Individual differences in processing temporarily ambiguous quantifiers

Sanne Tolboom

Master's thesis

MA Linguistics, Utrecht University

Supervisors: Dr. E. Kaan, University of Florida

Prof. dr. F.N.K. Wijnen, Utrecht University

Independent reader: Prof. dr. P.H.A. Coopmans, Utrecht University

September, 2012

Acknowledgements

I would like to thank Dr. Edith Kaan and the University of Florida for welcoming me as an exchange student. I am most grateful to her for providing me with the materials and unpublished results of the self-paced reading study that I replicated in this study, as well as for her guidance in designing and running the experiment described in this thesis.

I am indebted to Prof. dr. Frank Wijnen for his advice and support throughout my time as a student at Utrecht University.

I am grateful to Prof. dr. Peter Coopmans for taking time to serve as my independent reader.

Finally, I wish to thank three faculty members of the Linguistics program at the University of Florida: Dr. Lori Altmann for introducing me to the Shipley vocabulary test, Dr. Andrea Dallas for helping me prepare the self-paced reading task described in this thesis, and Dr. Wind Cowles for letting me use her version of the reading span task in preparation of administering a reading span task in the current study.

Abstract

In previous research, individual differences have been found in quantifier interpretation. When people are asked to complete sentences such as “Four flowers were put in the vase. Three…”, most people preferred a subset reading for the second sentence, i.e. a reading in which “three” refers to three of the four flowers that were mentioned previously. However, in studies on quantifier processing, only a subgroup of readers have shown effects that were associated with revision when the quantifier of the second sentence in the sequence was not compatible with the preferred subset reading (i.e. larger than the first). This subgroup of readers was characterized by their performance on comprehension statements that followed the sentence sequences in the quantifier processing study: Poor comprehenders did show a revision effect, while Good comprehenders did not.

We replicated a self-paced reading study of quantifier interpretation designed by Kaan, Alcocer, Barkley and Dallas (2007) and compared Poor and Good comprehenders on language and reading ability (as measured by the Shipley vocabulary task), working memory (as measured by an operation span and a reading span task) and attention (as measured by a situational motivation questionnaire). These measures were hypothesized to be factors in performance on the quantifier interpretation task by modulating reading comprehension and/or commitment to a particular reading of the sentence. Differences between Poor and Good comprehenders existed on tasks of working memory and language ability, but not on motivation. Unexpectedly, the best predictor of accuracy on the comprehension statements was the distracter task used in the operation span task. This distracter task involved the verification of simple math problems of the form “Is (8 : 2) – 3 = 1 ?”. However, performance on this equation verification task was not found to be related specifically to the interpretation of the quantities mentioned in the sentences, but was associated with general comprehension of the sentence sequences. This suggested that individual differences in quantifier interpretation were related to general cognitive capacities such as working memory and language skills.

Introduction

Since language perception is sequential, it is important to remember information given previously, while taking in new information. It is thought that we keep track of the entities, relationships between entities, and events that are described through a semantic representation. This representation of the discourse is updated whenever new information is provided by the speaker or in the text (Johnson-Laird, 1983; Van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). The semantic representation that is established also influences what the language comprehender will expect (Haviland & Clark, 1974). For example, in (1) one can guess that a new set of entities will be introduces in the second sentence.

1. Six apples were in the bag. Ten…

A possible continuation of the second sentence in (1) is “ten were used for a pie”, with “ten” referring to apples that were not mentioned yet. The term ‘quantifier’ denotes any word for an amount, such as “two”, “some” or “all”. In the ‘pie’ example a bare quantifier “ten” is used, i.e. a cardinal without an overtly expressed noun that specifies what it denotes. However, this quantifier cannot be interpreted on its own; it must be clear from the context what to take an amount of. In example (1), “ten” cannot refer to the set of entities mentioned previously, because ten cannot be a subset of six. However, in (2) the quantifier in the second sentence is ambiguous at this point. It could either refer to the known set of entities, or introduce a new set.

2. Six apples are in the bag. Five…

"Five" could refer to five of the six apples that are known in the discourse model, it could introduce five ‘new’ apples, or five entirely different entities (e.g. "boys", as in "Five boys rather had a pear"). Only when the referent is somehow made explicit, is the ambiguity of the bare quantifier removed.

Hendriks and De Hoop (2001) discuss a model that describes how expressions, including bare quantifiers, are interpreted. This model, based on Optimality Theory (Prince & Smolensky, 1997), consists of a set of constraints that are applied simultaneously to determine the optimal interpretation of a given utterance, assuming an, in principle, infinite number of possible interpretations for every utterance. The optimal interpretation is thought to be the preferred or "most unmarked" interpretation. One of the constraints Hendriks & De Hoop put forward is "Don't Overlook Anaphoric Possibilities" (DOAP). This constraint entails that entities that are already available in the discourse model must be chosen as the referent for the quantifier. As discussed, in (2) "five" could be taken to refer to the "apples in the bag" that have already been established in discourse. This interpretation is in line with DOAP by conforming to what Hendriks & De Hoop call the "Forward Directionality" constraint: "The topic range induced by the domain of quantification of a determiner (set A) is reduced to the topic range induced by the intersection of the two argument sets of this determiner (A ∩ B)" (p. 19). In the current example, this constraint comes down to taking "five" to denote a subset of the "apples in the bag" mentioned previously. These constraints are "soft" in the sense that they can be overruled by other "stronger" constraints. For example, in (1) "ten" cannot denote a subset of the "apples in the bag", because 10 is greater than 6. In this case the Forward Directionality constraint is outranked by a pragmatic constraint "Avoid Contradiction", which is stronger. However, in this case another constraint called "Parallelism" can be satisfied without violating Avoid Contradiction. Parallelism entails that a logically, structurally, or thematically parallel element from the preceding clause should be chosen as the referent of the quantifier (Hendriks & de Hoop, 2001 p. 26). Parallelism too is in accordance with DOAP. Under the Parallelism constraint, "ten" in example (1) can be taken to denote "apples" parallel to the "six apples" in the previous sentence. Note that although the Parallel interpretation is available in example (2) as well, the Forward Directional interpretation is preferred in this sentence sequence. The Forward Directionality constraint is stronger than the Parallelism constraint, which means the Forward Directional interpretation is 'optimal', hence preferred. This preference is confirmed in studies where participants were presented with a temporarily ambiguous quantifier like the one in (2). Most people preferred to complete the sentence using the Forward Directional reading, where the quantifier refers to a subset of the entities that have already been established in the discourse model (Frazier et al., 2005; Wijnen & Kaan, 2006). This is also in accordance with the suggestion that it is more difficult to process information that is not yet present in the discourse (Haviland & Clark, 1974; Murphy, 1984).

Kaan, Dallas and Barkley (2007) studied the process of quantifier interpretation in an ERP study. They compared sentences in which the Forward Directional interpretation is available

(3a), resulting in a subset interpretation of the second quantifier, to ones in which only the Parallel interpretation is available (3b), making it necessary to interpret the quantifier as introducing a new discourse referent. They call these conditions "Subset-compatible” (3a) and "New" (3b).

3.

a) Twelve flowers were put in the vase. Six had a broken stem and had to be cut very short.
b) Four flowers were put in the vase. Six had a broken stem and were trashed.

When the Forward Directional reading is preferred, one would expect signs of semantic integration difficulty at the quantifier “six” in the second sentence of (3b), where this reading turns out to be impossible. Furthermore, it is necessary to establish a new discourse referent when the second quantifier is larger than the quantifier in the first sentence. Therefore, extra processing effort is expected at the second quantifier of this sequence of sentences (3b), when compared to a sequence in which the second quantifier is smaller than the first, i.e. where a subset reading is possible (3a). Previous ERP studies that looked at difficulties of integration into a discourse model, have found a 300-600 ms negativity (Van Berkum, Brown & Hagoort, 1999), a Left Anterior Negativity (Anderson & Holcomb, 2005) and an N400 component (Streb et al. 1999, 2004). Kaan et al. found none of these early effects that indicate semantic integration difficulty. They did find a late positivity (900-1500 ms) at the second quantifier of the sentence sequence in the New condition compared to the Subset-compatible condition. A late positivity was reported previously in relation to the introduction of new discourse referents (Burkhardt, 2005), and retrieval and updating of discourse information (Van Petten, Kutas, Kluender, Mitchiner & McIsaac, 1991). Kaan et al. interpreted their finding as a Late Positive Complex that signals the establishment of a new referent. Findings of an eye-tracking study administered by Paterson and colleagues (as cited by Mousoulidou, 2009) seem to confirm this interpretation. They included an unambiguous control (4c) for the sentence in which the bare quantifier was ambiguous between a subset reading and a parallel reading (4a). In the control condition, the sentence was disambiguated by the phrase “of these”, which explicitly denotes a subset reading. The sentences in which the quantifier was incompatible with a subset reading (4b) started with “another” in the control condition (4d), explicitly denoting a parallel reading.

4.

a) The fishermen saw six ships appear on the horizon.
   Apparently, three ships had been bombarded by enemy fire.
b) The fishermen saw two ships appear on the horizon.
   Apparently, three ships had been bombarded by enemy fire.
c) The fishermen saw six ships appear on the horizon.
   Of these, three ships had been bombarded by enemy fire.
d) The fishermen saw two ships appear on the horizon.
   Another three ships had been bombarded by enemy fire.

They found that participants spent more time re-reading portions of text when the second sentence introduced a new referent (4b, d) compared to the sentence sets that were compatible with a subset reading. The sentences that were unambiguous at the quantifier (4c) yielded the same results as the sentences that were ambiguous (4a). Furthermore, the longer re-reading times were not specific to a particular portion of the sentence, but were found as a longer total reading time for the sentences in which a new referent was introduced. Since ambiguity of the quantifier was not found to evoke additional processing difficulty, Paterson et al. suggested that their findings indicate processing efforts of establishing a new referent in discourse.

The findings discussed here are compatible with Hendriks and De Hoop's (2001) model of interpretation. We have argued that the Forward Directional constraint is 'stronger' than the Parallelism constraint. The effects found by Kaan et al. and Paterson et al. could point to additional processing cost when the Forward Directional constraint is not met. However, it is not clear from these findings whether readers commit to one interpretation, and whether they can adjust when the interpretation they committed to turns out to be wrong at a later stage. In a self-paced reading experiment, Kaan, Alcocer, Barkley and Dallas (2007) studied this commitment of readers to an interpretation. In the sentences used in this study (5), the quantifier (underlined for clarity) was manipulated, as well as a critical word (in italics for clarity) that was presented at a later point in the sentence. The quantifier was either compatible with a subset interpretation (5a, c) or not (5b, d). Then the second sentence unfolded and the critical word was revealed to be either compatible with a subset reading (5a, b) or not (5c, d).

5.

a) Eight retirees had booked a trip to Europe. <u>Five</u> had chosen *France* for a week long tour. (<u>subset</u> – *subset*)
b) Four retirees had booked a trip to Europe. <u>Five</u> had chosen *Japan* for a week long tour. (<u>no subset</u> – *no subset*)
c) Eight retirees had booked a trip to Europe. <u>Five</u> had chosen *Japan* for a week long tour. (<u>subset</u> – #*no subset*)
d) Four retirees had booked a trip to Europe. <u>Five</u> had chosen *France* for a week long tour. (<u>no subset</u> – **subset*)

In terms of the constraints posed by Hendriks and De Hoop (2001), this resulted in one sentence with a preferred Forward Directional reading (a). Furthermore, there were two sentences with a Parallel reading (b, c) of which (b) was an 'early' parallel, and (c) a 'late' parallel, or 'garden path' sentence. The combination of the quantifier and the critical word in sentence (d) resulted in an anomaly. This experiment did not reveal processing difficulties at the quantifier or at the critical word in items that introduced a new referent (b, d) compared to items that were subset-compatible (a, c). Only at the end of the sentence were the anomalous (d) and garden path (c) conditions read more slowly than the correct sentences (a, b).

