Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis

Barbara Carretti *, Erika Borella, Cesare Cornoldi, Rossana De Beni

Department of General Psychology, University of Padova, Italy

A R T I C L E   I N F O

Article history:
Received 14 January 2008
Received in revised form 22 September 2008
Accepted 5 October 2008

Keywords:
Reading comprehension difficulties
Working memory
Inhibition

A B S T R A C T

It is well established that working memory (WM) is involved in reading comprehension. However, depending on the theoretical approach the structural and functional role of WM in reading comprehension is explained in different ways. Referring to the original model proposed by Baddeley and Hitch (1974), some researchers suggest that the relationship between WM and reading comprehension depends on domain-specific factors (see also Cornoldi & Vecchi, 2003). Consistently with this idea, it has been shown that verbal WM tasks (e.g., Reading Span Test by Daneman & Carpenter, 1980) are related to reading comprehension tasks since both draw on processes of a specific domain (the verbal domain), since visuo-spatial WM tasks are only moderately correlated to reading comprehension (Daneman & Tardif, 1987). However, other data have suggested that the involvement of WM in reading comprehension is domain-general, i.e. dependent on the attentional/executive control component of WM tasks (e.g., Engle, Kane & Tuholski, 1999; Turner & Engle, 1989). In fact, it has been reported that tasks requiring both maintenance and manipulation of information, or tasks associated to executive functions, correlate more strongly with reading comprehension – regardless of task modality (e.g. Daneman & Merikle, 1996 meta-analysis) – than storage-only tasks (e.g. forward digit span tasks). This relationship is explained by the fact that in understanding a text a reader is engaged not only in maintaining some piece of information but also, for example, in merging that information with previous knowledge, therefore actively processing incoming information.

Studies comparing good and poor comprehenders (i.e. readers with adequate decoding and intellectual abilities) may clarify the relationship between working memory and reading comprehension, establishing which mechanisms might become particularly crucial for the case of substantial reading comprehension difficulties. However, this controversy also arises in studies taking an individual-differences viewpoint (for a review see Savage, Lavers, & Pillay, 2007). Indeed, studies have shown contrasting results on whether or not a deficit in WM is present in poor comprehenders regardless of task modality: some report significant differences between good and poor comprehenders in visuo-spatial WM tasks (e.g. Cornoldi, De Beni, & Pazzaglia, 1996), albeit more modest than in verbal WM tasks, but others find no difference (e.g. Nation, Adams, Bowyer-Crane, & Snowling, 1999). To explain the nature of reading comprehension disabilities, the existence of differences also in visuo-spatial WM task performance would support a view where domain-general aspects of WM are specifically involved in higher-order cognition (Swanson & Siegel, 2001), whereas the opposite result would suggest that the lower performance of poor comprehenders is not completely independent of the nature of WM task (see for example Cornoldi & Vecchi, 2003).

To sum up, different approaches to WM result in different predictions about the nature of WM deficits in individuals who have...
Table 1
List of studies included with details about the selection reading comprehension task, reading decoding task, age, gender and N of participants, separated into good and poor comprehenders.

<table>
<thead>
<tr>
<th>Age Good comprehenders</th>
<th>Age Poor comprehenders</th>
<th>Reading comprehension</th>
<th>Reading decoding</th>
<th>Working memory task performances considered for the meta-analysis</th>
<th>Executive functions mechanisms</th>
</tr>
</thead>
</table>

Total 954 659

Note. LST: Listening Span Test; RST: Reading Span Test; CWMS: Categorization Working Memory Span Test.
specific reading comprehension difficulties but adequate intellectual functioning and level of reading (decoding) skills. This controversy raises serious problems in the field, not only in studying the relationship between WM and reading comprehension, but also as regards implementation of the most suitable procedures for assessment and intervention. Reading comprehension difficulties certainly have marked effects on school achievement (Taraban, Rynearson, & Kerr, 2000; Meneghetti, Carretti, & De Beni, 2006) and day-to-day activities. From a practical point of view, therefore, great benefit would derive from determining which aspects should be explored during clinical evaluation, with a view to devising an empowerment programme of activities for poor comprehenders.

