, $SD = 2.99$) over visual ($M = 9.90$, $SD = 3.12$) encoding. Conversely the main effect of recall mode reflects a benefit to visual ($M = 11.96$, $SD = 2.98$) over verbal ($M = 9.56$, $SD = 2.89$) recall.

The implication of these results is that Condition 1, which combines visual encoding with verbal recall, is the hardest condition in the experiment. In the Discussion we explain why this is an unsurprising finding. However, the immediate question that it raises is whether the relative reduction in the size of the phonological similarity effect in Condition 1 is simply the consequence of a constant proportional similarity effect operating on a lower baseline value than the other conditions. To address this question Table 3 also shows the proportional phonological similarity scores for each condition, coded as (dissimilar score – similar score) / dissimilar score (cf. Beaman et al., 2008), and Figure 1b plots these proportional scores by condition and grade level.

Figure 1b suggests that age related changes in the phonological similarity effect are eliminated by proportionalizing this score, with the exception of atypically low proportional scores for the children in grade K in Condition 1. The obviously larger standard deviation around the mean of this cell precludes an analysis of variance of the whole data shown in the figure. However, an ANOVA of the data from all conditions across grades 1 to 3 with the factors of encoding mode, recall mode, and grade level showed no significant main effects and no significant interactions, highest F and lowest p = 1.108 and .336 respectively. Similarly, an ANOVA of the data from across all grades in Conditions 1 to 3 with the factor of condition and grade level revealed non-significant main effects of each factor and no significant interaction, all Fs < 1.[3] In other words, while proportional scores are clearly lower for the children in grade K in Condition 1, they are comparable for all of the other data points represented in Figure 1b.

Modelling Condition 1 performance without assuming qualitative change with age

The results of the preceding analyses are largely consistent with the notion that developmental changes in the absolute size of the phonological similarity effect are simply a consequence of a constant proportional decrement caused by similarity being applied to different baseline levels of performance. However, while this account can readily explain age-related change in the absolute size of the phonological similarity effect in Conditions 2, 3 and 4 of the experiment, and between grade levels 1 and 3 in

Condition 1, it clearly fails to account for the particularly low proportional scores seen in Condition 1 among children in grade K (see Figure 1b). One reading of this latter finding is that this does indeed reflect a qualitative change in recoding, and by implication rehearsal, between the ages of 5 and 6; this being apparent in Condition 1 only because this is the condition that most requires recoding. In other words, while proportional scaling effects are undoubtedly operating in this dataset, there may still be evidence of a qualitative change in recoding and rehearsal from the condition traditionally used to explore this issue.

An alternative reading follows from the fact that the above analysis of dissimilar recall performance showed that Condition 1 was the hardest condition within the experiment. This means that dissimilar recall scores were particularly low for the children in grade K in this condition (see Table 2), raising the possibility that floor effects among these individuals reduced the size of their phonological similarity effect even more than would be predicted solely from proportional scaling. Specifically, if an individual has a low dissimilar partial credit score, then the similar score predicted by a proportional decrement may be below the functional floor for similar recall. The phonological similarity effect is known to arise as a result of the confusion of the *order* of the presented items and can actually aid memory for the identity of the *items* on a list (Fallon, Groves, & Tehan, 1999; Gupta, Lipinski, & Aktunc, 2005; Nairne & Kelley, 2004; Wicklegren, 1965). Order information is likely to be the major determinant of correct recall for longer lists, but will play a smaller role in constraining performance on short lists. Consequently the cost of phonological similarity may reduce on short lists, while the relative benefit from item cueing increases, reducing the difficulty of recalling information and resulting in a floor to performance on shorter lists. Indeed, one might well expect individuals to be able to successfully recall at least either the first or the last item on any similar list (cf. Jarrold et al., 2008), which would give them some credit under the kind of partial credit scoring scheme employed here. In addition, random variation cannot push this score below this functional floor but can operate in the other direction to increase the observed similar score. Consequently, at low levels of performance, a floor effect plus noise in the data will necessarily lead to a reduction, or even an inversion, of the expected similarity effect.

Given these arguments, in our final analysis we attempted to model the development of the phonological similarity effect seen in Condition 1 with just two main assumptions, namely that this effect is proportional to level of dissimilar recall and that there is a functional floor to the level of similar performance that can be seen. Figure 2 shows the proportional similarity effect plotted against dissimilar recall score for each individual in Condition 1. Crucially, these data were modelled by first determining the proportional size of the phonological similarity effect (proportion parameter), and the variance in this estimate (noise parameter), from the data from Conditions 2 to 4 (see appendix for details). In other words, we estimated the size of the proportional cost of similarity and the variance in this value from an entirely different set of participants who were tested in conditions that relied less, or not at all, on phonological recoding, and then applied these estimates to the prediction of Condition 1 performance.

These two fixed parameters, derived from the performance of the participants in Conditions 2 to 4, were then employed in a Monte Carlo simulation that generated 1000 hypothetical datasets based on the observed dissimilar scores of participants in Condition

1. For each Condition 1 participant a predicted similar score was determined by multiplying their dissimilar score by the proportion parameter extracted from the Condition 2 to 4 data. Then, in each separate simulation, individually sampled noise was added to each Condition 1 similar score, drawn from a Gaussian distribution that was scaled by the variance parameter extracted from the Condition 2 to 4 data. A floor effect to similar recall performance was then imposed on simulated Condition 1 scores. Any predicted similar score for Condition 1 that fell below the specified floor level was raised to this floor value. The floor value was a free parameter that was estimated in order to minimise the discrepancy between the modelled and observed Condition 1 data in terms of three variables. These were the intercept, the slope, and the fit of the modelled regression between dissimilar recall and proportional phonological similarity effect (cf. Figure 2), which were averaged across the 1000 simulations.