The three studies described here did not find early effects related to recovering from a commitment to the preferred subset interpretation, in the cases in which this interpretation was incorrect. This is especially surprising in the light of other studies that have found indications of a strong bias towards the subset-reading. A possible explanation is that the differences between studies were task induced. One of the studies that did find an early effect was an acceptability judgment task (Wijnen & Kaan, 2006), in which participants were asked to indicate at each word whether the sentence was acceptable up until that point. This might have induced a focus on interpreting and integrating individual words, rather than the sentence as a whole. In the ERP study, words were presented at a fixed rate. Although the relative speed of presentation compared to their 'natural' reading speed might differ between participants, the unusual way of reading might have caused the participants to use unusual reading strategies. In contrast, the self-paced reading task asks participants to move through the sentence at their own pace, possibly resulting in a more natural reading style than is allowed for in both the acceptability judgment and ERP studies. Incidentally, Tunstall (1998) in her dissertation used a self-paced reading task to study quantifier scope ambiguity. She did not find differences in reading times between the two possible readings of *a critic* in the sentences in (6), even though the plural reading (6b) is preferred by most people (Kurtzman & MacDonald, 1993).

6.

a) Kelly showed every photo to a critic last month. The *critic* was from a major gallery.
b) Kelly showed every photo to a critic last month. The *critics* were from a major gallery.

Kurtzman & Macdonald (1993) presented their participants with the first sentence of a sentence sequence like the ones in (6) and asked them to read this sentence at a fast pace. The participants were instructed to press a button when they finished reading, in order for the second sentence of the sequence to appear. As soon as they finished reading the second sentence at a fast pace, the participants indicated whether the second sentence made sense to them, as a continuation of the first sentence. Like the self-paced reading task, this task allowed participants to read in a fairly natural pace. However, similarly to the acceptability judgment task in the Wijnen & Kaan study (2006), participants were asked to indicate whether the sentence made sense or not. This assignment might have induced some form of evaluative reading. Therefore, a difference in reading style, or reading strategy, used by the participants still holds as a possible answer to the question why some studies have found indications of commitment to a subset reading, while other studies have not found evidence pointing in that direction.

We discussed how different tasks might have given rise to different reading strategies, possibly influencing the results of the study. Of course, differences might also exist between participants in the same study. Individual differences have been found in a number of ERP studies for example (e.g. Osterhout, 1997; Mecklinger, Schriefers, Steinhauer & Friederici, 1995; King & Kutas, 1995). King and Kutas (1995) were interested in investigating the ERP effects that were associated with comprehension of a text. They used object and subject relative sentences such as the ones in (7). These sentences are known to differ in syntactic complexity and working memory demands (e.g. Gibson, 1998).

7.

a) The reporter who harshly attacked the senator admitted the error.

b) The reporter who the senator harshly attacked admitted the error.

In previous studies, the sentence in which “the reporter” is the object of the relative clause (7b) is found to be more difficult than the subject relative sentence (7a). In the study by King and Kutas, these experimental sentences were accompanied by filler sentences in the form of simple declarative sentences. Furthermore, about 45% of the experimental sentences were followed by a comprehension probe in the form of a statement, as illustrated in (8).

8. The reporter attacked the senator.

Upon presentation of the statement, the participants were required to indicate whether it was true or false. Performance on these comprehension probes was used to determine whether a participant was a relatively good or poor comprehender. When object and subject relatives were compared, a difference in ERPs was found. Object relatives elicited a negativity at the verbs resembling a Left Anterior Negativity (LAN) when compared to the verbs of the subject relative sentences. The LAN has been associated with working memory load (Kluender & Kutas, 1993). In order to compare patterns of ERP components shown by participants that did correctly comprehend the sentence to patterns shown by participants that were less successful in comprehension of the sentence, they performed a median split on the basis of accuracy on the comprehension probes. King and Kutas found that the half of participants in the Good comprehension group showed a slow positive going wave at the multiword region “main verb phrase”, which in example (7) consists of the phrase “admitted the error”. This slow positivity is found in both the subject relatives and the declarative filler sentences, in comparison to the object relatives. On the basis of this finding, King and Kutas suggest that the slow positivity is a sign of ease of processing. Since the sentences differ mainly in terms of working memory demands (Gibson, 1998), a possible explanation for the difference between Good and Poor comprehenders in this study could lie in working memory capacity.

Kaan, Dallas and Barkley (2007) investigated whether any effects in their ERP study were obscured by individual differences. Following King and Kutas, they divided the participants in their ERP study into a group of Good comprehenders and a group of Poor comprehenders, by use of a median split based on the performance on comprehension questions that followed both experimental and filler sentences. These questions probed comprehension after 25% of the sentences (33% of the experimental items were probed, 17% of the filler items), in order to ensure that the participants were paying attention to the material. The comprehension question was either about the first sentence (9a), the second sentence (9b), or the relation between the first and second quantifier (9c).

9.

a) We picked eight tomatoes. Nine were not completely ripe, so we left them on the plant.
   Did we pick more than eight tomatoes?
b) Kim mixed five eggs in a bowl. Six were about to go bad so she did not use them.
   Was Kim afraid to use all the eggs?
c) Five lizards were missing their tails. Four had some scars down their back as well.
   Did all lizards with missing tails have scars?

Kaan et al. found that the Poor comprehenders showed an additional 500-700 ms positivity in the New condition. This resembled a P600 effect, which is thought to signal revision of a structure (Friederici, 2002). The Good comprehenders did not show this earlier effect in addition to the late positivity discussed earlier, which was found for both groups. This suggests that Good comprehenders committed to a particular reading at a later stage than Poor comprehenders.

In the self-paced reading study by Kaan, Alcocer, Barkley and Dallas (2007) the participants were divided in Good and Poor comprehenders as well. In this study comprehension statements were used following every sentence sequence. Most of these statements (**S**) probed both sentences of a sequence, as illustrated by examples (10a), (10b) and (10e), although some only probed either the first (10c) or the second sentence (10d). The information that was probed was either a temporal relation (10a), a location (10b), a quantity (10c), a description (10d), or a combination of a quantity and a description (10e). For a complete overview of the experimental items and their comprehension statements, the reader is referred to appendix E.

10.

a) Three classes in the course schedule were science classes.
   Five were advanced English and slotted for Monday mornings.
   S The science classes were on Monday morning.

b) Three birds in the tree were cardinals.
   Four were colorful canaries loudly singing a melodic song.
   S The canaries were in the tree.

c) Six wines at the tasting were Italian.
   Four were from Rome and known for their grapes.
   S Several wines were from Italy.

d) Six cups were filled with pink lemonade.
   Five were completely full because the server was careless.
   S The server carefully filled the cups with lemonade.

e) Eight people passed their driving test.
   Six did very poorly according to the examining official.
   S The official thought all of the people did poorly.

When Poor and Good comprehenders were compared, these groups were found to behave differently. When the second quantifier was subset incompatible (5b, d), only the Poor comprehenders had longer reading times at the quantifier. This can be seen as an indication that they had more trouble establishing a new referent than the Good comprehenders. However, it can also be interpreted as further evidence that Good comprehenders commit to an interpretation at a later stage.

In summary, ERP and self-paced reading studies on quantifier interpretation have not found clear effects of expectations or preferred readings. However, there do seem to be individual differences in the interpretation, and the timing of interpretation of quantifiers.

The question remains what factors explain the differences between Good and Poor comprehenders. Note that the groups are characterized by both a difference in accuracy on comprehension questions, and a difference in commitment to the subset-interpretation. The difference in commitment is indicated by a difference in reading times and ERP effects that are associated with revision, when a new discourse referent is introduced by the quantifier. As a possible explanation for these differences, we have discussed the idea that there may be task-induced differences, possibly related to manner of processing, i.e. a tendency to integrate individual words, and we have hypothesized that this might translate into intrapersonal processing differences in all tasks. In addition, we discussed the study by King and Kutas (1995), in which the differences between Good and Poor comprehenders were thought to be associated with working memory capacity. And finally, we suggested that general language and reading ability is an obvious candidate when differences in reading comprehension and processing of written language are concerned. We will discuss these possible explanations, (1) language and reading ability, (2) working memory capacity, and (3) depth or style of processing in more detail below.

Language and reading skills, obviously, might influence reading strategies. It is not clear however, in what way reading strategy would be influenced by reading and language skills. Better language and reading skills might include efficient updating of the discourse model. If this is true, we would expect participants with better language and reading skills to exhibit less processing effort as a result of establishing a new discourse referent.

The division between Poor and Good comprehenders in the ERP study by King and Kutas (1995) was largely based on assumptions of working memory load associated with the experimental items. Recall that the object relative sentences were more difficult to process, hence comprehend, and these sentences were thought to demand higher working memory capacity. Working memory as a result, might be related to accuracy on the comprehension questions in the quantifier interpretation studies as well. However, not only comprehension, but also commitment to an interpretation might be related to working memory. Studies on syntactic ambiguity indicated that working memory capacity had an effect on whether or not a reader committed to an interpretation. Working memory was tested by use of a reading span task. In a reading span task participants are asked to read series of unrelated sentences aloud, while memorizing the final word of each sentence of a series. Working memory capacity is expressed by the number of sentences in the longest series of which the participant correctly recalled all final words. High span readers, i.e. participants with high working memory capacity, are able to hold multiple interpretations simultaneously, while low span readers commit themselves to a particular reading at an early stage (e.g. MacDonald, Just & Carpenter, 1992). However, Friederici, Steinhauer, Mecklinger & Meyer (1998) reported the opposite effect. In their study they found ERP effects related to revision for high span readers, but not low span readers, at the word that disambiguated an ambiguous sentence.

Indications that retaining multiple interpretations is less effortful for high span readers also comes from an ERP study by St. George, Mannes and Hoffman (1997) on inference generation during reading of short texts (11). The average peaks of the N400 component in the ERPs for each word were compared. The generation of an inference was indicated by a reduced average N400 effect for the words in the explicit inference (11a, b; in italics for clarity) compared to the average N400 component for the same sentence when it was not an explicit inference (11c). The reduced N400 is taken to indicate a reduced effort of semantic integration when the statement was already inferred.

11. a) Pam set the dining room table.
She forgot about the turkey in the oven.
The guests were disappointed with the ruined meal.
It was too bad the turkey burned.
b) Pam set the dining room table.
She forgot about the turkey in the oven.
Pam was disappointed when the argumentative guests ruined the meal.
It was too bad the turkey burned.
c) Pam set the dining room table.
She put the turkey in the oven.
The guests were outside playing badminton.
It was too bad the turkey burned.

St. George et al. (1997) found that readers with high working memory capacity generated inferences not only when necessary (11a), but also when optional (11b), as was supported by the N400 being lower for the explicit inference in (11a) and (11b) versus the same statement in (11c). Readers with low working memory capacity on the other hand, only generated inferences when this was necessary for reading comprehension (11a). The average N400 for the explicit inference in (11b) did not differ from the components seen in (11c) in this group.

We have discussed the fact that retention of multiple syntactic and semantic representations has been found to be associated with working memory. When readers with high working memory capacity are indeed capable of retaining multiple interpretations, this would show in the self-paced reading task as a lack of effect when the quantifier is incompatible with a subset-reading, because revision of a commitment to an interpretation is not expected for this group. However, low working memory capacity readers, who are thought to be less capable of retaining multiple interpretations, would show signs of revision of a commitment in the self-paced reading task in the form of longer reading times, when the quantifier is incompatible with the preferred subset-reading.

We will now turn to a discussion of depth, or style, of processing. We have already suggested that slow, word-by-word reading, and/or evaluative reading might be related to interpreting and integrating individual words. In other words, we suggested that reading strategy and the intention of readers might play a role in processing style. Processing style might in turn influence the degree of commitment to an interpretation. Sanford and Sturt (2002) argue for the existence of shallow processing in language. They argue that each word in a sentence does not necessarily contribute its full meaning. Furthermore, they suggest that these word meanings are not always combined into higher-level phrase meanings. They hypothesize that shallow processing yields underspecified representation. This is a level of representation at which ambiguities remain unresolved. An underspecified representation might therefore surface as an apparent retention of multiple readings, or a lack of commitment to one reading of the sentence.

Sanford (2002) argues that words that are attended to are processed more fully, i.e. they contribute their full meaning and are integrated into a phrase meaning. Sanford illustrates this by use of the Moses Illusion (12a). It has been reported that many people state that (12a) is correct, even though it was Noah, not Moses who put animals on the ark (Van Oostendorp & Kok, 1995). Bredart and Modolo (as cited in Sanford, 2002) constructed two versions of the Moses Illusion (12a, b).

12. a) Moses put two of each kind of animal on the ark.

b) It was Moses who put two of each kind of animal on the ark.