The aim of this meta-analysis is to clarify the specificity of the role of WM in reading comprehension difficulties of individuals (children or adults) with normal decoding skills and intellectual abilities. The strength of the differences in reading comprehension performance between good and poor comprehenders in WM tasks is tested using the classic effect-size index (Cohen, 1988), which refers to the degree of association/correlation between two or more variables.

In line with a domain-specific view of WM, it is possible to predict that, if the relationship between WM and reading comprehension performance is mediated by task modality and attentional control, then verbal complex memory tasks should better discriminate between poor and good comprehenders than verbal simple memory tasks and visuo-spatial tasks. In contrast, a domain-general view of WM should predict that WM tasks, regardless of task modality, should better capture the differences between poor and good comprehenders in comparison to less demanding tasks in terms of attentional resources (e.g. short-term memory tasks). From an individual differences point of view, this would mean that the poorer WM performance of poor comprehenders depends little on task modality, and instead is mainly related to the attentional control involved.

2. Method

We carried out a search of the published literature to identify studies where WM was assessed in individuals (children ages 8–14 and young adults ages 18–30) with reading comprehension difficulties. Medline, Web of Science, ERIC and PsycINFO were searched from August 1980 to September 2006 (i.e. after introduction of Daneman and Carpenter’s Reading Span Test) using combinations of specific terms such as reading comprehension difficulties and disabilities, poor comprehenders, with the following keywords: WM, verbal span, visuo-spatial span, short-term memory, phonological loop, visual-spatial sketchpad, digit span. The lists of articles retrieved were also examined for further relevant publications. We limited our search to studies in the English language in peer-reviewed journals and in published books (Table 1 list the studies and administered tasks included in the meta-analysis).

3. Study selection

Several inclusion and exclusion criteria were used for identifying relevant articles. The criteria mainly refer to participant selection. Articles selected had to:

(1) include a group of poor comprehenders with reading comprehension learning disabilities with a normal-range level of general cognitive abilities (usually associated with level of IQ); specific impairment in the ability to understand the meaning of a text, not anticipated from the participant’s level of general cognitive abilities; reading comprehension difficulties not primarily due to hearing and/or visual problems, socio-economic factors, cultural or linguistic differences, lack of motivation, or ineffective teaching (see Cain & Oakhill, 2004 or Cornoldi & Oakhill, 1996);

(2) provide comparison with groups of normally developing children or children with reading comprehension score above expected age average score. Good and poor comprehenders had to be matched on measures of general cognitive functioning and, in particular for children, on measures of reading decoding;

These two above criteria allowed exclusion of participants from the poor comprehension group whose weak comprehension ability was due to low cognitive profile or word reading (decoding) difficulties.

(3) select good and poor comprehenders on the basis of a standardized reading comprehension task. The reading comprehension performance of the poor comprehenders group had to be two standard deviations below expected age average score or alternatively 12 months below their chronological age.

This procedure allowed us to find 18 articles; these are listed in Table 1. Despite the low number of studies found, we went ahead to

Table 2
Commonly used verbal complex span tasks.

<table>
<thead>
<tr>
<th>Material</th>
<th>Task requests</th>
<th>Storage request</th>
</tr>
</thead>
<tbody>
<tr>
<td>sentences, e.g. there are 18 h in a day; sea water is full of salt</td>
<td>true/false judgment of each sentence and memorization of thes last word of each sentence</td>
<td>final words (e.g. day, salt)</td>
</tr>
<tr>
<td>series of digits, e.g. 4 7 5; 9 3 2</td>
<td>reading aloud of digits and memorization of the last digit of each series</td>
<td>final digits (e.g. 5 2)</td>
</tr>
<tr>
<td>sequences of words, e.g. house, mother, dog, word, night; money, bull, minister, duck, strawberry</td>
<td>tapping whenever an animal noun appears and memorization of the last word of each sequence</td>
<td>final words (e.g. night, strawberry)</td>
</tr>
<tr>
<td>sequences of words, e.g. shirt, saw, trousers, hammer, shoes, nail</td>
<td>listening to the words and memorization of them in alphabetical order or memorization of the words that “go together”</td>
<td>ordered words (e.g. hammer, nail, trousers, saw, shirt) or grouped words (e.g. shoes, shirt, trousers);</td>
</tr>
</tbody>
</table>

Note. WM: working memory.