The left panel of Figure 2 shows the best-fit model of the Condition 1 data, which was obtained with a floor parameter of 4.45. The average intercept, slope, and r^2 values of the simulated datasets (see appendix) generated with this floor value were -0.325, 0.061, and .461, which compare to respective values of -0.318, 0.064, and .432 for the actual Condition 1 data. An obvious feature of this panel is that there appears to be an outlier in the real data that can be found towards the bottom left of the graph. It should be noted that at these low levels of dissimilar recall performance a small absolute difference in dissimilar and similar performance equates to a relatively large proportional difference, and that such negative proportional similarity effects are exactly what are predicted by our assumption of a floor on similar performance coupled with noise in this measure. Nevertheless, the right panel of Figure 2 replots the Condition 1 data without this data point. The best-fit model to these 28 Condition 1 data points was derived with a floor parameter value of 3.80. This led to average intercept, slope, and r^2 values of -0.112, 0.040, and .297, compared to respective values of -0.110, 0.044, and .434 for the actual data. Both panels of Figure 2 show that the actual data in Condition 1 can be closely modelled simply by assuming a fixed proportional phonological similarity effect, noise in this estimate, and a floor to similar recall performance, with the first two parameters being derived directly from the performance of an entirely separate group of participants from Conditions 2 to 4 of the experiment.

Discussion

The aim of this experiment was to critically evaluate the widely held view that the phonological similarity effect for visually presented information provides a test of individuals' ability to recode, and by implication rehearse, information in verbal short-term memory. In particular, we sought to examine whether the existing evidence of a change in the size of the phonological similarity effect with age, from typically non-significant levels before seven to clearly significant levels after seven, is evidence for a qualitative change in rehearsal status.

A first point to note about the current data is that this general finding was essentially replicated, with the size of the absolute phonological similarity effect increasing steadily with increasing grade level (see Figure 1a). However, this increase was not specific to Condition 1, the condition typically used to investigate the development of recoding and rehearsal, but rather was seen in all conditions. This pattern of findings already casts doubt on whether the developmental increase in the absolute size of the phonological similarity effect provides support for a qualitative change in recoding status. In particular, it would be very hard to argue that any degree of recoding is required in Condition 3 (verbal encoding and verbal recall).

Instead, a general pattern of a larger phonological similarity effect in older than younger children is consistent with the claim that this effect is simply a proportional one, with higher baseline levels of dissimilar recall in older children producing a larger phonological similarity effect in absolute terms. In addition, the analysis of levels of dissimilar recall across conditions showed significant main effects of both encoding and recall mode, with no sign of an interaction between them. This reflected a positive benefit to verbal presentation of material and a cost to verbal recall. Both of these effects are entirely in line with existing findings in the verbal short-term literature and were therefore to be expected (Harvey & Beaman, 2007; Murray, 1966). Crowder (1976) suggests that the modality effect, reflecting better recall of spoken over written versions of the same stimuli was noted as early as the start of the 20th century (Washburn, 1916). Similarly, output interference is an established phenomenon in verbal short-term memory (see Tan & Ward, 2007), and Harvey and Beaman (2007) have shown that the interference to still-to-be-recalled items in verbal short-term memory from outputted items is greater with verbal than with visual recall. Although therefore unsurprising, the key point about these input and output effects is that they combine to make Condition 1 the hardest condition in the current experiment. This, in turn, means that the reduction in absolute size of the phonological similarity effect seen in younger children as a result of proportional scaling will be most marked in this condition. In other words, it is likely to be particularly hard to observe a significant phonological similarity effect in young children with visual presentation and verbal recall of the material.

Indeed, when coded in proportional terms there was very little evidence that the phonological similarity effect differed across the various cells in the design of this experiment (see Figure 1b).[4] The one exception to this was the case of grade K children in Condition 1, where proportional phonological similarity effects were clearly lower than elsewhere. One might therefore argue that the current data do indeed provide evidence for a qualitative difference between Condition 1 and the other conditions of the experiment, and for a qualitative change in degree of recoding between approximately 5 and 6 years of age in this condition. However, we would argue for an alternative explanation of this particular finding, namely that floor effects, coupled with noise in the estimation of the phonological similarity effect lead to an additional reduction in the size of this effect at low levels of performance.

