Article:
Wang, S and Allen, RJ (2018) Cross-modal working memory binding and word recognition skills: how specific is the link? Memory, 26 (4). pp. 514-523. ISSN: 0965-8211

https://doi.org/10.1080/09658211.2017.1380835

Cross-modal working memory binding and word recognition skills: How specific is the link?

Shinmin Wang[a] and Richard J. Allen[b]

[a] National Taiwan Normal University

[b] University of Leeds

Word count: 5501

Correspondence concerning this article should be addressed to Shinmin Wang, Department of Human Development and Family Studies, National Taiwan Normal University, 162, Sec.1, Heping E. Rd., Taipei 106, Taiwan. TEL: +886 (2)77341450. Email: s.wang@ntnu.edu.tw; and Richard Allen, School of Psychology, University of Leeds, Leeds, LS2 9JT, UK. TEL: +44 (0)1133432667. Email: r.allen@leeds.ac.uk.

Abstract

Recent research has suggested that the creation of temporary bound representations of information from different sources within working memory uniquely relates to word recognition abilities in school-age children. However, it is unclear to what extent this link is attributable specifically to binding ability for cross-modal information. This study examined the performance of Grade 3 (8-9 yrs old) children on binding tasks requiring either temporary association formation of two visual items (i.e., within-modal binding) or pairs of visually presented abstract shapes and auditorily presented nonwords (i.e., cross-modal binding). Children's word recognition skills were related to performance on the cross-modal binding task but not on the within-modal binding task. Further regression models showed that cross-modal binding memory was a significant predictor of word recognition when memory for its constituent elements, general abilities, and crucially, within-modal binding memory were taken into account. These findings may suggest a specific link between the ability to bind information across modalities within working memory and word recognition skills.

Key words: working memory; cross-modal binding; episodic buffer; word acquisition; Mandarin

Cross-modal working memory binding and word recognition skills: How specific is the link?

Successful word learning depends on a range of cognitive and linguistic factors. From the perspective of memory, visual word learning requires constitution of memory for word forms (orthography), pronunciations (phonology), meanings (semantics) and crucially, their associations. These three distinguishable but interlinked constituents of word identities are acknowledged in the lexical constituency model for word reading (Perfetti, Liu, & Tan, 2005). The connectionist model of reading proposed that visual word learning occurs by incremental changes in the strength of association between orthography and phonology (Seidenberg & McClelland, 1989). At a neuronal level, learning to read involves creating an invariant visual representation of written words in posterior visual regions, and connecting it to brain areas coding for speech sounds (temporal/parietal and anterior areas) and meaning (posterior and anterior regions) (see Dehaene, 2009 for a comprehensive review). Disruptions of these connections (e.g., Shaywitz et al., 2002) and problems in audiovisual integration (e.g., Harrar et al., 2014; Jones, Branigan, Parra, & Logie, 2013) have been observed in populations with dyslexia. Taking a different approach, a similar pattern of finding is also suggested in a recent study where the ability to temporarily bind information drawn from the phonological loop and visuospatial sketchpad within working memory has been found to uniquely link to word learning outcomes in typically developing school-aged children (Wang, Allen, Lee, & Hsieh, 2015). What remains unclear is whether this represents a specific link between cross-modal binding ability and printed word learning outcomes. The aim of the current study is hence to address this issue by including a within-modal binding task as well as a cross-modal binding task to see whether the predictive power of cross-modal binding in word reading skills can go above and beyond that of within-modal binding.

Successful visual word recognition involves the ability to create arbitrary associations between symbol and sounds and to retrieve the corresponding sound from the visually presented symbol. This process is conventionally well captured by the RAN (the rapid automatized naming) task, where participants were required to name aloud printed lists of digits or letters as fast as they can (see Bowey, 2005 for a review). It therefore taps the speed of retrieval of the learned symbol-sound associations. Arguably, the rapid retrieval is only possible when such associations have been successfully stored in memory and thus highly-learned. In light of this observation, researchers have started to investigate how associative learning ability contributes to visual word acquisition, and found that the ability to form association between auditory-verbal and visual materials in long-term memory is important for visual word acquisition in school-age children across different languages (e.g., Hulme, Goetz, Gooch, Adams, & Snowling, 2007; Li, Shu, McBride-Chang, Liu, & Xue, 2009; Messbauer & de Jong, 2003; Mourgues et al., 2016; Nielsen & Juul, 2016; Warmington & Hulme, 2012).

In alphabetic languages such as English, letter knowledge has been consistently found to be a crucial factor influencing subsequent word recognition development. The learning of letter sounds depends on the ability to associate printed letters (visual) to their corresponding sounds (auditory-verbal) (Hulme et al., 2007). Association ability may also be involved in visual word development by assisting the learning of large number of inconsistent mappings between phonemes and graphemes, as suggested in a recent Danish study (Nielsen & Juul, 2016). Likewise, in non-alphabetic languages such as Chinese, initial character learning may largely requires association formation between written characters and their pronunciations given the fact that Chinese characters are not designed to spell out their corresponding sounds in nature (Li et al., 2009; Wang et al., 2015).