Table 3
Commonly used visuo-spatial complex span tasks.

<table>
<thead>
<tr>
<th>Material</th>
<th>Task request</th>
<th>Storage request</th>
</tr>
</thead>
<tbody>
<tr>
<td>series of dot arrays presented on a page</td>
<td>counting of the dots and memorization of the results of counting</td>
<td>numbers of dots</td>
</tr>
<tr>
<td>street map with lines connected to a number of dots/positions (pathway)</td>
<td>decision whether the presented dots/positions fall into a particular pattern and memorization of the pathway</td>
<td>positions of dots (pathway)</td>
</tr>
<tr>
<td>series of dots in a matrix</td>
<td>decision whether the dots presented fall in a particular position and memorization of the position of the dots</td>
<td>positions of dots</td>
</tr>
<tr>
<td>series of overlapping geometrical figures</td>
<td>task and storage requests: find where figures intersect maintaining the visual images of the figures</td>
<td></td>
</tr>
</tbody>
</table>

Figure intersection test (FIT)
calculate effect-size indexes, since no particular concerns emerge in the meta-analysis literature about sample size for a meta-analysis (Higgins & Green, 2006).

3.1. Categorization of WM measures

Although there are a variety of theoretical views of WM (Miyake & Shah, 1999), most of the models agree that WM tasks can be distinguished according to modality (verbal vs visuo-spatial) and attentional control (only storage, simple span task, vs tasks requiring storage plus processing, complex span tasks).

In this vein, simple span tasks (forward digit or word span) were considered as tasks requiring only storage, since participants are simply required to reproduce a sequence of items without any further manipulation. The simple span tasks therefore only involved storage of the material (verbal or visuo-spatial). As storage plus processing tasks (complex span tasks), we considered all the variations of the Daneman and Carpenter’s Reading or Listening Span Test that required the processing of some piece of information (reversing the order of item presented, or semantically judging sentences) and simultaneous memorization of the last word of each set (maintenance phase).

Complex span tasks were then distinguished according to task modality: verbal and visuo-spatial tasks (see Tables 2 and 3).

Furthermore, two WM measures involving executive functions were considered: updating, assessed through correct recall in WM updating tasks, and inhibition, with intrusion errors in WM tasks.

In the WM updating tasks, participants are required to remember a variable number of items following a criterion, for example the last word of each set (maintenance phase). Updating (Morris & Jones’ version, 1990) and Updating (semantic criterion) have been shown to be equivalent to a medium effect size, suggesting that group means differ of .5 indicated that the means differed by half a standard deviation, or dynamically change memory content.

Another measure often associated to poor comprehension performance is the intrusion error, i.e. the recall of no-longer-relevant information in WM tasks (Carretti, Cornoldi, De Beni, & Palladino, 2004). The occurrence of intrusion errors is considered an index of the inefficiency of inhibitory mechanisms (De Beni, Palladino, Pazzaglia, & Cornoldi, 1998). If poor comprehender WM deficit is related to difficulty in controlling the permanence of irrelevant or no longer useful information in WM (see Gernsbacher, 1993) – which can saturate WM capacity – then the difference with good comprehenders should also be large for the intrusion errors measure.

Both updating and intrusion errors were considered since they refer to ability to manage attentional control resources, maintain relevant information in an active state, suppress irrelevant information, or dynamically change memory content.

4. Analyses

The studies are numbered in Table 1 and summarized with reference to the tasks and measures used.

In particular, in calculating the effect size for the simple span task we considered Studies 1, 3, 7, 8, 13, 16 and 17 (see Table 1); for verbal complex span tasks, Studies 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17 and 18; and visuo-spatial complex span tasks, Studies 1, 3, 6, 10 and 16; for the high attentional controlled tasks, Studies 5, 13 and 17; and finally for intrusion error measure, Studies 1, 4, 5, 6, 7, 8 and 13.