This position was strongly supported by our modelling of the Condition 1 data. This modelling was based on the two assumptions discussed above, namely that the phonological similarity effect was a fixed proportion of dissimilar performance and that there was a functional floor to the absolute level of similar recall performance that could be observed. Importantly, this modelling involved estimating the proportional phonological similarity cost, and associated noise, from completely separate participants, and therefore involved only one free parameter. This was the functional floor level for similar recall, which was adjusted to provide the best fit to the observed Condition 1 data. The resultant estimate of floor partial credit score of 4.45 provided an excellent fit to the data (see Figure 2), and, we would argue, is not unreasonably high. If one accepts that on any trial an individual will recall at least one item, be it the first or the last depending on their strategy, then just a 33% chance of successfully recalling just one other item on each list would on average produce a partial credit score of 4.45, assuming that a given individual was successful in recalling a second item on at least one trial at list length 2 in order to move on to list length 3. Indeed, other experimental data provide support for a functional floor to similar recall that is at least as high. Hulme (1984) examined the phonological similarity effect for auditorily presented material in children aged 4 to 10 years, and found that the average *span* scores for similar items for 4, 5, and 6 year olds were 3.41, 3.48, and 3.47 respectively. This contrasted to average spans for dissimilar items that increased across this age range (3.73, 4.08, and 4.79 respectively), suggesting that there may be a 3-item floor to similar recall in young children under these testing conditions.[5]

Our modelling exercise therefore demonstrates that the particularly small phonological similarity effects shown by the grade K children in Condition 1 can be readily explained by assuming a proportional cost of similarity, coupled with floor effects asymmetrically truncating the range of the similarity effect that is seen at low levels of recall. The former assumption is entirely in line with the evidence for the proportional effects of phonological similarity seen in adult data (Beaman et al., 2008), and in the other conditions of the current experiment (see Figure 1b). The suggestion that there is a functional floor to the size of an individuals' phonological similarity effect is a novel one that is supported by the success of our model (see Figure 2), but one that is both plausible and suggested by previous data. We accept that our particular model of the Condition 1 data has not been directly tested against another computational model of performance. However, we believe that our explanation of these data is far more parsimonious than the traditional verbal account of a qualitative change in recoding occurring between the ages of 5 and 9.

Further data from the current study strongly support this interpretation. In particular, the evidence presented here of a comparable phonological similarity cost in proportional terms across Conditions 2 to 4 is completely inconsistent with the view that some participants in each condition were not recoding visually presented material. If this were the case then a smaller similarity effect would be observed, for example, in Condition 2 (visual encoding visual recall) than in Condition 4 (verbal encoding visual recall). Instead it appears that the ability to verbally recode visually presented items is in place in children of 5 years of age, the youngest age assessed here. Although this suggestion is at odds with current assumptions in the short-term memory literature, it is entirely in line with the observation that infants start naming objects and pictures from the point at which they begin to make single word utterances.

It should be noted that a number of the short-term memory studies that have previously argued for a qualitative change in degree of verbal recoding have also shown that young children differ from older individuals in showing a greater detrimental effect of visual similarity (Hitch et al., 1988; Hitch, Woodin, & Baker, 1989; Williams, Happé, & Jarrold, 2008). Clearly one cannot ascribe a *greater* visual similarity effect in younger individuals to the kind of scaling effects that we are claiming reduce the phonological similarity effect, and so one must assume that younger children do have a greater tendency to maintain information in its visual form when given the opportunity to do so (see also Brown, 1977; Hayes & Schulze, 1977). What the current data imply is that this is not occurring at the expense of phonological encoding, but, rather, that as children age they move from this form of dual coding to sole reliance on phonological representations to support their recall (Palmer, 2000a).

Another caveat is that while evidence of a reduced phonological similarity effect for visually presented information in young children can no longer be used as positive evidence to support the view that rehearsal undergoes a qualitative change in development, this is not the same as disproving that such a change occurs. Some might argue that one should distinguish between the processes of recoding and rehearsal, and if so then it is possible that young children might recode visually presented information without subsequently rehearsing it (cf. Crowder, 1976, p. 86). Such a view would then still be consistent with Flavell et al.'s (1966) seminal evidence that children start to show signs of rehearsal of to-be-remembered material at round the age of 7.

Our view is that rehearsal would certainly be required in tasks that impose a maintenance interval between the presentation of the memoranda and their recall, as is the case in delayed span tasks such as those used by Flavell et al. (1966) or in working memory tasks that potentially provide opportunities for rehearsal in between the presentation of the to-be-remembered items (Jarrold, Tam, Baddeley, & Harvey, 2011; Tam et al. 2010). It remains an open question whether young children attempt to maintain information by subvocal rehearsal during such intervals (see Tam et al., 2010), and, if not, whether any rehearsal failure in these children reflects either quantitative changes in rehearsal efficiency with age or a qualitative change in the use of this strategy. However, the current data call into question the assumption that rehearsal is needed to account for the development of *immediate* serial recall performance, and, by implication, whether rehearsal actually occurs in immediate serial recall tasks at all. Clearly further work is therefore needed to properly determine the various causes of verbal short-term memory development in children, and the extent to which top-down maintenance strategies interact with bottom-up changes in capacity (cf. Cowan, AuBuchon, Gilchrist, Ricker, & Saults, 2011; Cowan, Morey, AuBuchon, Zwilling, & Gilchrist, 2010). However, the current data take an important step in this direction by ruling out one previously strongly held assumption, namely that a failure to rehearse in young children is demonstrated by their reduced phonological similarity effect for visually presented material.