The role of long-term association learning ability in developing word recognition skills has been receiving increased attention across different language systems (e.g., Hulme et al., 2007; Li et al., 2009; Messbauer & de Jong, 2003; Mourgues et al., 2016; Nielsen & Juul, 2016; Warmington & Hulme, 2012). In these studies, individual differences in association learning ability have been typically indexed by paired-associated learning (PAL) tasks. The PAL task requires participants to learn association between materials through repeated exposures. It therefore taps a series of cognitive processes including short-term memory for constituent materials, initial formation for their association, learning speed, and long-term retention of the learned associations. To further tease apart these processes and their relative contributions to word recognition skills, a recent study by Wang and colleagues (2015) took a different approach by focusing on the cognitive process related to the ability to form initial associations between materials in working memory and how it may contribute to integration of orthographic and phonological information in long-term memory, or in other words, to visual word acquisition.

The approach taken is noteworthy given that working memory is a system that provides temporary maintenance of information necessary for long-term learning (Baddeley, 2003). One of the influential theoretical approaches to understand working memory structure and functions among others (e.g., Barrouillet, Bernardin, & Camos, 2004; Cowan, 1999; Unsworth & Engle, 2007) is the multi-component model of working memory (Baddeley, 1986, 1996; Baddeley, 2012; Baddeley & Hitch, 1974). This framework describes the working memory system as including separable phonological and visuospatial short-term stores and a central executive control process, along with the more recent addition of the episodic buffer, a modality-general store capable of integrating information drawn from different sources in the environment and from long-term memory (Baddeley, 2000; Baddeley, Allen, & Hitch, 2011). Basic temporary storage of verbal or visuospatial information is typically measured by simple span tasks that primarily require information retention, whereas complex span tasks are designed to capture the simultaneous storage and the central executive control process. Using such measures, substantial research has shown links between working memory capacity and word acquisition (e.g., Baddeley, Gathercole, & Papagno, 1998; Gathercole, Willis, Emslie, & Baddeley, 1992; Majerus, Poncelet, Greffe, & van der Linden, 2006; Wang & Gathercole, 2013).

The ability to hold temporary bound information in working memory is conventionally measured by experimental binding tasks that require integration of individual features to form a bound representation, typically following a single exposure (e.g., Allen, 2015; Brockmole & Franconeri, 2009). Following this paradigm, Wang et al. (2015) developed a binding task in which immediate memory for pairs of visually presented abstract shapes and auditorily presented nonwords was assessed to examine whether temporary binding ability is related to word recognition skills in Mandarin-speaking children aged 8-10. The results show that children's performance on this task is a strong correlate of their word recognition skills, even when chronological age, nonverbal ability, memory for individual features that constitute the binding task, and other reading-related factors were considered. Building on this, corroborative evidence for such a relationship has also been found in an adult word learning context using the same working memory binding paradigm (Wang, Allen, Fang, & Li, in press). These findings therefore provide preliminary evidence for the link between individuals' working memory binding skills and their capacity to form long-term orthography-phonology associations (also see Jones et al., 2013).

However, it remains unclear whether there is a specific link between cross-modal working memory binding and printed word learning outcomes, or if the link observed simply reflects the contribution from a more general binding ability. To address this issue, the current study included a binding task requiring temporary association formation of two visual materials (i.e., within-modal binding) as well as the task developed by Wang et al. (2015) involving immediate memory for pairs of visually presented abstract shapes and auditorily presented nonwords memory (i.e., cross-modal binding). Our hypothesis was that if there is a specific link between cross-modal working memory binding abilities and developing word recognition skills, we would expect that cross-modal working memory binding performances remain a significant predictor of word recognition skills above and beyond within-modal working memory binding performances.

Method

Participants

A total of 152 Grade 3 children (70 boys and 82 girls; mean age = 8 years and 11 months, *SD* = 3.69 months, range = 8 years and 4 months to 9 years and 7 months) were recruited from three state primary schools in Taipei County, Taiwan. 22 children with reported developmental disorders or additional educational needs were excluded. Then, 12 children who produced guessing error rates (see details below) higher than .50 in either binding memory task (see details below) were also excluded, as this means that they responded by guessing randomly for more than half of the trials. In the end, data from 118 children (43 boys and 75 girls; mean age = 8 years and 11 months, *SD* = 3.72 months, range = 8 years and 4 months to 9 years and 7 months) were used in the current analyses. None of the children had any known additional learning needs or sensory impairments. All children were native Mandarin speakers. Informed parental consent was obtained and completed prior to participating. The study was approved by the Center for Research Ethics of the National Taiwan Normal University.

Procedure

Children were tested across three sessions. In the first session, the Raven's Progressive Matrices and measure of word recognition were given to the whole class (at the second week after the beginning of the second semester). Then, children received the other two sessions for the computerized working memory binding battery (see below for details) individually in a quiet room of the school. The measures of memory for nonwords, memory for shapes and memory for aliens were administered in the second session lasting 30 minutes. The administration order of the tasks was balanced across children. In the third session lasting 20-30 minutes, the measures of cross-modal binding memory and within-modal binding memory were conducted in a counter-balanced order across children. The third session was given to the children within a week after the second session, to ensure that visual and auditory-verbal elements were equally familiar to children when performing the relevant binding tasks.