In the 18 studies selected, participant ages ranged from 8 to 30; in order not to confound developmental issues with reading comprehension difficulties (Cain & Oakhill, 2006), two separate meta-analyses were run, one for children (8–15), the other for young adults (18–30). Two studies (Floyd, Bergeron, & Alfonso, 2006; Swanson, Howard, & Sáez, 2006) with participants aged from 8 to 17/18 were excluded.

To establish the magnitude of the differences between poor and good comprehenders, the classic effect-size index ($d$) proposed by Cohen (1988) was calculated: this expresses “the degree to which the phenomenon is present in the population” (p. 9, Cohen, 1988). It is obtained either by subtracting experimental (poor comprehenders) from control group (good comprehenders) mean scores and dividing by the pooled (average) standard deviation ($\sigma$), or from tests of the significance of differences in performance between groups (e.g. $t$ or $F$, $\chi^2$ tests, exact $p$ value). This formula therefore allows evaluation of the overlap between the group means by expressing score distances in units of variability.

The magnitude of effect sizes was interpreted according to Cohen’s (1988) guidelines ($d = .20$ small; $d = .50$ medium; $d = .80$ large). The value $d$ described the mean standardized difference in WM performance between poor and good comprehenders. For example, a $d$ value of .5 indicated that the means differed by half a standard deviation, equivalent to a medium effect size, suggesting that group means differ

#### Table 4
WM memory updating tasks.

| Updating (Morris & Jones’ version, 1990) | Material: Sequences of words of unknown length, e.g. oven, milk, cathedral, breath, bath, pumpkin, juice, box, moustache | Task request: recall of the last four words | Storage request: final words (e.g. pumpkin, juice, box, moustache) |
| Updating (semantic criterion) | Material: Sequences of words, e.g. meeting, sense, woodpecker, passion, law, cow, happiness, amount, caterpillar, lamb, feast, frog | Task request: recall of the three smallest items | Storage request: ordered smallest items (e.g. woodpecker, caterpillar, frog) |

#### Table 5
Summary of indices obtained in the meta-analysis (the studies used for each index are mentioned in Table 1 and in the text).

<table>
<thead>
<tr>
<th>Memory task</th>
<th>Number of participants</th>
<th>$d$</th>
<th>95% CI</th>
<th>$r$</th>
<th>$t^2$</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Good comprehenders</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple span tasks</td>
<td>109</td>
<td>.29</td>
<td>.10-.47</td>
<td>.14</td>
<td>0%</td>
<td>0-51</td>
</tr>
<tr>
<td>Complex span tasks</td>
<td>718</td>
<td>.75</td>
<td>.64-.85</td>
<td>.34</td>
<td>40%</td>
<td>0-60</td>
</tr>
<tr>
<td>Visuo-spatial</td>
<td>447</td>
<td>.36</td>
<td>.19-.51</td>
<td>.18</td>
<td>0%</td>
<td>0-54</td>
</tr>
<tr>
<td>Executive functions mechanisms</td>
<td>235</td>
<td>1.07</td>
<td>.66-.87</td>
<td>.47</td>
<td>71%</td>
<td>14-85</td>
</tr>
<tr>
<td>WM updating measure</td>
<td>163</td>
<td>.91</td>
<td>.67-.59</td>
<td>.41</td>
<td>70%</td>
<td>30-82</td>
</tr>
<tr>
<td>Intrusion errors</td>
<td>87</td>
<td>.89</td>
<td>.64-.115</td>
<td>.41</td>
<td>17%</td>
<td>0-63</td>
</tr>
<tr>
<td>Children</td>
<td>517</td>
<td>.73</td>
<td>.60-.87</td>
<td>.32</td>
<td>48%</td>
<td>0-69</td>
</tr>
</tbody>
</table>

* In view of the high value of the heterogeneity index, the value of $d$ was calculated using a random effect analysis (see DerSimonian & Laird, 1986). In all other cases a fixed effect model was used (see Hedges & Olkin, 1985).
to some extent. From a correlational viewpoint, a higher level of $d$

coresponds to a higher degree of association between the variables

considered. To perform a statistical analysis of the strength of the
difference, the $d$ was transformed into the correlation index ($r$) and
then compared, to understand whether or not they might be
considered significantly different.