The detection rate of the error is much higher in (12b) than in (12a). In (12b), the cleft construction puts focus on the proper name “Moses”. This might result in more attention being allocated to processing the name, which in turn results in processing the meaning of the word more fully. When the full meaning of the word “Moses” is retrieved, the error can be noticed more easily. Linguistic structures are not the only factor in determining the amount of attention allocated to a particular word or sentence. For instance, the total amount of attentional resources and the amount of attention control influence the allocation of attention. These factors are subject to inter- and intrapersonal variability. For example, sleep deprivation is related to attention deficits (Alhola, & Polo-Kantola, 2007), and Attention Deficit and Hyperactivity Disorder is thought to be a deficit of attention control (Sonuga-Barke et al., 2011). Another influence of attention allocation to a task might be motivation. Motivation has been found to influence performance on a task that manipulated depth of processing of word meanings (Graham & Golan, 1991). Motivation might be explained in terms of initiative to allocate attentional resources. For example, Hertel and Rude (1991) argue that poor performance on cognitive tasks in individuals suffering from depression is explained by a lack of initiative to allocate attentional resources. Brose, Schmiedek, Lövdén and Lindenberger (2012) show that cognitive performance in healthy individuals is linked to control of attention and task-motivation as well. They found that day-to-day variability in accuracy on a working memory task correlated with fluctuations in self-reports of negative affect, motivation and subjective control of attention. We can hypothesize that motivation indirectly influences the degree of commitment to an interpretation in the quantifier task, by influencing the allocation of attention. However, this influence can go in two opposite directions. On the one hand, a lack of motivation is related to a lack of initiative to allocate attentional resources. We have argued that this possibly results in shallow processing, yielding underspecified representations, which could in turn surface as a lack of commitment to an interpretation. This means that low-motivation results in less commitment to a particular interpretation. Therefore, in the self-paced reading study, participants with high-motivation would show longer reading times when the quantifier is incompatible with a subset-reading than participants with low-motivation, due to a revision of their commitment to the preferred interpretation. On the other hand, a lack of motivation has been linked to poor working memory performance. As discussed previously, low working memory capacity might make it more difficult to retain multiple interpretations. Under this reasoning, one would expect a reader with low-motivation to commit to one interpretation. Therefore, in the self-paced reading study, participants with low-motivation would show longer reading times when the quantifier is incompatible with a subset-reading than participants with high-motivation, due to a revision of their commitment to the preferred interpretation.

In conclusion, in processing quantifier ambiguities some readers have shown an effect when the quantifier was not compatible with the preferred subset reading, while other readers did not show an effect. In the current study we attempt to shed more light on the issue of the individual differences. We have repeated the self-paced reading study designed by Kaan, Alcocer, Barkley and Dallas (2007) and divided the participants into groups of Good and Poor comprehenders on the basis of performance on the comprehension statements, in the same way they did. Since performance on cognitive tasks was expected to be subject to day-to-day variation, all additional tasks had to be administered in the same session. Limiting the study to one session restricted the time available for each task, which meant that the proposed factors (language and reading ability, working memory, and attention) needed to be measured in a concise task that was easy to administer. We will proceed to discuss the methods used to measure these factors.

Language and reading ability was measured in part by the comprehension statements. The division of participants in Good and Poor comprehenders in itself suggests a difference in ability. Additionally, the vocabulary task from the Shipley Institute of Living Scale (Shipley, 1940) was used as a measure of reading and language ability. It is obvious that vocabulary and reading are closely related, since one needs to know the words in order to understand a text. At the same time reading will give opportunities to learn new word meanings (Beck, Perfetti & McKeown, 1982). Furthermore, any skill is thought to improve with practice. Vocabulary size and knowledge of infrequent words could therefore give an indication of the language and reading abilities of the participants.

Working memory was tested by conducting two different types of complex span tasks, an operation span and a reading span task. As discussed, in a reading span task participants are asked to read a set of sentences aloud and memorize the last word of each sentence. The operation span task also required the participants to memorize words, but instead of reading sentences, the interference task involved solving simple math problems. The operation span task used in this experiment can be downloaded from the website of the Engle lab (http://psychology.gatech.edu/renglelab/tasks.htm). These complex span tasks measure verbal working memory and have been shown to predict individual differences in cognitive function including reading comprehension (see Conway et al., 2005 for a review).

In the operation span task, the actual solving of the math problem might involve more capacities than verbal working memory, like number sense. Many animals are capable of perceiving exact numbers smaller than 4, while most animals are only capable of approximation of larger quantities (Cordes, Gelman, Gallistel & Whalen, 2001; Hauser, Carey & Hauser, 2000). In contrast, humans have a precise sense of numbers larger than four. Although error rate and reaction time in the detection of the number of dots in an array have been found to climb with increasing numbers (Mandler & Shebo, 1982; Trick & Pylyshyn, 1994). Obviously, having a precise sense of the numbers in the equations in the operation span task is a prerequisite for correctly determining the outcome of the equation. An fMRI study by Clark & Grossman (2007) shows that cortical areas (inferior parietal cortex) associated with number sense were involved in a quantifier comprehension task. This suggests that number sense supports quantifier comprehension. We can hypothesize that performance on the distracter task of the operation span task is related to quantifier interpretation. Furthermore, participants that are less efficient in calculation, might experience a higher level of interference of this task on the memorization of the words. This suggests that memory span as calculated by the operation span task might be a better predictor of performance on the quantifier task than the reading span task is.

Attention was discussed as one of the factors in depth of processing. As was the case for general language and reading ability, attention might already be measured in part by the comprehension statements following the experimental items. Therefore, the distinction between groups might already be based on differences in the amount of attention allocated to the task. As discussed, motivation is related to attention levels. Motivation was chosen as an indirect measure of attention in this study, because it can be measured for the experiment as a whole. Actual levels of attention might fluctuate during the session, while motivation for participating in the experiment, in comparison, is more constant. Therefore, motivation level at the moment of measuring might be indicative of levels during the rest of the session, while actual attention levels are more likely to diverge from the measured level of attention at one moment in time. Furthermore, motivation might influence intention of reading. This too was discussed previously as a factor in processing style. Participants with low motivation levels are less likely to read the sentence sequences in the quantifier study with the intention to perform well. Therefore, they might show more shallow processing than the participants who are highly motivated and hence, more likely to intend to perform well. The motivation that is experienced while a task or activity is executed is called “situational motivation” (Vallerand, 1997). Situational motivation can be measured through a questionnaire called the situational motivation scale, or SIMS (Guay, Vallerand & Blanchard, 2000). The theory that underlies the questionnaire is that motivation is an effect of perceived levels of self-determination. Self-determination is the feeling of being free to do what one chooses to (Deci & Ryan, 1985). The scale differentiates between four types, or levels, of motivation, in order of increased self-determination: amotivation, external regulation, identified regulation and intrinsic motivation. In this model, amotivation means that one does not see a valid reason for doing something. When externally regulated one does something because one feels like one has to do it. Identified regulation refers to the situation in which external reasons are internalized. For instance, one does something because one feels one will benefit in some way from doing it. Intrinsic motivation for doing something means that one performs the activity upon free choice, solely for the pleasure one derives from it. These four types of motivation can be joined in an overall motivation score called the self-determination index (SDI), which is derived from a weighted calculation of the individual measures of motivation (Vallerand & Ratelle, 2002).

Method

Participants

Forty-two University of Florida undergraduate and graduate students (13 male; age 18-25, M = 20.0 years) participated in the experiment for course credit or a small payment. All participants gave informed consent in accordance with University of Florida Institutional Review Board regulations. Two participants were excluded from analysis due to acquisition of a second language before the age of 12. Excluding these participants prevented any differences as a result of bilingualism or non-native English from obscuring differences between Poor and Good comprehenders on the factors under investigation in the current experiment. All remaining participants where native speakers of English and did not learn another language before the age of 12 according to self-report. The participants did not have reading or language disorders, or neurological deficits, according to the questionnaire they completed prior to the experimental tasks.

Individual difference measures

Shipley vocabulary task

In a computerized version of the Shipley vocabulary task, participants were presented with a word and were asked to select one of four words that most closely matched it in meaning.
For each trial, five words were printed in a row in the center of the monitor. The first word in the row was in capital letters, this was the word for which a synonym needed to be selected from the next four words. The task consisted of 40 items of increasing difficulty from "talk" to "temerity". The list of items can be found in appendix A.

Participants were instructed to take as much time as they needed to respond by pressing one of four buttons on the response box. The buttons were associated with the four potential synonyms, i.e. the button marked "1" for the first word, "2" for the second etc. The button press triggered the appearance of the next item.

Working memory

As discussed in the introduction, working memory was tested by use of two complex span tasks.

Operation span: A math problem followed by a question mark and a word was printed in the center of the screen. A complete list of item sets can be found in appendix B. All items were of the form :

13. IS (8 : 2) – 3 = 1 ? BEAR

Participants were asked to read the equation out loud ("Is eight divided by two minus three equal to one?"), and to verify whether the answer that was provided was correct or not by saying "Yes" or "No", and then read the word out loud as well. Participants were encouraged to take as much time as they needed to solve the math, and accurately verify the provided answer. The verification response was recorded by the experimenter. As soon as the word was read, the experimenter pressed the space bar to make the next item appear. Participants were only allowed to pause reading at the question mark – in order to solve the math – and were encouraged to “keep it going” throughout the rest of the item. After each set of items, which consisted of either two, three, four or five items, three question marks “? ? ?” appeared in the center of the screen. These served as a cue for the participants to write down the words of the current set they remembered on a response sheet. The response sheet provided five blanks in a row for each set. The participants were asked to write each word on the blank corresponding to the sequential position that the word was presented at. The task consisted of twelve sets (three of each set size). The set sizes were randomized, so the participants did not know how many items there were in the current set until they saw the question marks. Participants were given as much time as they needed to recall the words, but were not allowed to take a break in between sets. They received extensive instruction on the task and completed a practice trial consisting of three two-item sets before starting with the actual task. The experimenter reminded the participants of the instructions during the task when needed. Scores were taken for the total number of correct words recalled in their correct sequential position.

Reading span task: The reading span task used resembled the traditional Daneman & Carpenter (1980) reading span task. To fit the procedure of the operation span task, the order of the set sizes was randomized. A sentence was printed in the center of the screen and participants were asked to read the sentence out loud and memorize the last word of each sentence. For example:

14. The entire town arrived to see the appearance of the controversial political candidate.

The sentences were unrelated, and were 11 to 17 words in length (M = 14.0 words). For a complete list of items the reader is referred to appendix C. Like the operation span task, this task consisted of twelve sets of varying size (two, three, four or five items) which were presented in a randomized order. After each set three question marks “? ? ?” cued the participants to write down the words they remembered in their correct sequential position on a response sheet. Instead of using the common scoring procedure in which a participant’s score is based on the highest set size at which all words were correctly recalled, scores were taken for each individual word that was correctly recalled in its original sequential position. This scoring method, which was used for both the operation span and the reading span task, provided a wider variety of different scores than a score based on set size would. This increased the chance of any existing differences to become apparent.

Motivation

The SIMS questionnaire consisted of 16 statements that were possible answers to the question “Why are you currently engaged in this activity?”. Participants were told it was a questionnaire about “the reasons you might have for participating in this experiment”. They were asked to indicate whether the statements corresponded to their reasons using a seven point scale from "1: corresponds not at all" to "7: corresponds exactly". The central question: “Why are you currently engaged in this activity?” was printed in the top third of the computer screen. In the middle third a statement was printed. See appendix D for an overview of the statements. For example:

15. Because I think it’s fun.

16. There probably are good reasons to do this, but I don’t see any.

The lower third of the screen was dedicated to the seven point scale. Seven square boxes were printed with a number (1-7) in the center of each box. Participants used the mouse to select a number on the scale. The selected box turned blue upon clicking, after which the statement was immediately replaced by the next. Participants were allowed to take as much time as they needed to complete the task.

Quantifier interpretation tasks

Self-paced reading task

The self-paced reading task was exactly the same as the task used in Kaan, Alcocer, Barkley & Dallas (2007) with the exception of two item sets that turned out to be less clear examples of the four conditions in their study. These items were replaced. As discussed, each item set consisted of four sentences, where the quantifier in the first sentence could either be larger than the quantifier in the second sentence (a, c), or smaller (b, d) and the critical word was either compatible (a, d) or incompatible (b, c) with a subset reading. The material consisted of 32 sets of four sentence pairs, as illustrated by example (5). The 32 sets were divided into four groups of eight sets. Each group used the same combinations of numbers, which ensured that ratio and distance between numbers were matched. Furthermore, critical words were matched for word length and frequency across the four conditions. The first sentence of each sequence was of varying length, the second sentence always was nine words long. A complete list of experimental items can be found in appendix E. One item set (number 26 in appendix E) was excluded from analysis because the second sentence was only eight words long. Kaan et al. (2007) tested the material in an acceptability judgment study prior to their self-paced reading study. This revealed that the plausibility of the sentences was perceived as intended.