Materials and Tasks

The Working memory binding battery

The computerized binding task designed by Wang et al. (2015) was adjusted to measure working memory for verbal-visual information binding (i.e., auditory nonword and shape - the cross-modal binding task) and working memory for visual-visual information binding (i.e., alien and shape – the within-modal binding task). In order to be able to separate participants' memory capacity for individual features from their ability to form associations between features in working memory, three corresponding feature memory tasks were also administered to measure memory for the constituent auditory-verbal materials (the auditory-verbal memory task) and visual materials (the alien memory task and the shape memory task).

In the two binding tasks, list length was set at 2 pairs[1]. In each of the feature memory tasks (nonwords; shapes; aliens), list length was set at 4 items to equate the number of individual features to be remembered between individual and binding memory conditions (Wang et al., 2015). Each task contained 10 experimental trials, preceded by 3 practice trials of a 2-item/2-pair sequence. Materials and procedures used in the five tasks were described in turn below.

The auditory-verbal memory task. The stimulus pool consisted of a set of 8 Mandarin auditory nonwords (*ga, bou, teng, mu, fao, hang, rei, se*) used in the study of Wang et al. (2015). The stimuli were sampled randomly without replacement within each trial and used for all participants.

At the study phase, a sequence of 4 auditory nonwords was presented via headphones. Each trial began with a black fixation cross presented at the upper centre of the screen for 500ms followed by a 250ms delay. Each to-be-remembered item was then presented for 1000ms, with inter-stimulus intervals of 250ms, with the screen

[1] We also ran a version of the binding tasks using sequences of 3 pairs, but this was too challenging for this age group, with children (N=130) performing at a very low level (cross-modal: *M* =.34, *SD* =.14; within-modal: *M* =.34, *SD* =.15). The current analysis was therefore limited to length 2 data sets.

remaining blank during presentation. A 1000ms delay followed offset of the final item in the sequence and was then followed by the test phase.

At the test phase, all 8 nonwords were displayed as response options in their visual forms--Zhu Yin Fu Hao (Fig. 1A) -- a phonetic symbol system designed to represent the sounds of Chinese characters in Taiwan. This design allows manual responses and therefore potentially rules out the possibility that requirement for spoken responses might drive any link observed between reading measures and other cognitive tasks (see Wang et al., 2015 for the design). The Zhu Yin Fu Hao were presented in the lower half of the screen, with each surrounded by a grey square outline (Fig. 1A). Cue items were selected in random order on each trial, thus nullifying the possible role of serial order mechanisms. The participants used the mouse to click the target items in any order. No serial order element was required as this was not an explicit part of the binding task. The grey square around the items turned green once selected and remained green till the end of the test phase as a reminder of which items had been selected. The trial ended when all responses had been made or when the total response time exceeded 24s, giving 6s on average for each response. The experimenter pressed the spacebar to start the next trial. Display locations of the 8 response options at test were randomized and changed across trials to prevent the potential use of location cues. The dependent variable was proportion of correct responses.

The shape memory task. The stimulus pool consisted of a set of 8 abstract and non-nameable six-point shapes drawn randomly from the study of Vanderplas and Garvin (1959) (number 3, 7, 13, 19, 20, 21, 29, 30). All stimuli were presented in black against a white background (Fig. 1B). They were sampled randomly without replacement within each trial and used for all participants.

The task procedure was identical to that employed in the auditory-verbal memory task, except that the experimental stimuli were replaced by a set of 8 shapes. At study, a sequence of 4 shapes was presented at the upper center of the screen at study. At test, the 8 possible choices were presented in the lower half of the screen (Fig. 1B). The participants had to click to select the target items in any order. The dependent variable was proportion of correct responses.

The alien memory task. The stimuli pool consisted of a set of 8 unfamiliar pictures of aliens taken from the study of Gupta et al. (2004) (set1B01, set1B02, set1B03, set1B04, set1C01, set1C02, set1C03, set1C04). All stimuli were presented in greyscale against a white background. They were sampled randomly without replacement within each trial and used for all participants (Fig. 1C).

The task procedure was identical to that employed in the visual memory task, except that the experimental stimuli were replaced by a set of 8 aliens. At study, a sequence of 4 aliens was presented at the upper center of the screen at study. At test, the 8 possible choices were presented and displayed evenly on the screen (Fig. 1C). The participants had to click to select the target items in any order. The dependent variable was proportion of correct responses.

The cross-modal binding task. The stimuli consisted of the 8 auditory nonwords and 8 shapes described above. At study, a sequence of 2 arbitrary pairs of auditory nonword and visual shape was presented (Fig. 1D). The task procedure was identical to that employed in the feature memory tasks, except that the presentation time for the constituent shape of each to-be-remembered pair was extended to 2000ms. The presentation time for each auditory nonword remained the same as that in the feature condition (i.e., 1000ms). This gave participants 2000ms to process each pair, and hence ensured equivalent feature processing for the individual feature and binding memory tasks.

At test, auditory nonwords were presented one at a time as retrieval cues. Simultaneously, participants saw all 8 possible shapes that made up the pair in the study phase displayed at the lower half of the screen, and were required to identify the target item by mouse clicking (Fig. 1D). The maximum response time for each cue was 6s. To prevent the role of serial order mechanism, cue items were randomly presented on each given trial. The dependent variable was proportion of correct responses. Additionally, guessing error rate was examined as a proportion of the total number of features that did not appear in the presented sequence but were selected. This response type represents children's tendency of random guesses when performing the binding task.