Furthermore, for each meta-analysis the $I^2$ statistic and 95%

confidence intervals were calculated according to Higgins and
Thompson (2002). The index is intended to describe the amount

of total variation across studies due to heterogeneity rather than chance.
The range of $I^2$ values lies between 0% (negative values are set to 0)

and 100%, with a value of 0% indicating no observed heterogeneity,

and larger values suggesting very little consistency in effect sizes

across the studies. As suggested by Higgins and Green (2006), the

value of $d$ was adjusted using random or fixed effect models after

considering the degree of heterogeneity.

5. Results

As can be seen in Table 5, the magnitude of $d$ varies as a function of
two characteristics of WM task: the modality (verbal vs visuo-spatial),
and attentional control involved (storage vs storage/processing).

Higher values are associated with tasks with a verbal domain
requiring both maintenance and manipulation of information. Indeed, $d$
values for verbal WM tasks can be considered to lie in the medium
range, while for WM updating and intrusion errors (measures of
executive functions), the measure was high (see Cohen, 1988,
grand, while for WM updating and intrusion errors (measures of
executive functions), the measure was high (see Cohen, 1988,
grand, while for WM updating and intrusion errors (measures of
executive functions), the measure was high (see Cohen, 1988,
grand, while for WM updating and intrusion errors (measures of
executive functions), the measure was high (see Cohen, 1988,
grand, while for WM updating and intrusion errors (measures of
executive functions), the measure was high (see Cohen, 1988,
grand, while for WM updating and intrusion errors (measures of
executive functions), the measure was high (see Cohen, 1988,
grand, while for WM updating and intrusion errors (measures of
executive functions), the measure was high (see Cohen, 1988,
grand, while for WM updating and intrusion errors (measures of
executive functions), the measure was high (see Cohen, 1988,
grand, while for WM updating and intrusion errors (measures of
executive functions), the measure was high (see Cohen, 1988,
grand, while for WM updating and intrusion errors (measures of
executive functions), the measure was high (see Cohen, 1988,
grand, while for WM updating and intrusion errors (measures of
executive functions), the measure was high (see Cohen, 1988,

6. Discussion and conclusion

The main goal of the current study was to clarify the role of WM
measures in distinguishing between performance of poor and good
comprehenders, by analyzing two possible factors that can account for
their differences: the modality of WM tasks (verbal vs visuo-spatial),
and attentional control involved (storage vs storage/processing)
during reading comprehension.

Examining the $d$ values (see Table 5), it is clear that the association
between WM measures and reading comprehension abilities varies as
a function of the modality and attentional control required. It should
be noted that poor comprehenders are more disadvantaged in
complex span tasks than good comprehenders only when tasks
involve verbal material. In contrast, the performance of poor
comprehenders is comparable to that of good comprehenders when
considering visuo-spatial complex span tasks and simple span tasks.
It therefore seems plausible that the poor WM performance of poor
comprehenders depends partially on WM modality, but it may also be
caused by failure in attentional control component of WM. Consis-
tently with this view, the difference in performance between poor and
good comprehenders in updating and inhibitory (intrusion errors)
executive functions mechanisms proved high.

Overall our findings confirm that WM tasks that involve a high
demand in terms of attentional resources are a better predictor of
reading comprehension performance than measures of simple span
tasks (see Daneman & Merikle, 1996). The results thus appear to be
consistent with non-unitary models of WM (see for example Baddeley
& Logie, 1999 or Cornoldi & Vecchi, 2003), suggesting that deficits in
reading comprehension by poor comprehenders can also be partly
attributed to inefficiencies in WM control mechanisms, which are
failing to support specifically the verbal processing (see for example
Swanson & Berninger, 1995). Substantial differences between groups
were indeed isolated to verbal complex span measures, whereas
weaker group differences between poor and good comprehenders
emerged on visual-spatial complex span measures.