References

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development, 77*, 1698-1716. doi: 10.1111/j.1467-8624.2006.00968.x

Al-Namlah, A. S., Fernyhough, C., & Meins, E. (2006). Sociocultural influences on the development of verbal mediation: Private speech and phonological recoding in Saudi Arabian and British samples. *Developmental Psychology, 42*, 117-131. doi: 10.1037/0012-1649.42.1.117

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1-29. doi: 10.1146/annurev-psych-120710-100422

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

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General, 132*, 71-92. doi: 10.1037/0096-3445.132.1.71

Bayliss, D. M., Jarrold, C., Baddeley, A. D., Gunn, D. M., & Leigh, E. (2005). Mapping the developmental constraints on working memory span performance. *Developmental Psychology, 41*, 579-597. doi: 10.1037/0012-1649.41.4.579

Beaman, C. P., Neath, I., & Surprenant, A. M. (2008). Modeling distributions of immediate memory effects: No strategies needed? *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, 34, 219-229. doi: 10.1037/0278-7393.34.1.219

Bird, H., Franklin, S, & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments & Computers. 33*, 73-79. doi: 10.3758/BF03195349

Brown, G. D. A., & Hulme, C. (1995). Modeling item length effects in memory span: no rehearsal needed. *Journal of Memory and Language, 34*, 594-621. doi: 10.1006/jmla.1995.1027

Brown, R. M. (1977). An examination of visual and verbal coding processes in preschool children. *Child Development, 48*, 38-45. doi: 10.2307/1128878

Camos, V., & Barrouillet, P. (2011). Developmental change in working memory strategies: From passive maintenance to active refreshing. *Developmental Psychology, 47*, 898-904. doi: 10.1037/a0023193

Cerella, J. (1985). Information processing rates in the elderly. *Psychological Bulletin, 98*, 67-83. doi: 10.1037/0033-2909.98.1.67

Conrad, R. (1971). The chronology of the development of covert speech in children. *Developmental Psychology, 5*, 398-405. doi: 10.1037/h0031595

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55*, 429-432. doi: 10.1111/j.2044-8295.1964.tb00928.x

Cowan, N., & Alloway, T. P. (2009). Development of working memory in childhood. In M. Courage & N. Cowan (Eds.), *The development of memory in infancy and childhood* (pp. 303-342). London: Psychology Press.

Cowan, N., AuBuchon, A. M., Gilchrist, A. L., Ricker, T. J., & Saults, J. S. (2011). Age differences in visual working memory capacity: not based on encoding limitations. *Developmental Science, 14,* 1066-1074. doi: 10.1111/j.1467-7687.2011.01060.x

Cowan, N., Morey, C. C., AuBuchon, A. M., Zwilling, C. E., & Gilchrist, A. L. (2010). Seven-year-olds allocate attention like adults unless working memory is overloaded. *Developmental Science, 13*, 120-133. doi: 10.1111/j.1467-7687.2009.00864.x

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review, 12*, 769-786. doi: 10.3758/BF03196772

Crowder, R. G (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.

Dempster, F. N. (1981). Memory span: sources of individual and developmental differences. *Psychological Bulletin, 89*, 63-100. doi: 10.1037/0033-2909.89.1.63

Fallon, A. B., Groves, K., & Tehan, G. (1999). Phonological similarity and trace degradation in the serial recall task: When CAT helps RAT, but not MAN. *International Journal of Psychology, 34*, 301-307. doi: 10.1080/002075999399602

Flavell, J. H., Beach, D. R., & Chinsky, J. M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development, 37*, 283-299. doi: 10.2307/1126804

Ford, S., & Silber, K. P. (1994). Working memory in children: A developmental approach to the phonological coding of pictoral material. *British Journal of*

Developmental Psychology, 12, 165-175. doi: 10.1111/j.2044-835X.1994.tb00626.x

Gathercole, S. E. (1998). The development of memory. *Journal of Child Psychology and Psychiatry, 39*, 3-27. doi: 10.1111/1469-7610.00301

Gathercole, S. E., Adams, A. M., & Hitch, G. J. (1994). Do young children rehearse? An individual differences analysis. *Memory and Cognition, 22*, 201-207. doi: 10.3758/BF03208891

Gathercole, S. E., Briscoe, J., Thorn, A., Tiffany, C., & the ALSPAC Study Team. (2008). Deficits in verbal long-term memory and learning in children with poor phonological short-term memory skills. *The Quarterly Journal of Experimental Psychology, 61*, 474 - 490. doi: 10.1080/17470210701273443

Gilhooly, K. J., & Logie, R. H. (1980) Age of acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research, Methods, and Instrumentation, 12*, 395–427. doi: 10.3758/BF03201693

Gupta, P., Lipinski, J., & Aktunc, E. (2005). Reexamining the phonological similarity effect in immediate serial recall: the roles of type of similarity, category cuing, and item recall. *Memory and Cognition, 33*, 1001-1006. doi: 10.3758/BF03193208

Halliday, M. S., Hitch, G. J., Lennon, B., & Pettipher, C. (1990). Verbal short-term memory in children: The role of the articulatory loop. *European Journal of Cognitive Psychology, 2*, 23-38. doi: 10.1080/09541449008406195

Harvey, A. J., & Beaman, C. P. (2007). Input and output modality effects in immediate serial recall. *Memory, 15*, 693-700. doi: 10.1080/09658210701644677

Hayes, D. S., & Rosner, S. R. (1975). The phonetic effect in preschool children: the influence of overt rehearsal and verbal instructions. *Journal of Experimental Child Psychology, 20*, 391-399. doi: 10.1016/0022-0965(75)90114-9

Hayes, D. S., & Schulze, S. A. (1977). Visual encoding in preschoolers' serial retention. *Child Development, 48*, 1066-1070. doi: 10.2307/1128362