The 32 experimental items were distributed across four participant lists through a Latin Square design, making sure each participant only saw one version of each sentence set. As a result, each list consisted of a total of 32 experimental items, eight sentence sequences for each of the four conditions. In each list the experimental items were combined with 64 filler items, yielding a total of 96 trials. Of the fillers, 16 contained a number in the first and second sentence, 16 only had a number in the first sentence, and 32 sentences did not have any numbers. In half of the filler items the second sentence was an implausible continuation of the first sentence. The other half of the items were plausible. As the reader might recall from the discussion of Kaan et al. in the introduction, each trial was followed by a comprehension statement. A complete list of experimental items and fillers, including comprehension statements, can be found in appendices E and F, respectively.

The sentences were presented in a self-paced moving window format using E-Prime (PST software), using a non-cumulative display. At the beginning of each trial the participants were shown a rendition of the sentence sequence, in which each letter was replaced by a dash. Pressing "ENTER" on a response box revealed the first word of the sentence. By pressing again, the next word was presented, while at the same time the current word turned back into dashes. In this way, the participants could read the entire sentence sequence word by word, each time pressing a button to move on to the next word. The two sentences of each item were printed on separate lines. The second line started with two meaningless symbol sequences "#X#X X#X#" to prevent wrap up effects of the first sentence, and eye-movements from affecting reading times of the beginning of the second sentence. This means that the first word of the second sentence was presented in the third position on the second line. Reading times were recorded for each word, from the presentation of the word on the screen until the button press. After reading the last word of the item, the participant pressed “ENTER” again causing a screen with a statement to appear, probing comprehension of the sentence sequence. The participant indicated whether the statement was true or false by pressing “YES” or “NO” on the response box. After this button press, feedback was presented automatically in the form of a “Correct” or “Incorrect” message. The feedback was followed by a screen with the message “Press X for next”. The participants were instructed to take a break at this point if they needed one. Upon pressing “X” on the response box the next item appeared, again in the form of a rendition of the sentence sequence. Each participant saw five practice items before starting the experiment. Participants were instructed to read as naturally as possible.

Sentence completion task

A sentence completion task was administered. This allowed us to check whether the participants had a bias towards a particular reading of a sentence. The completion task took the form of a pen-and-paper assignment, consisting of ten items. The items were similar to the sentence sets used in the self-paced reading task, but the second sentence was cut off after the quantifier (17).

17. a) Five books were on the shelves. Three………………
    b) Eight players were in the stadium. Ten………………

In 50 percent of the sentences the second number was smaller than the first number (17a), in the other half the second number was larger (17b). Participants were instructed to complete the “ultra-short stories” and to write down the first ending of the sentence that came to mind.

Procedure

Upon arrival at the lab participants were asked to give informed consent, and to complete the language questionnaire. Next, they were seated in front of a computer screen in a dimly lit room. In this room they performed all experimental tasks, starting with the individual differences tasks. The participants first completed the Shipley vocabulary task, followed by the operation span task. After taking a mandatory break for three to ten minutes, participants were asked to complete the reading span task, followed by the motivation questionnaire. This part of the experiment took about 45 minutes to be completed.

The quantifier interpretation tasks were next, consisting of the self-paced reading task, followed by the pen-and-paper sentence completion task. The second part of the experiment took 30 to 45 minutes to complete.

Results

Quantifier interpretation tasks

<u>Self-paced reading task</u>

Reading times of the second sentence and the symbol sequence preceding it were analyzed. Reading times longer than 5000ms or shorter than 50ms were regarded as missing data. These constituted 0.5% of the data. After these outliers were taken out, mean reading times at each word position were calculated for each participant separately. Reading times longer than the mean plus 2.5 times the standard deviation were replaced by the mean plus 2.5 times the standard deviation. This affected 3.5% of the data.

Reading times

Reading times of all participants taken together were analyzed. Table 1 gives an overview of the results. A repeated measures ANOVA was carried out with Quantifier (subset versus no subset) and Critical Word (subset versus no subset) as factors. This analysis was done both by subjects (F_1) and by items (F_2), for each word position separately.

Table 1. – Reading times (mean; standard deviation in parentheses) at each word position as a function of condition (cf. example 5).

Condition	**Word Position**									
	Hash *X#X#X*	**Q** *Five*	**Q+1** *had*	**Q+2** *chosen*	**CW** *France*	**CW1** *for*	**CW2** *a*	**CW3** *week*	**CW4** *long*	**CW5** *trip.*
A: subset – subset	305 (135)	396 (209)	364 (196)	355 (168)	378 (176)	388 (182)	386 (177)	381 (164)	408 (228)	712 (543)
B: no subset – no subset	312 (179)	395 (239)	350 (136)	353 (167)	396 (218)	382 (167)	373 (176)	377 (151)	395 (167)	678 (523)
C: subset – no subset	295 (87)	374 (194)	352 (137)	356 (162)	376 (202)	396 (197)	379 (176)	376 (149)	380 (141)	685 (523)
D: no subset – subset	310 (106)	384 (196)	351 (139)	356 (189)	411 (248)	391 (176)	404 (208)	390 (159)	408 (241)	678 (508)

No differences between conditions were found at the quantifier (Q) and the following two positions. At the critical word (CW) an effect of Quantifier was found. When the quantifier was incompatible with the subset reading (b,d) this resulted in longer reading times at the critical word. However, this failed to reach significance in the by items analysis [$F_1(1,39) = 4.43$, $p = .042$; $F_2(1,30) = 2.79$, $p = .105$]. At the second word after the critical word (CW+2) there was an effect of Critical Word, where critical words that were compatible with a subset reading (5a, d) resulted in longer reading times. Again, this was not significant in the by items analysis [$F_1(1,39) = 8.16$, $p = .007$; $F_2(1,30) = 1.91$, $p = .178$]. The Critical Word effect can be explained by a Quantifier and Critical Word interaction at the same position, which was not significant in the by items analysis either [$F_1(1,39) = 4.33$, $p = .044$; $F_2(1,30) = .71$, $p = .407$]. Paired samples *t*-tests showed that the *no subset -subset* condition (5d) was read, on average, 31 milliseconds slower than the *no subset - no subset* condition (5b) [paired $t_1(39) = 3.50$, $p = .001$; $t_2(30) = -1.61$, $p = .118$]. No other significant differences were found at CW+2 between conditions [all p's > .13]. Reading times at the third word after the critical word (CW+3) did not show significant differences between conditions. An effect of Critical Word was found at the second to last word position (CW+4). This word was read slower when the critical word was compatible with a subset reading. This approached significance in the by items analysis only [$F_1(1,39) = .87$, $p = .36$; $F_2(1,30) = 3.44$, $p = .07$]. Finally, the sentence final word (CW+5) did not reveal any significant differences between conditions.

In summary, similarly to Kaan, Alcocer, Barkley & Dallas (2007), no significant effects were found when the participant group as a whole was taken into consideration.

Comprehension questions

One sentence was excluded from the analysis because the comprehension statement was ambiguous. The rest of the comprehension questions were correctly answered 91 percent of the time (fillers 90%, experimental conditions 93%). Paired samples *t*-tests showed that the four experimental conditions did not differ in terms of percentage of correct responses [all p's > .05; mean percentage correct condition A: 90% (SD 3.0%); B: 94% (SD 2.4%); C: 94% (SD 2.4%); D: 93% (SD 2.6%)].

The participants were divided into Poor and Good comprehenders by means of a median split of the comprehension scores on the experimental conditions. Fourteen participants obtained the median score (93%). These participants were further divided into good and poor on the basis of the comprehension scores including fillers. For a complete overview of the division into Poor and Good comprehenders on the basis of performance on the comprehension statements, the reader is referred to Table 2.

Table 2. – Percentage of correct answers to the comprehensions statements for both Poor and Good comprehenders.

Poor comprehenders			**Good comprehenders**		
	Percentage correct			**Percentage correct**	
Participant	**Experimental items**	**All items**	**Participant**	**Experimental items**	**All items**
42	77%	72%	*40*	93%	91%
37	77%	86%	*15*	93%	93%
21	80%	81%	*32*	93%	93%
36	80%	85%	*34*	93%	93%
11	80%	87%	*14*	93%	94%
4	83%	85%	*31*	93%	94%
8	83%	87%	*33*	93%	95%
6	83%	89%	*7*	93%	96%
23	87%	85%	*35*	93%	96%
39	87%	90%	*25*	97%	90%
13	87%	93%	*18*	97%	91%
3	90%	89%	*9*	97%	93%
2	90%	90%	*5*	97%	94%
19	90%	91%	*22*	97%	94%
30	90%	91%	*27*	97%	95%
16	93%	87%	*20*	97%	98%
10	93%	88%	*17*	100%	94%
24	93%	88%	*26*	100%	96%
29	93%	90%	*1*	100%	97%
12	93%	91%	*38*	100%	97%

In the analyses of correlation between comprehension score and the individual difference measures in this study, percentage correct on the experimental items will be used. Poor comprehenders had a mean score of 87% correct (SD 5.8 %) on the experimental conditions, while Good comprehenders had a mean score of 96 % (SD 2.6 %) on the experimental conditions.

Reading times Poor comprehenders versus Good comprehenders

The reading times of Poor comprehenders and Good comprehenders were analyzed separately (see Table 3. for an overview of reading times at each word position).

Table 3. – Reading times (mean; standard deviation in parentheses) for Poor and Good comprehenders at each word position as a function of condition (cf. example 5).

Condition	**Group**	**Word Position**									
		Hash *X#X#X*	**Q** *Five*	**Q+1** *had*	**Q+2** *chosen*	**CW** *France*	**CW1** *for*	**CW2** *a*	**CW3** *week*	**CW4** *long*	**CW5** *trip.*
A: subset – subset	**Poor**	300 (165)	375 (173)	336 (113)	334 (152)	375 (202)	378 (175)	355 (129)	354 (123)	390 (144)	669 (458)
	Good	310 (97)	416 (239)	391 (250)	376 (179)	381 (147)	398 (189)	417 (210)	407 (193)	426 (288)	754 (615)
B: no subset – no subset	**Poor**	294 (108)	374 (213)	338 (136)	331 (138)	347 (137)	383 (205)	362 (173)	369 (153)	363 (113)	605 (440)
	Good	331 (228)	416 (262)	363 (135)	374 (189)	445 (268)	380 (119)	384 (179)	384 (150)	427 (203)	751 (587)
C: subset – no subset	**Poor**	288 (87)	354 (168)	320 (103)	345 (177)	359 (187)	376 (178)	363 (167)	361 (143)	373 (134)	634 (437)
	Good	301 (87)	396 (216)	384 (158)	367 (146)	394 (216)	416 (214)	395 (185)	391 (154)	387 (148)	736 (595)
D: no subset – subset	**Poor**	300 (114)	359 (182)	337 (141)	360 (220)	393 (237)	370 (171)	400 (220)	385 (164)	383 (131)	637 (467)
	Good	319 (96)	409 (207)	364 (137)	352 (154)	428 (258)	411 (179)	408 (196)	396 (155)	433 (213)	717 (545)

The effects of Critical Word at positions CW+2 and CW+4 we discussed for all of the participants together were not significant either in the by items and by subjects analyses when each group was analyzed separately. The effect of Critical Word at CW+2 was significant only in the by subjects analysis of the Poor comprehenders [Poor comprehenders: $F_1(1,19) = 7.42$, $p = .013$; $F_2(1,30) = 2.26$, $p = .143$. Good comprehenders: $F_1(1,19) = 2.05$, $p = .169$; $F_2(1,30) = .97$, $p = .333$.]. The effect of Critical Word at position CW+4 was significant only in the by items analysis of the Good comprehenders [Poor comprehenders: $F_1(1,19) = .87$, $p = .362$; $F_2(1,30) = 1.07$, $p = .310$. Good comprehenders: $F_1(1,19) = 2.02$, $p = .171$; $F_2(1,30) = 9.10$, $p = .005$.]. However, the effect of Quantifier at the critical word (CW) that was significant in the by subjects analysis of all of the participants was found to be significant in both analyses for the Poor comprehenders [$F_1(1,19) = 5.88$, $p = .03$; $F_2(1,30) = 6.19$, $p = .02$]. When the quantifier was incompatible with a subset reading (i.e. larger than the first number), this resulted in longer reading times at the critical word. Good comprehenders did not show this effect [$F_1(1,19) = .02$, $p = .89$; $F_2(1,30) = .01$, $p = .94$]. However, Quantifier by Group interaction failed to reach significance in the by items analysis [$F_1(1,38) = 4.22$, $p = .05$; $F_2(1.60) = 2.63$, $p = .11$]. No other significant effects were found.