The within-modal binding task. The stimuli consisted of the 8 aliens and 8 shapes described above. The task procedure corresponded to the cross-modal binding task. At study, a sequence of 2 arbitrary pairs of pictures of alien and shape was presented (Fig. 1E). At test, pictures of alien were presented one at a time as retrieval cues. Participants simultaneously saw all 8 possible shapes that made up the pair in the study phase displayed at the lower half of the screen, and were required to identify the target item by mouse clicking (Fig. 1E). The dependent variable was proportion of correct responses. Guessing error rate was also calculated.

Other measures

Word recognition. Graded Chinese Character Recognition Test (Huang, 2001) was used as a standardized group administered and untimed reading measure with 200 single-syllable characters increasing in difficulty. This task is widely used in Taiwan to index children's word recognition abilities. Children were asked to write down the pronunciation of the character next to it using Zhu-Yin-Fu-Hao, with the dependent variable being the number of characters answered correctly. The raw score was transformed into a *T* score with population mean of 50 and standard deviation of 10. The test-retest reliability has been previously estimated at .89 for grade 2 and .84 for grade 3 (Huang, 2001).

Nonverbal ability. The Raven's Progressive Matrices for nonverbal intelligence ability (Raven, Court, & Raven, 2006). The raw score was converted to a standard score (*M*=100, *SD*=15) and used as the dependent variable.

Results

Descriptive statistics for the principle measures are displayed in Table 1. The simple correlations (Table 2, lower half) show that the measure of word recognition was significantly correlated with nonverbal ability, auditory memory, alien memory and crucially, with cross-modal binding memory but not within-modal binding memory. Outcomes from a Bayesian correlation analysis (JASP Team, 2016) are in line with this (Table 1, upper half), producing a Bayes Factor of around 48 for the correlation between word recognition and cross-modal binding (thus representing 'very strong' evidence for the correlation). This contrasts with BF = .77 for the correlation between word recognition and within-modality binding, which represents evidence (albeit 'anecdotal') for the absence of correlation in this case.

However, we note here that contrasting the presence vs. absence of correlation does not mean that these two correlations themselves differ (e.g., Gelman & Stern, 2006; Nieuwenhuis, Forstmann, & Wagenmakers, 2011). A direct (one-sided, overlapping, dependent groups) comparison of the correlations between word recognition and 1) cross-modal binding and 2) within-modal binding was therefore carried out using Dunn and Clark's (1969) z test within the *cocor* package (Diedenhofen & Musch, 2015). This indicated that the two correlations do not significantly differ, and that the null hypothesis should be retained ($z = 1.30$, $p = .096$). Thus, while we observed the predicted correlation between cross-modal binding and word recognition, and the absence of such a correlation in the case of within-modality binding, there was in fact little evidence that these two correlations themselves differ.

Three sets of hierarchical regression analyses were then carried out with word recognition as a dependent measure (Table 3). In regression 1 & 2, nonverbal ability was entered at Step 1 and corresponding feature memory at Step 2, to control effects of general ability and memory for individual feature. Regression 1 revealed that cross-modal binding memory was a significant predictor of word recognition above and beyond general ability and its constituent feature memory, and accounted for unique 4.4% of variance in word recognition. Together all variables explained 20.0% of the variance in word recognition. On the contrary, the second set of regression analysis indicated that within-modal binding memory was not a significant predictor of word recognition. In regression 3, within-modal binding memory was also entered in the model together with memory for the constituent parts of cross-modal binding memory (i.e., verbal memory and shape memory). The results indicated that cross-modal binding memory remained a unique predictor of word recognition even when within-modal binding memory was included in the model. Thus far, the regression results suggest a specific link between children's ability to bind information drawn across modalities in working memory and their word recognition skills.

However, we noted that the proportion correct data from measures of verbal memory (W=0.93, p<.001), alien memory (W=0.97, p=.006) and cross-modal binding memory (W=0.96, p=.001) were not normally distributed. To deal with this issue, we firstly tried to transform the data using three different ways including the log transformation, the square root transformation and the reciprocal transformation. Neither of them succeeded to correct the non-normality. We then applied the bootstrapping process (Efron & Tibshirani, 1994) using R software (R Core Team, 2017) to deal with the impact of bias due to the non-normality. As shown in table 3, all the bootstrap confidence intervals are close to the plug-in confidence intervals, suggesting that we did not have a problem of non-normal distribution in the model

(Field, Miles, & Field, 2012).

Following this, Bayesian regression analyses (JASP Team, 2016) were conducted to further explore this finding, examining either cross-modal or within-modal binding. In each case, word recognition was entered as the dependent variable, and nonverbal ability and feature memory classed as nuisance variables (to include them within the null model). For cross-modality binding, this produced a Bayes Factor of 4.6 (i.e. 'moderate' evidence for an effect), whereas the BF for within-modality binding was .324, or 3 to 1 in favour of the null. Finally, a Bayesian regression focusing on cross-modal binding that also included within-modality binding within the null model (alongside nonverbal ability and feature memory) produced a BF of 4.1. Thus, these additional Bayesian regression analyses corroborate the hierarchical regression outcomes and suggest a specific link between children's ability to bind information drawn across modalities in working memory and their word recognition skills.