Henry, L. A. (1991). The effects of word length and phonemic similarity in young children's short-term memory. *The Quarterly Journal of Experimental Psychology, 43A*, 35-52. doi: 10.1080/14640749108400998

Hitch, G. J., Halliday, M. S., Schaafstal, A. M., & Heffernan, T. M. (1991). Speech, ‘inner speech’, and the development of short-term memory: Effects of picture-labelling on recall. *Journal of Experimental Child Psychology, 51*, 220-234. doi: 10.1016/0022-0965(91)90033-O

Hitch, G. J., Woodin, M. E., & Baker, S. (1989). Visual and phonological components of working memory in children. *Memory and Cognition, 17*, 175-185. doi: 10.3758/BF03197067

Howard, D., & Franklin, S. E. (1990). Memory without rehearsal. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory* (pp. 287-318). Cambridge: Cambridge University Press.

Hulme, C. (1984). Developmental differences in the effects of acoustic similarity on memory span. *Developmental Psychology, 20*, 650-652. doi: 10.1037/0012-1649.20.4.650

Hulme, C. (1987). The effects of acoustic similarity on memory in children: A comparison between visual and auditory presentation. *Applied Cognitive*

Psychology, 45-51. doi: 10.1002/acp.2350010107

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434-446. doi: 10.1016/j.jml.2007.11.007

Jarrold, C., Baddeley, A. D., Hewes, A. K., Leeke, T., & Phillips, C. (2004). What links verbal short-term memory performance and vocabulary level? Evidence of changing relationships among individuals with learning disability. *Journal of Memory and Language, 50*, 134-148. doi: 10.1016/j.jml.2003.10.004

Jarrold, C., Cocksey, J., & Dockerill, E. (2008). Phonological similarity and lexicality effects in children's verbal short-term memory: Concerns about the interpretation of probed recall data. *The Quarterly Journal of Experimental Psychology, 61*, 324-340. doi: 10.1080/17470210701202210

Jarrold, C. & Tam, H. (2010). Rehearsal and the development of working memory. In P. Barrouillet, & V. Gaillard, (Eds.) *Cognitive Development and Working Memory: A Dialogue between Neo-Piagetian and Cognitive Approaches* (pp. 177-199). Hove: Psychology Press.

Jarrold, C., Tam, H., Baddeley, A. D., & Harvey, C. E. (2011). How does processing affect storage in working memory tasks? Evidence for both domain-general and domain-specific effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 688-705. doi: 10.1037/a0022527

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of*

Experimental Psychology: General, 133, 189-217. doi: 10.1037/0096-3445.133.2.189

Leather, C. V., & Henry, L. A. (1994). Working memory span and phonological awareness as predictors of early reading ability. *Journal of Experimental Child Psychology, 58*, 88-111. doi: 10.1006/jecp.1994.1027

Leclercq, A.-L., & Majerus, S. (2010). Serial-order short-term memory predicts vocabulary development: Evidence from a longitudinal study. *Developmental Psychology, 46*, 417-427. doi: 10.1037/a0018540

Lépine, R., Barrouillet, P., & Camos, V. (2005). What makes working memory spans so predictive of high level cognition? *Psychonomic Bulletin and Review, 12*, 165-170. doi: 10.1080/09541440440000014

Logie, R. H., Della Sala, S., Laiacona, M., Chambers, P., & Wynn, V. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. *Memory and Cognition, 24*, 305-321. doi: 10.3758/BF03213295

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *The Quarterly Journal of Experimental Psychology, 50A*, 528-559. doi: 10.1080/027249897392017

Mosse, E. K., & Jarrold, C. (2008). Hebb learning, verbal short-term memory, and the acquisition of phonological forms in children. *The Quarterly Journal of Experimental Psychology, 61*, 505-514. doi: 10.1080/17470210701680779

Murray, D. J. (1966). Vocalization-at-presentation and immediate recall, with varying recall methods. *Quarterly Journal of Experimental Psychology, 18*, 9-18. doi:

10.1080/14640746608400002

Nairne, J. (1990). A feature model of immediate memory. *Memory and Cognition, 18*, 251-269. doi: 10.3758/BF03213879

Nairne, J. S., & Kelley, M. R. (2004). Separating item and order information through process dissociation. *Journal of Memory and Language, 50*, 113-133. doi: 10.1016/j.jml.2003.09.005

Page, M. P. A., & Norris, D. (1998). The primacy model: a new model of immediate serial recall. *Psychological Review, 105*, 761-781. doi: 10.1037/0033-295X.105.4.761-781

Palmer, S. (2000a). Working memory: A developmental study of phonological recoding. *Memory, 8*, 179-193. doi: 10.1080/096582100387597

Palmer, S. (2000b). Development of phonological recoding and literacy acquisition: A four-year cross-sequential study. *British Journal of Developmental Psychology, 18*, 533-555. doi: 10.1348/026151000165841

Swanson, H. L. (2008). Working memory and intelligence in children: What develops? *Journal of Educational Psychology, 100*, 581-602. doi: 10.1037/0022-0663.100.3.581

Tam, H., Jarrold, C., Baddeley, A. D., & Sabatos-DeVito, M. (2010). The development of memory maintenance: Children's use of phonological rehearsal and attentional refreshment in working memory tasks *Journal of Experimental Child Psychology, 107*, 306-324. doi: 10.1016/j.jecp.2010.05.006

Tan, L., & Ward, G. (2007). Output order in immediate serial recall. *Memory and Cognition, 35*, 1093-1106. doi: 10.3758/BF03193481

Vallar, G., & Papagno, C. (1995). Neuropsychological impairments of short-term memory. In A. D. Baddeley, B. A. Wilson & F. Watts (Eds.), *Handbook of memory disorders* (pp. 135-165). Chichester: John Wiley & Sons.