Sentence completion task

Answers on the sentence completion task were analyzed to find out whether Poor and Good comprehenders differed in their bias towards a particular interpretation for quantifiers that were or were not compatible with a subset reading. The type of completion was coded for each sentence. Possible types were parallel (as illustrated by 18a), subset (18b), non-anaphoric (18c), and superset (anomalous; 18d).

18. a) Five books were on the shelves. Three books were lying on the floor.
b) Five books were on the shelves. Three of the books were novels.
c) Five books were on the shelves. Eight magazines were on the shelves as well.
d) Five books were on the shelves. Eight books were on the top shelf.

When the second quantifier is compatible with a subset reading (18a,b), the group as a whole preferred the subset reading (18b; 53%) followed by the parallel reading (18a; 38%). Good comprehenders chose a subset reading in 54% of the cases and the parallel reading in 37%, Poor comprehenders had 51% subset readings and 38% parallel. None of the differences between groups were significant [all p's > .36].

When the second quantifier was subset incompatible (18c,d), the participants most often used a parallel reading (18a) to complete the sentence (53% of the number completions across participants), followed by a non-anaphoric reading (18c; 33%). Note that the non-anaphoric reading is not available in the self-paced reading task. Both Good and Poor comprehenders show the same pattern, 55% parallel, 35% non-anaphoric, and 51% parallel and 31% non-anaphoric completions respectively.

No differences between groups in bias towards an interpretation were found in the completion study. Possibly, the answer as to why the two groups behaved differently on the self-paced reading task can be discovered in the analysis of the individual differences measures, which will be discussed in the next section.

Individual differences measures

Performance of the group as a whole was analyzed for each of the tasks. In addition, performance of the Poor and the Good comprehenders were compared for each task. Furthermore, we tested for correlations between performance on the individual differences measures and performance on the comprehension questions. A complete overview for each group of mean scores on all tasks can be found in Table 4. The results for each task will be discussed in detail below.

Table 4. – Scores (mean; standard deviation in parentheses) for the participant group as a whole and Poor and Good comprehenders for performance on the comprehension statements for the experimental items, and the measures of individual difference described in this study: Shipley vocabulary test – percentage correct, Operation span – percentage of words correctly recalled, Operation span – percentage correct math verification, Reading span – percentage of words correctly recalled, SIMS – overall motivation score. Differences between Poor and Good comprehenders as described by *p* values of independent *t*-tests for each measure.

	All participants	**Poor comprehenders**	**Good comprehenders**	**Poor vs. Good**
Comprehension	91% (6.5%)	87% (5.8%)	96% (2.6%)	$p < .001$
Shipley	77% (9.5%)	73% (8.2%)	81% (8.8%)	$p = .003$
Op. span – recall	68% (14.0%)	62% (13.5%)	73% (12.9%)	$p = .021$
Op. span – math	97% (3.1%)	96% (3.4%)	97% (2.8%)	$p = .285$
Read. span - recall	63% (12.7%)	58% (10.5%)	68.6% (12.4%)	$p = .004$
SIMS – overall	10 (18.9)	6 (20.1)	14 (17.1)	$p = .183$

Shipley vocabulary test

Percentages of correct answers on the vocabulary test were calculated. The mean score was 77% (SD 9,5%) for the whole group, Good comprehenders scored 81% (SD 8.8%) correct, Poor comprehenders 73% (SD 8.2%). The difference between groups was significant [$t(38) = 3.21$ $p = .003$].

Furthermore, accuracy on the comprehension questions following the experimental items of the self-paced reading task correlated with percentage of correct answers on the Shipley task, $r = .33$, $p = .04$. The complete overview of correlations between measures can be found in Table 5.

Table 5. – Pearson *r* Correlations between performance on the comprehension statements and the individual difference tasks

	1	2	3	4	5
1. Comprehension					
2. Shipley	.33 *				
3. Op. Span - recall	.17	.40*			
4. Op. Span - math	.42**	.15	.15		
5. Read. Span - recall	.31	.36*	.64***	.06	
6. SIMS - overall	.22	.14	.48**	.14	.30

* p < .05. ** p < .01. *** p < .001.

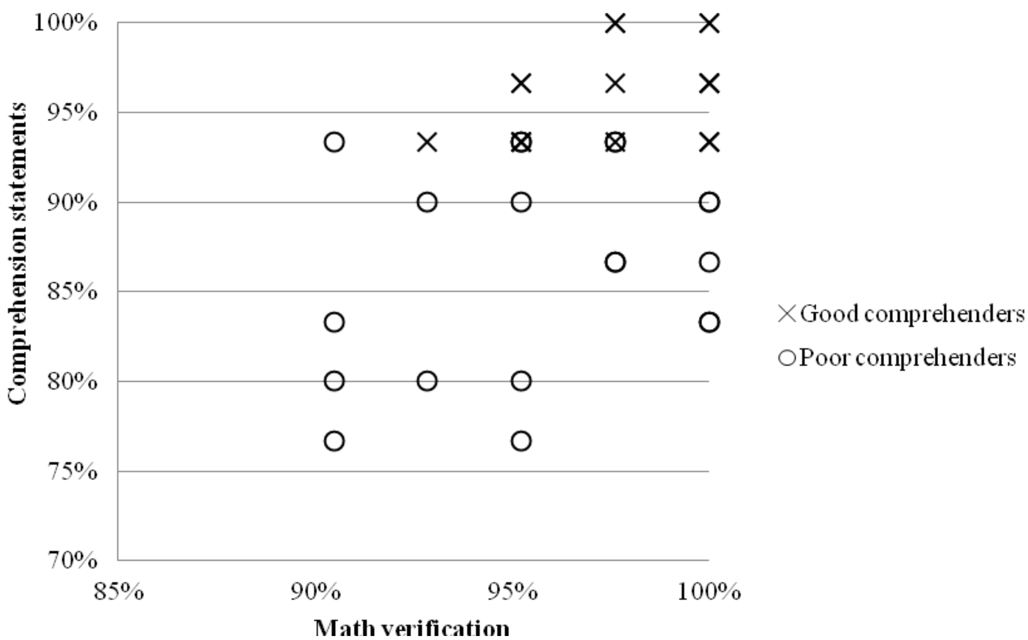
<u>Working memory</u>

Operation span task

The total number of words that were recalled in their correct sequential position were scored. One participant obtained a perfect score, but was excluded from the analysis because of subvocal rehearsal of the memorized words, resulting in prolonged pauses between verification of the equation and reading the word. The mean percentage of correctly recalled words was 68% (SD 14.0%). Mean score was 73% (SD 12.9%) for Good comprehenders and 62% (SD 13.5%) for the Poor comprehenders. This constitutes a significant difference [$t(37) = 2.41$, $p = .02$]. However, no significant correlation between performance on the operation span task and the comprehension questions was found, $r = .174$, $p = .29$.

The distractor task in the operation span task involved verifying simple math problems. The participant that was excluded from analysis of the recalled words, was included in this analysis, since performance on the distracter task was in accordance with instructions. The participants answered these verification questions correctly 97% (SD 3.1%) of the time. Poor comprehenders had a score of 96% (SD 3.4%), while Good comprehenders scored 97 % (SD 2.8%) correct on the math task. Accuracy on the verification of the mathematical equations did not significantly correlate with accuracy of word recall, r = .15, p = .36. There was no significant difference between Poor and Good comprehenders on the math task [t(38) = 1.085, p = .29]. However, performance on the math questions was significantly correlated with accuracy of responses to the comprehension statements of the experimental items, r = .42, p < .01. A scatter plot summarizes the results (Figure 1).

Figure 1. – Scatter plot of comprehension and math performance, by group.



The comprehension statements probe different types of information. To investigate whether the correlation between performance on the equations and the comprehension questions was directly related to interpretation of the quantifier, we looked at differences in accuracy between the different types of probes. No significant differences in comprehension between the information probes were found when the group as a whole was analyzed in a paired samples t-test [all p's > .09]. When each group of participants was analyzed separately, no differences in accuracy were found between probes either [all p's > .06]. A repeated measures ANOVA revealed that there was no significant Probe by Group interaction [F(4,152) = .66, p = .624]. Furthermore, performance on the equations significantly correlated with both performance on the questions that probed the quantifier, r = .41, p = .008, and with questions that probed the description of one of the referent sets, $r = .54$, $p < .001$. This suggests that performance on the math problems was not related to the interpretation of the quantifiers specifically, but rather to general comprehension performance.

Reading span task

The number of correctly recalled words in the correct sequential position was calculated. Mean performance for the whole group was 63% (SD 12.7%), Good comprehenders had a mean score of 69% (SD 12.4%), while the mean score of Poor comprehenders was 58% (SD 10.5%). The difference between groups was significant [$t(38) = 3.04$, $p = .004$]. The correlation between performance on the operation span task and accuracy on the comprehension questions was almost significant, $r = .31$, $p = .051$.

As was to be expected, a significant correlation existed between reading span and operation span scores, $r = .64$, $p < .001$.

<u>SIMS questionnaire</u>

For each type of motivation the sum of the ratings were taken. As a result of the 7 point scale, the score range for each motivation type was 7 to 28. Then, the overall score for motivation (self-determination index, SDI) was calculated following the procedure discussed by Vallerand and Ratelle (2002). The SDI was calculated by taking the sum of the score on the amotivation statements multiplied by -2, the score on external regulation multiplied by -1, identified regulation multiplied by 1, and intrinsic motivation multiplied by 2. This resulted in a score range for overall motivation from -72 to +72.

There were no differences between groups on any of the individual types of motivation [all p's $>.19$], nor on the SDI score [$t(38) = 1.36$, $p = .18$]. Furthermore, none of the motivation scores significantly correlated with accuracy on the comprehension questions [all p's $>.17$].

Amotivation was negatively correlated with both operation span, $r = -.49$, $p = .002$ and reading span results, $r = -.45$, $p = .003$. Furthermore, the overall motivation score correlated with the operation span, $r = .48$, $p = .002$. This suggests that a lack of motivation was related to lower working memory capacity.

Summary of the results

General reading and language ability, as measured by the Shipley vocabulary task, and working memory, as measured by the Operation span task and the Reading span task were found to be related to better performance on comprehension statements following the experimental items in the quantifier interpretation task. The math verification task that was used as the distracter task of the Operation span task was found to be a factor as well. Performance on the math task was associated with general comprehension, there was no relation to the interpretation of the quantifiers specifically. We have not found that differences in level of motivation were related to performance on the comprehension statements. Performance on working memory tasks, however, was related to levels of amotivation. A lack of motivation resulted in lower working memory capacity scores. Stepwise multiple regression analysis showed that scores on the math distracter task explained 15,6% of the variance in comprehension accuracy in the self-paced reading task [Model 1; adjusted $R^2 = .156$, $F(1,37) = 8.01$, $p = .007$]. When performance on the reading span task is added to the model, this results in a total of 26.9% of variance explained [Model 2; adjusted $R^2 = .228$ $F(1, 37) = 6.62$, $p =.004$]. Note, however that reading span scores did not quite significantly correlate with comprehension accuracy on the self paced reading study, $r = .31$, $p = .051$. The significant variables in the two models are shown in table 6.

Table 6. – Coefficients of the significant variables in the Stepwise Regression models

Predictor variable	**Model 1**		**Model 2**	
	Beta	***p***	**Beta**	***p***
Op. span – math	.422	.007	.417	.006
Read. Span – word recall			.301	.041

Discussion

In processing quantifier ambiguities some readers have shown an effect when the quantifier was incompatible with the preferred subset reading, while other readers did not show such an effect. This study aimed to shed light on the nature of these individual differences. We replicated the self-paced reading study designed by Kaan, Alcocer, Barkley and Dallas (2007) and added tasks that we argued would measure language and reading ability (Shipley vocabulary task), working memory (operation span and reading span task) and attention (SIMS motivation questionnaire). These measures were hypothesized to be factors in the quantifier interpretation task by modulating reading comprehension and/or commitment to a particular reading of a sentence.