Discussion

The aim of the present study was to evaluate the specificity of the link observed between cross-modal working memory binding and word recognition skills by incorporating within-modal as well as cross-modal working memory binding tasks. The results show that cross-modal working memory binding ability was uniquely associated with word recognition skills after nonverbal abilities, memory for the constituent materials, and crucially, within-modal binding memory were taken into account. This finding appears to suggest that this observed link between cross-modal binding task performance and character recognition skills could be attributable specifically to the binding ability for cross-modal information to some extent.

Our results point out that cross-modal binding ability based on a single exposure to pairs of visual and phonological features (that is, in the very early stage of associative learning) is specifically linked to word reading ability. The similar pattern of findings also appears in previous reports of disrupted performances on cross-modal association tasks in English-speaking populations with single word reading difficulties (e.g., Harrar et al., 2014), though discussed in the framework of multisensory integration and attention shifting rather than memory. In the study of Harrar et al., the authors found that dyslexic individuals have difficulties performing the cross-modal task only when the mapping is in the visual-verbal direction. Nevertheless, such finding does not necessarily preclude the possibility that individual differences in cross-modal task involving verbal-visual mapping direction will account for variance in word reading in a wider range of typically developing children, as shown in the present data. Of course, it will be fruitful to employ latent variable approach, where multiple measures of different types of cross-modal binding are obtained, to investigate whether tasks with different mapping directions will load on the same or different latent variables, and whether they make differential contributions to word reading skills in typically developing children.

The two binding tasks used in the present study were both selected to capture forms of extrinsic or relational binding (Ecker, Maybery, & Zimmer, 2013; Parra et al., 2013; (see Allen, 2015 for a review; also, Ecker, Maybery, & Zimmer, 2013; Parra et al., 2013). Within such tasks, the individual must encode and retain the associations between contextual features that do not form part of the same perceptual unit, a process potentially required by new word learning. As suggested by Ecker et al. (2013) and Parra et al. (2013), this broad category of task may be effortful and possibly draws on additional brain regions (e.g. the hippocampus), relative to tasks requiring intrinsic or conjunctive binding (such as the binding of colour and shape within a single object, e.g. Allen, Baddeley, & Hitch, 2006). The current results indicate that the possible relationship between cross-modal binding and word recognition reflects the role for this specific form of associative memory, rather than the operation of any type of extrinsic/relational binding. Little research has directly contrasted within- and cross-modal binding, and this work has typically focused on commonalities rather than differences (e.g. Baddeley et al., 2011; Gao, Wu, Qiu, He, Yang, & Shen, 2017). While future work might fruitfully explore the underlying mechanisms in each case, we would identify critical differences associated with variations in source modality, including involvement of distinct perceptual systems, and the challenges inherent in coordinating and binding temporally co-occurring visual and auditory information, versus information within multi-item visual arrays. We would also add that, while we did not test intrinsic/conjunctive binding in the present study, the absence of a relationship between word recognition and extrinsic within-modality visual binding means any link with intrinsic binding would be similarly unlikely to emerge.

A useful theoretical framework to understand our results is the multi-component model proposed by Baddeley and colleagues (Baddeley, 1986, 1996; Baddeley, 2012; Baddeley & Hitch, 1974). This model specifies distinctions between visuospatial and phonological modalities, and the means to integrate such information within a modality-general episodic buffer (see also Barrouillet & Camos, 2014). This latter component is assumed to comprise a storage capacity based on a multidimensional code. Binding materials from different domains or modalities (as examined by the cross-modal binding task in the current study) may particularly require the episodic buffer for their formation and retention, as implied by Baddeley's (2000) original proposal. In addition, this buffer may serve as a storage and modelling space that is informed by but separable from long-term memory (Allen, Havelka, Falcon, Evans, &

Darling, 2015; Langerock, Vergauwe, & Barrouillet, 2014), and may form an important stage in long-term episodic learning (Baddeley, 2003). The episodic buffer concept may therefore serve a useful purpose in understanding the observed relationship between cross-modal binding within working memory and word acquisition in long-term memory, through its proposed position at the interface of phonological processing, visuospatial processing, and long-term memory. This is not to claim that any task broadly linked with the episodic buffer will be involved in word learning, as indeed indicated by the absence of relationship with within-modality binding in the present study. Processes feeding into the initial formation and retention of representations will differ depending on a range of factors including type of material and source modality. It is the specific mechanisms involved in initially associating visual and auditory-verbal information that we would suggest are shared in both working memory and word learning tasks. The concept of the episodic buffer simply captures how this associative processing might be temporarily retained and connected to LTM.

We would note that, based on the current findings, the contributions of phonological memory and cross-modal binding in visual word learning are not mutually exclusive, but complementary. As shown in our regression results (see table 3), both verbal memory and cross-modal binding memory make specific and distinguishable contributions in predicting word recognition skills. This pattern is in line with the idea that learning to read involves creating visual representations of word forms and connecting these to brain areas coding phonological representation of speech sounds (Dehaene, 2009). The current findings do not intend to revisit the role of phonological loop in learning phonological structure of a given word. Instead, they highlight the additional contribution that cross-modal binding ability may make towards developing visual word recognition.

Of course, alternative perspectives on working memory are also likely to prove useful in considering how cross-modal binding might be achieved, and how this might relate to longer-term word acquisition. Within the embedded processes model of Cowan (e.g., 1999; 2014) for example, information from disparate sources could be temporarily retained together within a limited capacity focus of attention. While the current work was not designed to differentiate between theories of working memory, it may prove fruitful for different perspectives to consider how information from disparate sources is brought together and how this might relate to broader cognition.