Verhaeghen, P. (2000). The parallels in beauty's brow: Time-accuracy functions and their implications for cognitive aging theories. In T. J. Perfect and E. A. Maylor (Eds.), Models of cognitive aging (pp. 50-86). Oxford: Oxford University Press.

Washburn, M. F. (1916). *Movement and mental imagery*. Boston: Houghton Mifflin.

Wicklegren, W. A. (1965). Short-term memory for phonemically similar lists. *American Journal of Psychology, 78*, 567-574.

Williams, D., Happé, F., & Jarrold, C. (2008). Intact inner speech use in autism spectrum disorder: evidence from a short-term memory task. *Journal of Child Psychology and Psychiatry, 49*, 51-58. doi: 10.1111/j.1469-7610.2007.01836.x

Author Note

Preparation of this paper was supported by grants from the Economic and Social Research Council of the UK: RES-062-23-2467 and RES-062-23-1199. We are grateful to Debbora Hall and to Klaus Oberauer for their helpful comments on this work.

Footnotes

[1] Having said this, Henry's (1991) study of the development of the phonological similarity effect used a probed recall procedure that was designed to reduce output demands. Jarrold, Cocksey, and Dockerill (2008) noted that the combination of this particular procedure and the use of shorter list lengths for younger than for older children in this study reduced the power to detect the expected effect in younger children. Indeed, a replication using an alternative testing procedure which equated levels of baseline performance showed no evidence of a smaller phonological similarity effect for verbally presented material in 5- to 6- as opposed to 8- and 9-year-olds (Jarrold et al., 2008).

[2] Main effects and interactions from this three-factor analysis are statistically identical to the corresponding higher-order interactions with stimulus type (dissimilar vs. similar) in the four-factor analysis.

[3] Concerns are sometimes raised about the appropriateness of analysis of variance of proportional data (e.g., Jaeger, 2008). This follows from the fact that proportional scores are often truncated by their inability to exceed 1 or to fall below zero. However, in this case it is possible for proportional scores to be less than zero, and none of the mean values subjected to the just described ANOVAs were close to 1 (see Figure 1b). Indeed, formal tests of the homogeneity of variance for the two analyses just reported showed no evidence of any violation of the assumptions of analysis of variance, $F(11, 76) = 1.221$, $p = .288$, $F(11, 75) = 1.162$, $p = .328$. Furthermore, corresponding analyses using arcsine transformed proportional scores produced entirely comparable results.

[4] One set of data that appear not to fit with the current findings is that of Hulme (1987). This study essentially presented Conditions 2 and 3 of the current experiment to groups

of 4-, 7-, and 10-year-olds. Levels of dissimilar recall across these two conditions were similar, consistent with the current results (see Table 3). Despite this, Hulme (1987) found a significantly larger phonological similarity effect, when coded in absolute terms, with verbal presentation and verbal recall than with visual presentation and visual recall. Although these effects were not presented in terms of proportions, it appears that the proportional cost of similarity was greater for verbal presentation and recall in that previous study.

[5] Halliday et al. (1990) argued that floor effects could not explain the reduction of the phonological similarity effect in young children on the basis of a replication of this 'standard' finding using a fixed list length procedure, which they argued would allow for above floor performance to be observed. However, only 3 items were presented to 5-year-olds in this study and, as the Hulme (1984) data suggest, this list length might be close to the functional floor for recall of phonologically similar items in this age group.

Appendix

The derivation of the fixed parameter estimates employed in the simulation of Condition 1 data

The estimate of the proportional similarity effect shown by participants in Conditions 2, 3 and 4 was derived from the intercept of the regression line between the proportional phonological similarity effect and similar recall score for these individuals Although much shallower than the slope for the Condition 1 data shown in Figure 2, the slope of the regression line for the combined sample of individuals in the other conditions was significantly different from zero, $F(1, 85) = 4.862$, $p = .030$, $r^2 = .054$. Given that this deviation from flatness could be caused by floor effects on similar recall performance, we successively removed the individual with the lowest dissimilar score in Conditions 2 to 4 to investigate the effect of their removal on the regression line. Only two individuals had to be removed from the dataset for the slope of the regression line to become non-significantly different from zero. To estimate the proportional phonological similarity effect, the intercept of this not-significantly-sloped regression line was examined as participants were successively removed. The value of the intercept stabilized at .30, which was within the 95% confidence intervals for the whole range of sample sizes assessed (details available from the first author on request). This value was therefore adopted as a reasonable estimate of the fixed proportional effect of phonological similarity in Conditions 2 to 4 (see also Figure 1b). Predicted similar recall scores for Condition 1 were therefore estimated as .70 of each individual's observed dissimilar performance.