This study replicated the previous findings that suggested individual differences in quantifier interpretation exist. Only poor comprehenders showed longer reading times at the critical word, when the quantifier was incompatible with a subset reading (b, d). This is in accordance with Kaan, Alcocer, Barkley & Dallas (2007), who found that only the Poor comprehenders had longer reading times at the quantifier when the second quantifier was subset incompatible (b, d). Even though the effect was visible at a later stage in the current study, it can be interpreted in the same way. Poor comprehenders might have more trouble establishing a new referent than Good comprehenders. Alternatively, Good comprehenders do not commit to an interpretation until after the critical word, while Poor comprehenders commit to a reading at an earlier stage causing the need to reanalyze at the critical word.

We have found that differences between Poor and Good comprehenders can be partially explained by general reading and language ability, and working memory. Reading comprehension has been linked to both working memory and reading and language ability, so this result was to be expected. However, the exact nature of the difference between Poor and Good comprehenders is difficult to pinpoint, because the vocabulary task is an indirect measure of reading and language ability. In addition, the complex span tasks used to measure working memory have been considered to be debatable in terms of construct-validity (Savage, Lavers & Pillay, 2007).

The best predictor of performance on the comprehension statements in the self-paced reading task was found to be performance on the math distracter task that was a part of the operation span task. This is surprising, since both Poor and Good comprehenders have high mean scores on the equation verification task that did not differ significantly (96% and 97% correct, respectively). As discussed in the introduction, it has been suggested that number sense is involved in quantifier interpretation (Clark & Grossman, 2007). However, performance on the math problems was not related to the interpretation of the quantifiers specifically, but rather to general comprehension performance. It is unclear why solving simple math problems is correlated to language comprehension. One possible explanation is that both performance on the equations and on the comprehension questions are related to attention. Situational motivation was included in this study as an indicator of attention. However, motivation did not significantly correlate with performance on the math verification task, and was not found to be a factor in explaining differences between Poor and Good comprehenders.

Motivation might not have been a good measure of attention. Ideally, one would measure attention levels at each sentence, in order to reliably relate this to performance on the comprehension statements. This might be possible with electroencephalogram (EEG) recordings during a quantifier interpretation task. The EEG waveforms alpha and beta are thought to be related to relaxation and active concentration, respectively. In order to test for attention levels, one might measure these waveforms during the self-paced reading task. In contrast to the motivation level tested in the current study, higher levels of attention as measured by EEG waveforms might be found to correlate with both higher scores on the comprehension statements and higher scores on the math verification task.

As discussed, the measures of reading and language ability in this study are only indicative of any differences in reading and language capacity. More extensive tests of reading and language are needed to confirm the findings in the current study. For instance, we have used the Shipley vocabulary task as a measure of reading ability. However, standardized reading fluency tests and reading comprehension tests are thought to be more direct measures of reading skill. The use of these measures could confirm the findings in the current study.

In conclusion, we suggest that differences between Poor and Good comprehenders are characterized by differences in reading comprehension, or a difference in commitment to an interpretation. These differences might be explained in terms of general cognitive capacities such as working memory, language ability and attention.

References

Alhola, P., & Polo-Kantola, P. (2007). Sleep deprivation: Impact on cognitive performance. *Neuropsychiatric Disease and Treatment*, 3, 553–567.

Anderson, J. E., & Holcomb, P. J. (2005). An electrophysiological investigation of the effects of coreference on word repetition and synonymy. *Brain & Language*, 94, 200-216.

Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, 74, 506–521.

Brose, A., Schmiedek, F., Lövdén, M., & Lindenberger, U. (2012). Daily variability in working memory is coupled with negative affect: The role of attention and motivation. *Emotion*, 12, 605–617.

Burkhardt, P. (2005). *The syntax-discourse interface: Representing and interpreting dependency*. Philadelphia: John Benjamin.

Clark, R., & Grossman, M. (2007). Number sense and quantifier interpretation. *Topoi*, 26, 51-62.

Conway, A. R. E., 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 & Review*, 12, 769-786.

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguishing verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin Review*, 8, 698-707.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, 19, 450-466.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

Frazier, L., Clifton, C. J., Rayner, K., Deevy, P., Koh, S., & Bader, M. (2005). Interface problems: Structural constraints on interpretation? *Journal of Psycholinguistic Research*, 34, 201-231.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Neuroscience,* 6, 78-84.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1-76.

Graham, S., & Golan, S. (1991). Motivational influences on cognition: Task involvement, ego involvement and depth of information processing. *Journal of Educational Psychology*, 83, 187-194.

Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). *Motivation and Emotion*, 24, 175-213.

Hauser, M. D., Carey, S., & Hauser, L. B. (2000). Spontaneous number representation in semi-free-ranging rhesus monkeys. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 267, 829-833.

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, 13, 512-521.

Hendriks, P. & de Hoop, H. (2001). Optimality theoretic semantics. *Linguistic Inquiry*, 7, 391-426.

Hertel, P. T., & Rude, S. S. (1991). Depressive deficits in memory: Focusing attention improves subsequent recall. *Journal of Experimental Psychology: General*, 120, 301–309.

Johnson-Laird, P. N. (1983). *Mental models: Toward a cognitive science of language, inference and consciousness*. Cambridge: Harvard University Press.

Kaan, E., Dallas, A. C., & Barkley, C. M. (2007). Processing bare quantifiers in discourse. *Brain Research*, 1146, 199-209.

Kaan, E., Alcocer, P., Barkley, C. M., & Dallas, A. C. (2007). *Semantic garden pathing: Subsets and new discourse referents* [Abstract]. Retrieved from http://crl.ucsd.edu/cuny2007/program/141_Abstract.pdf

King, J., & Kutas, M. (1995). Who did what when? Using word- and clause-related ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 7, 378-397.

Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, 5, 196-214.

Kurtzman, H. S., & MacDonald, M. C. (1993). Resolution of quantifier scope ambiguities. *Cognition*, 48, 243-279.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory contraints on the processing of syntactic ambiguity. *Cognitive Psychology*, 24, 56-98.

Mandler, G., & Shebo, B. J. (1982). Subitizing: an analysis of its component processes. *Journal of Experimental Psycholinguistics: General*, 111, 1-21.

Mecklinger, A., Schriefers, H., Steinhauer, K., & Friederici, A. D. (1995). Processing relative clauses varying on syntactic and semantic dimensions: an analysis with even-related potentials. *Memory & Cognition*, 23, 477-494.

Mousoulidou, M. P. (2009). *Referential processes in children's sentence comprehension: evidence from numerically quantified expressions* (PhD dissertation). Retrieved from: http://vfr.academia.edu/MarilenaMousoulidou/Papers/1115217/Referential_Processes_in_Childrens_Sentence_Comprehension_Evidence_from_Numerically_Quantified_Expressions

Murphy, G. (1984). Establishing and accessing referents in discourse. *Memory and cognition*, 12, 489–497.

Osterhout, L. (1997). On the brain response to syntactic anomalies: manipulations of word position and word class reveal individual differences. *Brain & Language*, 59, 494-522.

Prince A., & Smolensky, P. (1997). Optimality: From neural networks to universal grammar. *Science*, 275, 1604-1610.

Sanford, A. J. (2002). Context, attention and depth of processing during interpretation. *Mind & Language*, 17, 188-206.

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Sciences*, 6, 382-386.

Sonuga-Barke, E. J. S., Koerting, J., Smith, E., McCann, D. C., & Thompson, M. (2011). Early detection and intervention for attention-deficit/hyperactivity disorder. *Expert Review of Neurotherapeutics*, 11, 557-563.

Savage, R., Lavers, N. & Pillay, V. (2007). Working memory and reading difficulties: what we know and what we don't know about the relationship. *Educational Psychology Review*, 19, 185-221.

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *Journal of Psychology: Interdisciplinary and Applied*, 9, 371-377.

St. George, M., Mannes, S., & Hoffman, J. E. (1997). Individual differences in inference generation: An ERP analysis. *Journal of Cognitive Neuroscience*, 9, 776-787.

Streb, J., Rösler, F., & Henninghausen, E. (1999). Event-related responses to pronoun and proper name anaphors in parallel and nonparallel discourse structures. *Brain & Language*, 70, 273-286.

Streb, J., Henninghausen, E., & Rösler, F. (2004). Different anaphoric expressions are investigated with Event-related brain potentials. *Journal of Psycholinguistic Research*, 33, 175-201.

Trick, L., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited capacity preattentive stage in vision. *Psychological Review*, 101, 80-102.

Tunstall, S. L. (1998). *The interpretation of quantifiers: Semantics and processing* (PhD dissertation). Retrieved from: http://www.sfs.uni-tuebingen.de/~cebert/teaching/10Unterspezifikation/papers/The%20Interpretation%20of%20Quantifiers%20-%20Semantics%20and%20Processing%20(Tunstall%20Diss).pdf

Vallerand, R. J. (1997). Toward a hierarchical model of intrinsic and extrinsic motivation. In M. P. Zanna (Ed.), *Advances in experimental social psychology*, pp. 271-360. New York: Academic Press.

Vallerand, R. J., & Ratelle, C. F. (2002). Intrinsic and extrinsic motivation: A hierarchical model. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research*, pp. 37-63. Rochester, NY: The University of Rochester Press.

Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: evidence from event-related brain potentials. *Journal of Memory and Language*, 41, 147-182.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies in discourse comprehension*. New York: Academic Press.

Van Oostendorp, H., & Kok, I. (1990). Failing to notice errors in sentences. Language and Cognitive Processes, 5, 105–113.

Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractionating the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience*, 3, 131-150.

Wijnen, F. N. K., & Kaan, E. (2006). Dynamics of semantic processing: the interpretation of bare quantifiers. *Language and Cognitive Processes*, 21, 684-720.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123, 162-185.

Appendix A – Shipley vocabulary task

TALK	draw	eat	speak	sleep
PERMIT	allow	sew	cut	drive
PARDON	forgive	pound	divide	tell
COUCH	pin	eraser	sofa	glass
REMEMBER	swim	recall	number	defy
TUMBLE	drink	dress	fall	think
HIDEOUS	silvery	tilted	young	dreadful
CORDIAL	swift	muddy	leafy	hearty
EVIDENT	green	obvious	skeptical	afraid
IMPOSTER	conducto	officer	book	pretender
MERIT	deserve	distrust	fight	separate
FASCINATE	welcome	fix	stir	enchant
INDICATE	defy	excite	signify	bicker
IGNORANT	red	sharp	uninform	precise
FORTIFY	submerge	strengthe	vent	deaden
RENOWN	length	head	fame	loyalty
NARRATE	yield	buy	associate	tell
MASSIVE	bright	large	speedy	low
HILARITY	laughter	speed	grace	malice
SMIRCHED	stolen	pointed	remade	soiled
SQUANDER	tease	belittle	cut	waste
CAPTION	drum	ballast	heading	ape
FACILITATE	help	turn	strip	bewilder
JOCOSE	humorou	paltry	fervid	plain
APPRISE	reduce	screw	inform	delight
RUE	eat	lament	dominate	cure
DENIZEN	senator	inhabitan	fish	atom
DIVEST	disposses	intrude	rally	pledge
AMULET	charm	orphan	dingo	pond
INEXORABLE	untidy	inviolate	rigid	sparse
SERRATED	dried	notched	armed	blunt
LISSOM	moldy	loose	supple	convex
MOLLIFY	mitigate	direct	pertain	abuse
PLAGIARIZE	intent	revoke	maintain	misappro
ORIFICE	brush	hole	building	lute
QUERULOUS	maniacal	curious	devote	complain
PARIAH	outcast	priest	lentil	locker
ABET	waken	ensue	incite	placate
TEMERITY	rashness	timidity	desire	kindness
PRISTINE	vain	sound	first	level