However, some caution should be applied in interpreting the data,
since in some cases heterogeneity values are high, in particular in the
case of $d$ values for WM updating measures and intrusion errors. On the
one hand, this could depend on the paucity of studies considering
updating and inhibition in relation to reading comprehension difficul-
ties; but equally it could be due to the different procedures adopted,
for example, for measuring the updating process. Swanson et al. (2006)
and Palladino, Cornoldi, De Beni, and Pazzaglia (2001, Exp. 1), for instance,
used a variant of the classic Morris and Jones (1990) task in which
participants are required to recall the last four items of each of a set of
lists. However, some authors have pointed out that this classic memory
updating task does not necessarily imply an updating process, since
participants could adopt a more passive strategy by waiting until the end
of the presentation and then retrieving the correct items on the basis of
a recency criterion (Ruiz, Eloua, & Lechuga, 2005). These latter aspects
may contribute to the heterogeneity obtained.

It should be noted here that only published studies were included in
the meta-analysis. Ignoring unpublished studies, which tend to
reveal smaller effect sizes, may have biased our research by increasing
the chances of finding larger effect sizes (Rosenthal, 1991). Never-
theless, overall our findings should be considered as an attempt to
summarize available data.

Another issue to consider is that the majority of studies were
conducted by only a few research teams, exposing the results to a
range of biases. Nevertheless, the results were obtained with different
groups of subjects and are very consistent not only internally but also
with studies analyzing correlations within the random population
(e.g. Daneman & Merikle, 1996).

In conclusion, our study shows that the WM deficit of poor
comprehenders relates mainly to those tasks requiring storage and
processing of information while inhibiting off-goal information and/or
updating memory content information (Carretti et al., 2005). However,
this is true only when considering verbal complex span tasks or tasks
involving executive functions. In fact, the performance of poor compre-
henders differs only marginally from that of good comprehenders
in visuo-spatial WM complex span or verbal simple span tasks. Consistently
with the common features of “working memory” (see Miyake & Shah,
we suggest that also in the case of individual differences in reading comprehension (poor vs good comprehenders) WM cannot be considered as completely unitary, since both domain-specific and attentional control factors place constraints on WM performance.

As regards to possible implications for the field of learning disabilities, our results point to WM as a possible important marker of reading comprehension difficulties (see Gathercole, Alloway, Willis, & Adams, 2006), suggesting the relevance of its evaluation for characterizing reading comprehension deficit profile. An important question arising is its potential clinical application in remedial activities designed to increase WM capacity and thus reading comprehension performance in poor comprehenders. Consequently, the crucial question is whether or not WM can be enhanced. Benefits from WM training have been examined, although to date only on adulthood (young and older adults), showing that WM performance can be improved (e.g. Carretti, Borella, & De Beni, 2007; Cavallini, Pagnin, & Vecchi, 2003; McNamara & Scott, 2001; Turley-Ames, & Whitfield, 2003). It is worth noting that most of these studies trained participants to use mnemonic strategies, such as imagery or semantic processing. These results therefore seem to indicate that WM enhancement might relate to a better use of strategies (see Ericsson, 1996), rather than a direct modification of WM structures.

Another key issue is the possibility of generalizing the effect of WM training to other related cognitive processes. In this case too, some encouraging results are present in the literature. For example, Jaeggi, Buschkuehl, Jonides, and Perrig (2008) demonstrated that WM training produces a transfer effect on fluid intelligence scores in young adults.

Overall, these data lead to the hypothesis that, also in the case of poor comprehenders, enhancement of WM could produce a transfer effect on specific components of reading comprehension ability. For example, some studies (e.g. Daneman & Carpenter, 1980; Hamm & Hasher, 1992) allow prediction that an increase in WM capacity could produce an improvement in the ability to maintain content-relevant information for construction of a mental model despite the presence of irrelevant material, or to connect different parts of a text. Addressing this crucial issue, future research should offer further demonstration of the central role of WM in learning disabilities, and thereby supporting the need to introduce WM evaluation in common clinical practice.

References