To determine the noise inherent in the estimate of the proportional phonological similarity effect we examined the residual difference between the observed similar recall score and the predicted similar recall score (.70 of the observed dissimilar recall score) of participants in Conditions 2 to 4, across the range of observed dissimilar scores. This allowed us to determine whether the noise in the proportional phonological similarity effect was constant across this range, or rather increased with increased dissimilar scores. For the positive residuals there was no evidence that these residuals varied in magnitude with dissimilar performance, $F(1, 30) < 1$. For the negative residuals there was a significant relationship between residual size and dissimilar score, $F(1, 53) = 7.011$, $p = .011$, $r^2 = .117$, due to a reduction in the size of negative residuals at lower levels of dissimilar performance. This presumably reflects the operation of a floor effect on similar performance at these lower levels, which reduces the possibility of showing the expected phonological similarity effect. Given this, the degree of variation in similar recall scores in Conditions 2 to 4, over and beyond the fixed 30% phonological similarity effect, was estimated by reflecting the positive residuals and deriving the population estimate of the standard deviation of the resultant symmetrical distribution. This produced a population estimate for the standard deviation of the variation of similar recall scores due to noise of 1.148 partial credit score points.

Table 1

Participant descriptive statistics. Digit spans are partial credit scores.

		Age (years)		Digit span	
Condition	Grade level	*M*	*SD*	*M*	*SD*
1	K	5.46	0.13	14.59	1.35
Visual-	1	6.50	0.28	17.32	3.11
verbal	2	7.53	0.22	18.86	2.14
	3	8.59	0.26	22.14	3.26
2	K	5.46	0.35	14.75	0.86
Visual-	1	6.52	0.33	17.54	3.91
visual	2	7.34	0.28	19.31	4.47
	3	8.34	0.40	22.56	4.46
3	K	5.37	0.28	14.43	3.64
Verbal-	1	6.55	0.23	17.78	4.42
verbal	2	7.15	0.19	18.81	5.71
	3	8.54	0.33	23.33	4.62
4	K	5.57	0.29	14.67	1.68
Verbal-	1	6.61	0.26	17.29	3.59
visual	2	7.19	0.21	19.24	4.18
	3	8.21	0.27	21.99	3.19

Table 2

Recall of phonologically dissimilar and similar material (partial credit scores).

		Dissimilar		Similar	
Condition	Grade level	*M*	*SD*	*M*	*SD*
1	K	5.01	2.02	4.90	1.44
Visual-	1	8.29	1.78	5.58	1.29
verbal	2	9.91	1.14	6.13	1.01
	3	11.49	1.47	7.92	1.38
2	K	8.38	1.63	5.51	0.77
Visual-	1	9.94	1.06	6.68	1.15
visual	2	11.95	1.62	7.63	1.57
	3	14.43	3.06	9.63	1.39
3	K	7.91	0.76	5.65	0.88
Verbal-	1	9.61	1.40	6.19	1.01
verbal	2	10.96	3.18	7.25	1.72
	3	13.42	1.41	8.95	1.11
4	K	9.34	0.91	6.55	1.02
Verbal-	1	12.30	2.21	8.78	1.48
visual	2	13.63	1.75	9.84	2.72
	3	15.96	1.18	10.25	0.93

Table 3

Summary by condition of dissimilar partial credit scores, absolute, and proportional phonological similarity effects (PSE).

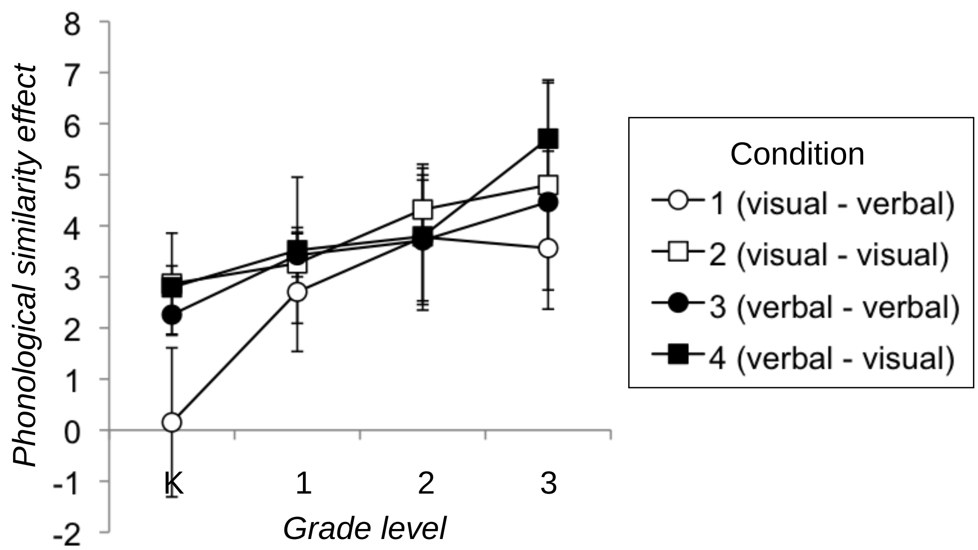
	Dissimilar		PSE absolute		PSE proportional	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	8.67	2.84	2.55	2.23	.237	.277
2	11.13	2.93	3.79	1.77	.331	.123
3	10.44	2.70	3.46	1.31	.326	.071
4	12.79	2.84	3.94	1.86	.302	.118