Moreover, it is worth discussing how cross-modal binding abilities might be involved in the process of word acquisition. A possible answer to this question has been discussed in the context of alphabetic languages. For example, in opaque orthography, such as English and Danish, there are many inconsistent mappings between phonemes and graphemes. It has therefore been suggested that learning sound-letter mapping may draw on the ability to form association between them (e.g., Nielsen & Juul, 2016). How might this relate to Chinese, as in the present study? As an important contrast to alphabetic systems, Chinese is considered as a morpho-syllabic writing system where characters map onto phonology at syllable level (e.g., Newman, Tardif, Huang, & Shu, 2011; Tan, Laird, Li, & Fox, 2005). The correspondences between Chinese characters and syllables are opaque and without reliable rules to follow, particularly for beginning readers (Shu, 2003). Using the characters taught at Grade 1 (approx. 700 characters; Wang, Hung, Chang, & Chen, 2008) as an example, less than half of them (45%) are semantic-phonetic compounds that contain a component signaling semantics and a component signaling pronunciation. Only 22% of the semantic-phonetic compounds taught at Grade 1 contain phonetics that provide useful pronunciation cues (Shu, Chen, Anderson, Wu, & Xuan, 2003). In addition, most of the phonetics themselves are stand-alone characters (e.g., 青 in the case of 清) that do not provide pronunciation cues. Children have to memorize the sounds of the phonetics before they use phonetics as cues. Consequently, acquisition of characters in early stages of learning to read Chinese apparently needs to rely on children’s ability to bind information drawn across verbal (i.e., pronunciation) and visual (i.e., character) modalities. Given that the processes of word acquisition are expected to differ across orthographies, it will be worthwhile for future work to examine whether a similar relationship is observable in other languages.

To sum up, from the working memory perspective, we identify a specific link between temporary binding of information across modalities (i.e., auditory-verbal and visual) and word acquisition by incorporating within-modal as well as cross-modal binding tasks and comparing their relative contributions to word recognition skills. This finding is consistent with the idea that successful word learning requires constitution of memory for word forms (orthography), word pronunciations (phonology), and crucially, their associations. As such, it replicates and extends the findings of Wang et al. (2015). The significance of the current study is to take a step further by indicating the specificity of the link between cross-modal working memory binding and word acquisition. The present findings indicate that this observed link is more likely to represent the unique contribution from cross-modal binding, rather than a more general binding ability.

Acknowledgements

This research is partially supported by the “Aim for the Top University Project” and “Center of Learning Technology for Chinese” of National Taiwan Normal University (NTNU), sponsored by the Ministry of Education, Taiwan, R.O.C.. Partial support is also provided by the Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. MOST 104-2911-I-003-301 (“International Research-Intensive Center of Excellence Program” of NTNU) and Grant no. MOST 105-2410-H-003-091. We thank Peng-Wen Lin for her assistance with running the experiment.

References

Allen, R. J. (2015). Memory binding. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioural Sciences* (2 ed.): Elsevier.

Allen, R. J., Havelka, J., Falcon, T., Evans, S., & Darling, S. (2015). Modality specificity and integration in working memory: Insights from visuospatial bootstrapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(3), 820.

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

Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology: Section A, 49*(1), 5-28.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417-423.

Baddeley, A. D. (2003). Working memory and language: an overview. *Journal of Communication Disorders, 36*(3), 189-208. doi: 10.1016/S0021-9924(03)00019-4

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

Baddeley, A. D., Allen, R. J., & Hitch, G. J. (2011). Binding in visual working memory: The role of the episodic buffer. *Neuropsychologia, 49*(6), 1393-1400. doi: 10.1016/j.neuropsychologia.2010.12.042

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent Advances in Learning and Motivation* (Vol. 8). New York: Academic Press.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133*(1), 83.

Barrouillet, P., & Camos, V. (2014). *Working memory: Loss and reconstruction*: Psychology Press.

Bowey, J. A. (2005). Predicting individual differences in learning to read. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading* (pp. 155-172). Oxford, UK: Blackwell Publishing Ltd.

Brockmole, J. R., & Franconeri, S. L. (2009). *Binding: A special issue of visual cognition*. Hover, UK: Psychology Press.

Cowan, N. (1999). An embedded-processes model of working memory. *Models of working memory: Mechanisms of active maintenance and executive control, 20*, 506.

Cowan, N., Saults, J. S., & Blume, C. L. (2014). Central and Peripheral Components of Working Memory Storage. *Journal of Experimental Psychology: General, 143*(5), 1806-1836.

Dehaene, S. (2009). *Reading in the brain: The new science of how we read*: Penguin.

Diedenhofen, B., & Musch, J. (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PloS one, 10*(4), e0121945. doi: 10.1371/journal.pone.0121945

Dunn, O. J., & Clark, V. (1969). Correlation coefficients measured on the same individuals. *Journal of the American Statistical Association, 64*(325), 366-377.

Ecker, U. K. H., Maybery, M., & Zimmer, H. D. (2013). Binding of intrinsic and extrinsic features in working memory. *Journal of Experimental Psychology: General, 142*(1), 218-234. doi: 10.1037/a0028732

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*: CRC press.