Appendix B – Operation span task

IS (10 ÷ 2) - 3 = 2 ? SEA

IS (10 ÷ 10) - 1 = 2 ? CLASS

IS (7 ÷ 1) + 2 = 7 ? PAINT

IS (3 ÷ 1) - 2 = 3 ? CLOUD

IS (2 x 1) - 1 = 1 ? PIPE

IS (10 ÷ 1) + 3 = 13 ? EAR

IS (9 x 2) + 1 = 18 ? FLAME

IS (9 ÷ 1) - 7 = 4 ? BIKE

IS (8 x 4) - 2 = 32 ? BEAN

IS (9 x 3) - 3 = 24 ? ARM

IS (4 ÷ 1) + 1 = 4 ? GROUND

IS (10 ÷ 1) - 1 = 9 ? HOLE

IS (8 x 4) + 2 = 34 ? DAD

IS (6 x 3) + 2 = 17 ? KID

IS (6 ÷ 3) + 2 = 5 ? FORK

IS (6 x 2) - 3 = 10 ? JAIL

IS (8 ÷ 2) + 4 = 2 ? HAT

IS (8 ÷ 2) - 1 = 3 ? LAMP

IS (9 ÷ 1) - 5 = 4 ? CAVE

IS (6 ÷ 2) - 2 = 2 ? BACK

IS (7 x 2) - 1 = 14 ? HALL

IS (6 x 2) - 2 = 10 ? FERN

IS (2 x 2) + 1 = 4 ? MAN

IS (7 x 1) + 6 = 13 ? WORLD

IS (3 ÷ 1) + 3 = 6 ? DRILL

IS (10 ÷ 1) + 1 = 10 ? CALF

IS (4 x 4) + 1 = 17 ? FISH

IS (3 x 3) - 1 = 8 ? CHEEK

IS (3 x 1) + 2 = 2 ? BREAD

IS (4 ÷ 2) + 1 = 6 ? GERM

IS (5 ÷ 5) + 1 = 2 ? DOCK

IS (2 x 3) + 1 = 4 ? GAME

IS (9 ÷ 3) - 2 = 1 ? NERVE

IS (10 ÷ 2) - 4 = 3 ? WAX

IS (5 ÷ 1) + 4 = 9 ? TIN

IS (10 x 2) + 3 = 23 ? CHURCH

IS (7 ÷ 1) + 6 = 12 ? BEACH

IS (3 x 2) + 1 = 6 ? CARD

IS (6 x 4) + 1 = 25 ? JOB

IS (9 ÷ 3) - 1 = 2 ? CONE

IS (8 ÷ 1) - 6 = 4 ? BRASS

IS (9 x 1) + 9 = 1 ? STREET

Appendix C – Reading span task

It was shortly after this that an unusual pressure of business called me into town.
I imagine that you have a shrewd suspicion of the object of my early visit.
I turned my memories over at random like pictures in a photograph album.

Despite the unusually cold weather, the campers continued their canoe trip.
I should not be able to make any one understand how exciting it all was.
There was still more than an hour before breakfast, and the house was silent and asleep.
I imagined that he had been thinking things over while the secretary was with us.
The young business executive was determined to develop his housing projects within the year.

Filled with these dreary forebodings, I fearfully opened the heavy wooden door.
The taxi turned up Michigan Avenue, where they had a clear view of the lake.
I'm not certain what went wrong but I think it was my cruel and bad temper.

The weather was very unpredictable that summer so no one made plans too far in advance.
The entire town arrived to see the appearance of the controversial political candidate.

The stories all deal with a middle-aged protagonist who attempts to withdraw from society.
The lumbermen worked long hours in order to obtain the necessary amount of wood.
In comparison to his earlier works, the musician had developed a unique enthralling style.
The boisterous laughter of the children was disturbing to the aged in the building.
The sound of an approaching train woke him, and he started to his feet.

At the conclusion of the musicians' performance, the enthusiastic crowd applauded.
Without any hesitation, he plunged into the difficult mathematics assignment blindly.
The devastating effects of the flood were not fully realized until months later.
When I got to the big tobacco field I saw that it had not suffered much.

According to the results from the survey, Robert Redford is the most liked Hollywood star.
The lieutenant sat beside the man with the walkie-talkie and stared at the muddy ground.

The products of digital electronics will play an important role in your future.
The old lady talked to her new neighbor on her weekly walks from church.
One problem with this explanation is that there appears to be no defense against cheating.
To determine the effects of the medication, the doctor hospitalized his patient.

He pursued this theme, still pretending to seek for information to quiet his own doubts.
In order to postpone the business trip, he canceled his engagements for the week.
The courses are designed as much for professional engineers as for amateur enthusiasts.

These splendid melancholy eyes were turned upon me from the mirror with a haughty stare.
The intervals of silence grew progressively longer; the delays became very maddening.
He sometimes considered suicide but the thought was too oppressive to remain in his mind.
He had patronized her when she was a schoolgirl and teased her when she was a student.
Slicing it out carefully with his knife, he folded it without creasing the face.

Jane's relatives had decided that her gentleman friend was not one of high status.
After passing all the exams, the class celebrated for an entire week without resting.

Sometimes the scapegoat is an outsider who has been taken into the community.
The entire construction crew decided to lengthen their work day in order to have lunch.
Without tension there could be no balance either in nature or in mechanical design.
Two or three substantial pieces of wood smoldered on the hearth, for the night was cold.

Appendix D – SIMS

Why are you currently engaged in this activity?

1. Because I think that this activity is interesting
2. Because I am doing it for my own good
3. Because I am supposed to do it
4. There may be good reasons to do this activity, but personally I don’t see any
5. Because I think that this activity is pleasant
6. Because I think that this activity is good for me
7. Because it is something that I have to do
8. I do this activity but I am not sure if it is worth it
9. Because this activity is fun
10. By personal decision
11. Because I don’t have any choice
12. I don’t know; I don’t see what this activity brings me
13. Because I feel good when doing this activity
14. Because I believe that this activity is important for me
15. Because I feel that I have to do it
16. I do this activity, but I am not sure it is a good thing to pursue it

Appendix E – Self-paced reading task: experimental items

1 A Four children were continuously whining.
Three were very loud throughout the entire nap period.
B Two children were continuously whining.
Three were very quiet throughout the entire nap period.
C Four children were continuously whining.
Three were very quiet throughout the entire nap period.
D Two children were continuously whining.
Three were very loud throughout the entire nap period.
S Some children were whining.

2 A Six cups were filled with pink lemonade.
Five were completely full because the server was careless.
B Three cups were filled with pink lemonade.
Five were completely empty because the server was careless.
C Six cups were filled with pink lemonade.
Five were completely empty because the server was careless.
D Three cups were filled with pink lemonade.
Five were completely full because the server was careless.
S The server carefully filled the cups with lemonade.

3 A Six wines at the tasting were Italian.
Four were from Rome and known for their grapes.
B Three wines at the tasting were Italian.
Four were from France and known for their grapes.
C Six wines at the tasting were Italian.
Four were from France and known for their grapes.
D Three wines at the tasting were Italian.
Four were from Rome and known for their grapes.
S Several wines were from Italy.

4 A Eight movies in the festival focused on America.
Six were about Texas with a focus on politics.
B Four movies in the festival focused on America.
Six were about Africa with a focus on politics.
C Eight movies in the festival focused on America.
Six were about Africa with a focus on politics.
D Four movies in the festival focused on America.
Six were about Texas with a focus on politics.
S None of the movies focused on politics.

5 A Ten babies were asleep in the nursery.
Eight were loudly snoring which concerned the attending nurse.
B Five babies were asleep in the nursery.
Eight were loudly crying which concerned the attending nurse.
C Ten babies were asleep in the nursery.
Eight were loudly crying which concerned the attending nurse.
D Five babies were asleep in the nursery.
Eight were loudly snoring which concerned the attending nurse.
S The babies were in the nursery.

6 A Eight sandwiches in the cooler were cheese.
Five were lean swiss due to strict dietary restrictions.
B Four sandwiches in the cooler were cheese.
Five were lean ham due to strict dietary restrictions.
C Eight sandwiches in the cooler were cheese.
Five were lean ham due to strict dietary restrictions.
D Four sandwiches in the cooler were cheese.
Five were lean swiss due to strict dietary restrictions.
S The lean sandwiches were in the cooler.

7 A Five diamond necklaces in the window were pricey.
Four were very expensive but were actually rather ugly.
B Three diamond necklaces in the window were pricey.
Four were very cheap but were actually rather ugly.
C Five diamond necklaces in the window were pricey.
Four were very cheap but were actually rather ugly.
D Three diamond necklaces in the window were pricey.
Four were very expensive but were actually rather ugly.
S All the necklaces were cheap.

8 A Eight kids passed their physicals.
Six were really healthy according to the examining doctor.
B Five kids passed their physicals.
Six were really ill according to the examining doctor.
C Eight kids passed their physicals.
Six were really ill according to the examining doctor.
D Five kids passed their physicals.
Six were really healthy according to the examining doctor.
S A doctor examined a few of the kids.

9 A Four relatives at her birthday party were men.
Three were distant uncles who were smoking and drinking.
B Two relatives at her birthday party were men.
Three were distant aunts who were smoking and drinking.
C Four relatives at her birthday party were men.
Three were distant aunts who were smoking and drinking.
D Two relatives at her birthday party were men.
Three were distant uncles who were smoking and drinking.
S None of her relatives smoked.

10 A Six classes in the course schedule were science classes.
Five were advanced chemistry and slotted for Monday mornings.
B Three classes in the course schedule were science classes.
Five were advanced English and slotted for Monday mornings.
C Six classes in the course schedule were science classes.
Five were advanced English and slotted for Monday mornings.
D Three classes in the course schedule were science classes.
Five were advanced chemistry and slotted for Monday mornings.
S The science classes were on Monday morning.

11 A Six deserts were pies.
Four were rich apple arranged in a beautiful display.
B Three deserts were pies.
Four were rich cakes arranged in a beautiful display.
C Six deserts were pies.
Four were rich cakes arranged in a beautiful display.
D Three deserts were pies.
Four were rich apple arranged in a beautiful display.
S Some desserts were arranged in a beautiful display.

12 A Eight students were born in America.
Six were from Florida but later moved to California.
B Four students were born in America.
Six were from Asia but later moved to California.
C Eight students were born in America.
Six were from Asia but later moved to California.
D Four students were born in America.
Six were from Florida but later moved to California.
S The students from America moved to Florida.

13 A Ten bananas in the bowl were unripe.
Eight were quite green and unusable for the recipe.
B Five bananas in the bowl were unripe.
Eight were quite brown and unusable for the recipe.
C Ten bananas in the bowl were unripe.
Eight were quite brown and unusable for the recipe.
D Five bananas in the bowl were unripe.
Eight were quite green and unusable for the recipe.
S The unusable bananas were green.

14 A Eight paintings were hanging throughout the museum.
Five were in halls since the curator acquired them.
B Four paintings were hanging throughout the museum.
Five were in storage since the curator acquired them.
C Eight paintings were hanging throughout the museum.
Five were in storage since the curator acquired them.
D Four paintings were hanging throughout the museum.
Five were in halls since the curator acquired them.
S All the paintings were in halls.

15 A Five birds in the tree were cardinals.
Four were colorful males loudly singing a melodic song.
B Three birds in the tree were cardinals.
Four were colorful canaries loudly singing a melodic song.
C Five birds in the tree were cardinals.
Four were colorful canaries loudly singing a melodic song.
D Three birds in the tree were cardinals.
Four were colorful males loudly singing a melodic song.
S The male birds were in the tree.

16 A Eight primates at the zoo were gorillas.
Six were older males all kept in separate cages.
B Five primates at the zoo were gorillas.
Six were older chimps all kept in separate cages.
C Eight primates at the zoo were gorillas.
Six were older chimps all kept in separate cages.
D Five primates at the zoo were gorillas.
Six were older males all kept in separate cages.
S The gorillas were in the wild.

17 A Four employees at the institute were health care professionals.
Three were licensed nurses who worked the night shift.
B Two employees at the institute were health care professionals.
Three were licensed plumbers who worked the night shift.
C Four employees at the institute were health care professionals.
Three were licensed plumbers who worked the night shift.
D Two employees at the institute were health care professionals.
Three were licensed nurses who worked the night shift.
S The nurses worked at the institute.

18 A Six vehicles on the highway were trucks.
Five were small pickups which were painted bright colors.
B Three vehicles on the highway were trucks.
Five were small motorcycles which were painted bright colors.
C Six vehicles on the highway were trucks.
Five were small motorcycles which were painted bright colors.
D Three vehicles on the highway were trucks.
Five were small pickups which were painted bright colors.
S All of the vehicles were small pickups.

19 A Six pieces of furniture on display were chairs.
Four were large armchairs that were currently on sale.
B Three pieces of furniture on display were chairs.
Four were large tables that were currently on sale.
C Six pieces of furniture on display were chairs.
Four were large tables that were currently on sale.
D Three pieces of furniture on display were chairs.
Four were large armchairs that were currently on sale.
S The armchairs were on display.