Figure Captions

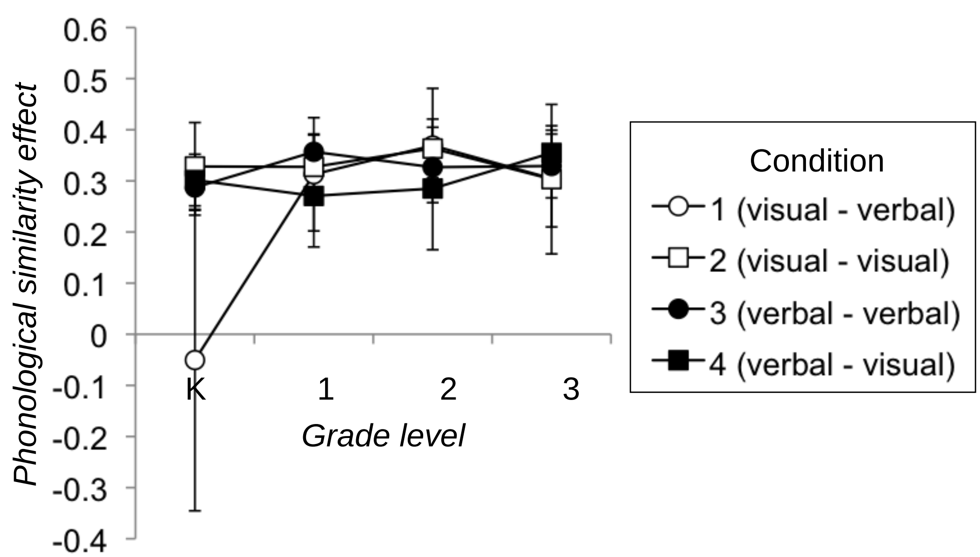
Figure 1. Developmental change in the size of the phonological similarity effect by condition. Upper panel plots absolute size of the phonological similarity effect (dissimilar score - similar score). Lower panel plots proportional size of the effect ((dissimilar score – similar score) / dissimilar score). 95% confidence intervals shown.

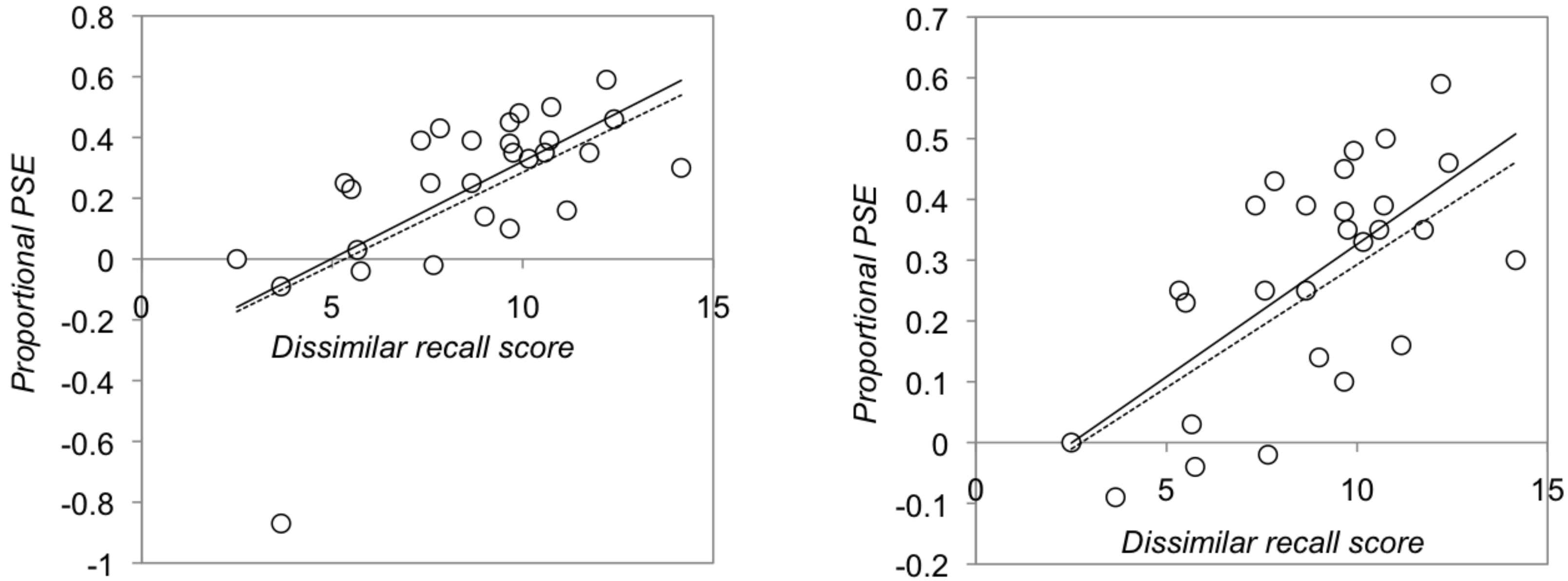
Figure 2. Modelling Condition 1 data: solid line is regression line for actual data, dotted line is regression line for model. Left panel shows all Condition 1 data, right panel removes one potential outlier.

1a Absolute size of the phonological similarity effect



1b Proportional size of the phonological similarity effect




0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
5
10
15
Proportional PSE
Dissimilar recall score
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
0
5
10
15
Proportional PSE
Dissimilar recall score