Field, A., Miles, J., & Field, Z. (2012). Discovering statistics using R: Sage, Thousand Oaks.

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28*(5), 887-898. doi: 10.1037/0012-1649.28.5.887

Gelman, A., & Stern, H. (2006). The Difference between "Significant" and "Not Significant" Is Not Itself Statistically Significant. *The American Statistician, 60*(4), 328-331.

Gupta, P., Lipinski, J., Abbs, B., Lin, P.-H., Aktunc, E., Ludden, D., . . . Newman, R. (2004). Space aliens and nonwords: stimuli for investigating the learning of novel word-meaning pairs. *Behavior Research Methods, Instruments, & Computers: A Journal Of The Psychonomic Society, Inc, 36*(4), 599-603.

Harrar, V., Tammam, J., Perez-Bellido, A., Pitt, A., Stein, J., & Spence, C. (2014). Multisensory Integration and Attention in Developmental Dyslexia. *Current Biology*(5), 531. doi: 10.1016/j.cub.2014.01.029

Huang, H.-S. (2001). *Graded Chinese character recognition test*. Taipei, Taiwan: Psychology Publication.

Hulme, C., Goetz, K., Gooch, D., Adams, J., & Snowling, M. J. (2007). Paired-associate learning, phoneme awareness, and learning to read. *Journal of Experimental Child Psychology, 96*(2), 150-166. doi: 10.1016/j.jecp.2006.09.002

JASP Team. (2016). JASP (Version 0.8.0.0)[Computer software].

Jones, M. W., Branigan, H. P., Parra, M. A., & Logie, R. H. (2013). Cross-modal binding in developmental Dyslexia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi: 10.1037/a0033334

Langerock, N., Vergauwe, E., & Barrouillet, P. (2014). The maintenance of cross-domain associations in the episodic buffer. *Journal of experimental psychology. Learning, memory, and cognition, 40*(4), 1096-1109. doi: 10.1037/a0035783

Li, H., Shu, H., McBride-Chang, C., Liu, H. Y., & Xue, J. (2009). Paired associate learning in Chinese children with dyslexia. *Journal of Experimental Child Psychology, 103*(2), 135-151. doi: 10.1016/j.jecp.2009.02.001

Majerus, S., Poncelet, M., Greffe, C., & van der Linden, M. (2006). Relations between vocabulary development and verbal short-term memory: The relative importance of short-term memory for serial order and item information. *Journal of Experimental Child Psychology, 93*(2), 95-119.

Messbauer, V. C. S., & de Jong, P. F. (2003). Word, nonword, and visual paired associate learning in Dutch dyslexic children. *Journal of Experimental Child Psychology, 84*(2), 77-96. doi: 10.1016/S0022-0965(02)00179-0

Mourgues, C., Tan, M., Hein, S., Ojanen, E., Reich, J., Lyytinen, H., & Grigorenko, E. L. (2016). Paired associate learning tasks and their contribution to reading skills. *Learning and Individual Differences, 46*, 54-63. doi: 10.1016/j.lindif.2014.12.003

Newman, E. H., Tardif, T., Huang, J., & Shu, H. (2011). Phonemes matter: The role of phoneme-level awareness in emergent Chinese readers. *Journal of Experimental Child Psychology, 108*(2), 242-259.

Nielsen, A.-M. V., & Juul, H. (2016). Predictors of early versus later spelling development in Danish. *Reading and Writing, 29*(2), 245-266. doi: 10.1007/s11145-015-9591-y

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: a problem of significance. *Nature Neuroscience, 14*(9), 1105-1107.

Parra, M. A., Fabi, K., Luzzi, S., Cubelli, R., Hernandez Valdez, M., & Della Sala, S. (2013). Relational and conjunctive binding functions dissociate in short-term memory. *Neurocase*, 1-11. doi: 10.1080/13554794.2013.860177

Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The Lexical Constituency Model: Some Implications of Research on Chinese for General Theories of Reading. *Psychological Review, 112*(1), 43-59.

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Raven, J. C., Court, J. H., & Raven, J. (2006). *Raven's Progressive Matrices*. U.S.A.: Harcourt Assessment, Inc.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523-568. doi: 10.1037/0033-295X.96.4.523

Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., Mencl, W. E., Fulbright, R. K., Skudlarski, P., . . . Gore, J. C. (2002). Disruption of posterior brain systems for reading in children with developmental dyslexia. *Biological Psychiatry, 52*(2), 101-110. doi: http://dx.doi.org/10.1016/S0006-3223(02)01365-3

Shu, H. (2003). Chinese writing system and learning to read. *International Journal of Psychology, 38*(5), 274-285. doi: 10.1080/00207590344000060

Shu, H., Chen, X., Anderson, R. C., Wu, N., & Xuan, Y. (2003). Properties of school Chinese: Implications for learning to read. *Child Development, 74*(1), 27-47.

Tan, L. H., Laird, A. R., Li, K., & Fox, P. T. (2005). Neuroanatomical correlates of phonological processing of Chinese characters and alphabetic words: A meta-analysis. *Human brain mapping, 25*(1), 83-91.

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: an examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin, 133*(6), 1038.

Vanderplas, J. M., & Garvin, E. A. (1959). The associative values of random shapes. *Journal of Experimental Psychology, 57*, 147-154.