20 A Eight utensils on the kitchen counter were for cutting.
Six were large knives that belonged to the chef.
B Four utensils on the kitchen counter were for cutting.
Six were large spoons that belonged to the chef.
C Eight utensils on the kitchen counter were for cutting.
Six were large spoons that belonged to the chef.
D Four utensils on the kitchen counter were for cutting.
Six were large knives that belonged to the chef.
S All of the utensils were spoons.

21 A Ten books at the sale were novels.
Eight were interesting stories about life in ancient Greece.
B Five books at the sale were novels.
Eight were interesting textbooks about life in ancient Greece.
C Ten books at the sale were novels.
Eight were interesting textbooks about life in ancient Greece.
D Five books at the sale were novels.
Eight were interesting stories about life in ancient Greece.
S The books about ancient Greece were on sale.

22 A Eight retirees had booked a trip to Europe.
Five had chosen France for a week long tour.
B Four retirees had booked a trip to Europe.
Five had chosen Japan for a week long tour.
C Eight retirees had booked a trip to Europe.
Five had chosen Japan for a week long tour.
D Four retirees had booked a trip to Europe.
Five had chosen France for a week long tour.
S All the retirees were going to Japan.

23 A Five soldiers were stationed in the middle east.
Four were in Israel since the Iraq war began.
B Three soldiers were stationed in the middle east.
Four were in Italy since the Iraq war began.
C Five soldiers were stationed in the middle east.
Four were in Italy since the Iraq war began.
D Three soldiers were stationed in the middle east.
Four were in Israel since the Iraq war began.
S Some of the soldiers were not fighting in Iraq.

24 A Eight people passed their driving test.
Six did very well according to the examining official.
B Five people passed their driving test.
Six did very poorly according to the examining official.
C Eight people passed their driving test.
Six did very poorly according to the examining official.
D Five people passed their driving test.
Six did very well according to the examining official.
S The official thought all of the people did poorly.

25 A Four trees at the nursery were tropical.
Three were tall palms that almost reached the roof.
B Two trees at the nursery were tropical.
Three were tall oaks that almost reached the roof.
C Four trees at the nursery were tropical.
Three were tall oaks that almost reached the roof.
D Two trees at the nursery were tropical.
Three were tall palms that almost reached the roof.
S The palm trees were at the nursery.

26 A Six doors were just painted in bright colors.
Five were completely red with brass door knobs.
B Three doors were just painted in bright colors.
Five were completely black with brass door knobs.
C Six doors were just painted in bright colors.
Five were completely black with brass door knobs.
D Three doors were just painted in bright colors.
Five were completely red with brass door knobs.
S The red doors were painted a long time ago.

27 A Six newspapers for sale were South American.
Four were from Brazil but reported news from England.
B Three newspapers for sale were South American.
Four were from Russia but reported news from England.
C Six newspapers for sale were South American.
Four were from Russia but reported news from England.
D Three newspapers for sale were South American.
Four were from Brazil but reported news from England.
S The papers from Brazil were for sale.

28 A Eight teachers were extremely disappointed after grading the tests.
Six were quite upset by the students' unexpected performance.
B Four teachers were extremely disappointed after grading the tests.
Six were quite impressed by the students' unexpected performance.
C Eight teachers were extremely disappointed after grading the tests.
Six were quite impressed by the students' unexpected performance.
D Four teachers were extremely disappointed after grading the tests.
Six were quite upset by the students' unexpected performance.
S All of the teachers were pleased with the grades.

29 A Ten instruments in the orchestra were strings.
Eight were loud violins which sounded magnificent to us.
B Five instruments in the orchestra were strings.
Eight were loud horns which sounded magnificent to us.
C Ten instruments in the orchestra were strings.
Eight were loud horns which sounded magnificent to us.
D Five instruments in the orchestra were strings.
Eight were loud violins which sounded magnificent to us.
S The instruments sounded magnificent.

30 A Eight pants were made of synthetic fibers.
Five were grey nylon which made them quite popular.
B Four pants were made of synthetic fibers.
Five were grey wool which made them quite popular.
C Eight pants were made of synthetic fibers.
Five were grey wool which made them quite popular.
D Four pants were made of synthetic fibers.
Five were grey nylon which made them quite popular.
S Some pants were nylon/wool.

31 A Five boxes at the farmers market were filled with fruit.

Four had fresh oranges grown at a local farm.

B Three boxes at the farmers market were filled with fruit.

Four had fresh carrots grown at a local farm.

C Five boxes at the farmers market were filled with fruit.

Four had fresh carrots grown at a local farm.

D Three boxes at the farmers market were filled with fruit.

Four had fresh oranges grown at a local farm.

S The boxes of oranges were rotten.

32 A Eight beverages on the menu were non-alcoholic.

Six were nice juices made from entirely organic ingredients.

B Five beverages on the menu were non-alcoholic.

Six were nice whiskies made from entirely organic ingredients.

C Eight beverages on the menu were non-alcoholic.

Six were nice whiskies made from entirely organic ingredients.

D Five beverages on the menu were non-alcoholic.

Six were nice juices made from entirely organic ingredients.

S All drinks contained alcohol.

Appendix F – Self-paced reading task: fillers

1 Four musicians were playing some country music on the square.
Five played guitar and wore cowboy hats.
S Some musicians wore cowboy hats.

2 Six different sausages were hanging from the hooks behind the counter.
Seven were Polish and had a strong aroma.
S The Polish sausages were on the counter.

3 Six pencils were in the pencil holder on my desk.
Seven were special drawing pencils that my mother gave me.
S All of the pencils were in the pencil holder.

4 Eight players were on the field.
Ten were getting into position so the game could begin.
S The game had already begun.

5 Two roses were getting ready to bloom.
One was already drooping in the vase.
S Some roses were still buds.

6 Three new bars opened downtown last year.
Two were not doing too well, though.
S None of the bars was doing great.

7 Three apples were bruised and had to be trashed.
Two were bought at a local supermarket.
S The bruised apples were not for sale.

8 Four bright stars can be seen in the west.
Three are only in view during this time of the year.
S The bright stars can be seen year round.

9 Ten girls were at the party.
Twenty friends came later and they chatted for a while.
S Most people at the party came later.

10 Eight rare ducks were sitting on the grass.
Fifteen geese all of a sudden flew up when we approached.
S The rare ducks suddenly flew up.

11 Five peppers in the bowl were finally ripe.
Six avocados were ready to eat as well.
S The avocados and the peppers were ripe.

12 Eight different cookies are usually for sale at this bakery.
Nine pies are made daily and are sold by the slice.
S Whole pies are usually for sale.

13 Five presents were in each child's stocking.
Three packages were under the Christmas tree.
S More presents were in the stockings.

14 Four parents were cheering from the bleachers.
Three players had done really well in the junior league that season.
S Several parents performed in the league.

15 Three mugs contained strong black coffee.
Two glasses were filled with rum.
S Most of the drinks were in mugs.

16 Five kids were playing in the pool.
Four adults were watching them from the side.
S The adults were in the water.

17 Four sunglasses were on display in the window.
They were inspired by the fashion of the '50-s.
S All of the glasses were on display.

18 Six bottles of juice were carried into the house.
Some were still in the car.
S All of the bottles were in the car.

19 Six clocks were mounted on the office wall.
Some were broken which annoyed the employees.
S The employees could see the clocks.

20 Eight colorful roosters were for sale at the fair.
They could be seen in the exhibit hall next door.
S The roosters in the exhibit hall were all brown.

21 Two donuts were gone within a few minutes.
Most were very popular amongst the office workers.
S Office workers ate the donuts.

22 Three cashiers were working the registers.
They were very busy due to the holiday tomorrow.
S Few cashiers were busy.

23 Three tourists were touring the fort.
Most had read extensively about the battles that had occurred there.
S Some tourists knew the history of the fort.

24 Four small towns were located near the river.
They all had friendly residents and quaint historic areas.
S Few towns near the river had friendly residents.

25 Ten computers were out of order.
The techs had not come yet to fix them.
S The computers were still broken.

26 Eight flowers were on each table.
The vases were made of colorful china.
S The flowers were all in white vases.

27 Five cartoons are produced every week.
The animators are based in Korea and work around the clock.
S The Korean animators produce cartoons every week.

28 Eight piglets have just been born at the petting zoo.
The children can come over to see them tomorrow.
S The children can see the piglets in a week.

29 Five crows were perched on the tree.
The sparrows were trying to chase them away from their nests.
S The crows were all in the tree.

30 Four swans were swimming in the lake.
Several boaters were anxiously trying to avoid them.
S All the boaters were disturbing the swans.

31 Three passengers got sick on the plane.
The airline blamed it on the salad provided by the catering company.
S The food made the passengers sick.

32 Five horses were still in the race.
Many spectators were screaming and shouting to cheer on the jockeys.
S The horse race had just ended.

33 About half of the students have signed up for the class.
Most are interested in language courses.
S Language courses are popular.

34 People at the beach wore a lot of sunblock.
Many had pale skin so they might burn easily.
S Most people at the beach had tan skin.

35 Few voters know the details of the proposal.
Many haven't read any of the text.
S The proposal was not read by many voters.

36 Most professors went to the meeting.
A few stayed in their offices grading exams.
S Many professors stayed in their offices.

37 Some bridesmaids were walking behind the bride.
A few were talking with a small boy, who was the groom's nephew.
S Some bridesmaids were not in the ceremony.

38 Few outdoor pools open before March.
Some have heated water and can open early.
S Pools without heated water open early.

39 All the players ran into the field to celebrate.
Some jumped up and down because they had won the championship.
S The players were pleased with the win.

40 Few trees survived the harsh winter.
Most suffered a lot from the ice storm.
S The ice storm caused no damage to the trees.

41 Some planes were delayed or cancelled due to the bad storm.
Many won't be going out today which angered the passengers.
S The passengers were upset over the bad storm.

42 Many people have one or two pets.
Few have exotic ones such as monkeys or alligators.
S Many people have monkeys for pets.

43 A few beads are needed to make this ornament.
They should match the fabric of the star.
S Not many beads will be on the ornament.

44 The entire audience applauded for over five minutes.
Many were also standing to show how good it had been.
S The applause were very brief.

45 A lot of turtles were on display in the main tank at the aquarium.
A few were quite large and often scared small children.
S Children are often scared at the main tank.

46 Many eggs are produced daily by John's chickens.
Most are sold to the grocery store downtown.
S The grocery store buys few eggs from John.

47 A few guests were watching TV at the bar.
Some were interested in breaking news about the political scandal.
S Breaking news interested a few people at the bar.

48 The kids took part in the community egg hunt.
Many were searching in the parking lot but there were no eggs there.
S Most of the eggs were in the parking lot.

49 A lot of diskettes were in the box.
The labels were dated and coded by subject.
S The diskettes were labelled.

50 Several ships sank in the ocean that night.
The storm was too violent for the rescue team to do anything.
S The storm spared all the ships.

51 Few people had seen the announcement.
The demolition of the building took many by surprise.
S The demolition of the building was announced.

52 The singer performed in the mall.
The production company had arranged everything with support from local merchants.
S The singer lacked local support.

53 Few candy bars are low in calories.
The amount of fat is often more than forty percent.
S Fat content is high in most candy bars.

54 Several potatoes needed to be peeled for tonight's feast.
A few tomatoes needed to be washed and sliced as well.
S The feast will consist of at least one potato dish.

55 The surfing competition was cancelled.
The wind was very strong and the waves were too violent.
S Safety issues prompted the cancellation.

56 All of the people were crying.
The funeral was very emotional, especially when the music started.
S Few of the people at the funeral were moved.

57 The room was filled with flowers.
The florist had delivered them in time for the newlyweds to see.
S The flowers were for the newlyweds.

58 Many patients could not be seen in the clinic.
Most appointments were all taken for that day and nobody had cancelled.
S Many patients had cancelled their appointments.

59 Everyone carried umbrellas with them.
Several weathermen were predicting wind and rain.
S Most people thought it would rain.

60 Several trees were planted along the boulevard.
A few streets were part of a project to make the town look better.
S Trees will make the town look worse.

61 The trash smelled badly.

The family had prepared fish for dinner the other night.

S Fish caused the trash to stink.

62 Some of our neighbors have plants on their porch.

Most ferns are easy to maintain with all the rain we get in summer.

S All of the neighbors have ferns on their porches.

63 Few kids like this educational program on TV.

Many cartoons are much more popular.

S Most kids like cartoons.

64 A few people at work take the stairs every day.

The elevator is too slow for them, and it keeps them in shape.

S A few people take the fast elevator.