Wang, C.-C., Hung, L.-Y., Chang, Y.-W., & Chen, H.-F. (2008). Number of characters school students know from grade 1 to 9. *Bulletion of Education Psychology (in Chinese), 39*(4), 555-568.

Wang, S., Allen, R. J., Fang, S.-Y., & Li, P. (in press). Cross-modal working memory binding and L1-L2 word learning. *Memory and Cognition*. doi: 10.3758/s13421-017-0731-2

Wang, S., Allen, R. J., Lee, J. R., & Hsieh, C.-E. (2015). Evaluating the developmental trajectory of the episodic buffer component of working memory and its relation to word recognition in children. *Journal of Experimental Child Psychology, 133*, 16-28. doi: 10.1016/j.jecp.2015.01.002

Wang, S., & Gathercole, S. E. (2013). Working memory deficits in children with reading difficulties: Memory span and dual task coordination. *Journal of Experimental Child Psychology, 115*(1), 188-197. doi: 10.1016/j.jecp.2012.11.015

Warmington, M., & Hulme, C. (2012). Phoneme awareness, visual-verbal paired-associate learning, and rapid automatized naming as predictors of individual differences in reading ability. *Scientific Studies of Reading, 16*(1), 45-62. doi: 10.1080/10888438.2010.534832

Table 1

Descriptive Data for The Principle Measures

	Mean	SD	Min	Max
Nonverbal Ability (standard score)	104.14	13.77	71.00	134.00
Word Recognition (raw, Max=200)	86.40	25.50	8.00	137.00
Word Recognition (*T* score)	59.43	11.13	25.00	81.71
Auditory Memory (ACC)	0.91	0.07	0.68	1.00
Visual Memory (ACC)	0.70	0.10	0.40	1.00
Alien Memory (ACC)	0.74	0.09	0.53	1.00
Cross-Modal Binding Memory (ACC)	0.72	0.18	0.25	1.00
Within-Modal Binding Memory (ACC)	0.64	0.18	0.20	1.00

Table 2

Correlations between Measures, with Pearson's r (lower half) and Bayes Factors (upper half)

	1	2	3	4	5	6	7
1. Nonverbal Ability	1	34.89	0.43	53.35	940.07	103.28	3.06
2. Word Recognition	.309**	1	13.19	0.37	268.05	47.57	0.77
3. Auditory Memory	.151	.282**	1	0.12	0.13	0.17	1.86
4. Shape Memory	.319**	.142	-.008	1	52497	203.56	182.4
5. Alien Memory	.382**	.357**	.051	.452**	1	1997.29	2017
6. Cross-Modal Binding Memory	.335**	.317**	.081	.351**	.397**	1	38.82
7. Within-Modal Binding Memory	.236**	.181	.218*	.348**	.397**	.311**	1

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

Table 3

Hierarchical Regressions

	Step	Independent Variables	Dependent Variable: Character Recognition										
				(CI 95%, plug-in)		(CI 95%, bootstrap)							
			b	LB	UB	LB	UB	SE *B*	final β	*p*	Total R^2	ΔR^2	*p*
Regression 1		Constant	-1.197	-29.381	26.988	-30.284	28.638	14.226					
	1	Nonverbal Ability	0.158	0.010	0.307	-0.001	0.323	0.075	.196	.037	.095	.095	.001
	2	Auditory Memory	37.200	10.328	64.071	12.510	64.560	13.564	.234	.007	.156	.061	.019
		Shape Memory	0.047	-19.717	19.812	-23.401	20.131	9.976	.000	.996			
	3	Cross-Modal Binding Memory	14.264	2.938	25.590	3.030	26.590	5.717	.232	.014	.200	.044	.014
Regression 2		Constant	18.669	-0.171	37.509	0.940	39.590	9.509					
	1	Nonverbal Ability	0.170	0.018	0.322	-0.032	0.314	0.077	.210	.028	.095	.095	.001
	2	Alien Memory	36.490	10.808	62.171	7.130	58.280	12.963	-.071	.479	.165	.070	.010
		Shape Memory	-7.637	-28.937	13.662	-34.450	12.760	10.751	.293	.006			
	3	Within-Modal Binding Memory	2.459	-9.381	14.300	-0.130	25.220	5.977	.039	.682	.166	.001	.682
Regression 3		Constant	-0.898	-29.506	27.709	-30.724	29.713	14.438					
	1	Nonverbal Ability	0.157	0.008	0.307	-0.006	0.322	0.075	.195	.039	.095	.095	.001
	2	Auditory Memory	36.774	9.151	64.397	10.630	65.890	13.941	.231	.010	.156	.061	.019
		Shape Memory	-0.345	-20.920	20.231	-24.945	22.149	10.384	-.003	.974			
	3	Within-Modal Binding Memory	0.850	-10.873	12.573	-11.610	11.807	5.917	.014	.886	.158	.003	.563
	4	Cross-Modal Binding Memory	14.111	2.542	25.680	3.070	26.440	5.839	.229	.017	.200	.042	.017

Note. LB denotes lower bound, and UB denotes upper bound. Boostrap confidence interval calculations are based on 2000 bootstrap replicates.

Table Captions

Table 1 *Descriptive Data for The Principle Measures*

Table 2 *Correlations between Measures, with Pearson's r (lower half) and Bayes Factors (upper half)*

Table 3 *Hierarchical Regressions*