BL

VBL DIG
9/1/12

FOR REFERENCE ONLY

10 JAN 2006

40 0761590 0

ProQuest Number: 10183503

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

ProQuest 10183503

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

WORKING MEMORY AND MULTI-DIGIT ARITHMETIC: THE EFFECTS OF DUAL-TASKS, CARRIES, FORMAT AND STRATEGY.

STEPHEN ROGER CROSS

A thesis submitted in partial fulfilment of the requirements of

The Nottingham Trent University for the degree of

Doctor of Philosophy

October 2004

Abstract

Following a review of working memory (WM), cognitive arithmetic and associated cognitive models, this study identified and investigated the lack of consensus on WM involvement in cognitive arithmetic. Two series of dual-task experiments extended previous studies through the inclusion of multi-digit problems from all four arithmetic operations, and through various manipulations of number of carries, visual format of presentation and solution difficulty. The effect of strategy was also explored, as was the effectiveness of different secondary tasks. The following conclusions are offered:

The involvement of the WM components described in the Baddeley and Hitch WM model (1974, 2000) is supported, but this involvement is heavily dependent not only on task and problem factors but also on the type of arithmetic operation. The phonological loop was implicated in multiplication, and the central executive in addition, while subtraction and division provided no evidence implicating either of these components.

The studies of Heathcote (1994) and Trbovitch and LeFevre (2003) are supported, in that there is evidence that linear (horizontal) format of presentation does appear to support phonological *reading* processes, while columnar (vertical) format does appear to engage visuo-spatial processing, with the latter providing greater support for the carry operation.

These results provide partial support for the existence of a separate cognitive module for number processing, (Butterworth, 2000), but also suggest that the triple-code model, (Dehaene and Cohen, 1995), needs to accommodate the possibility that the phonological loop is only involved when rote-learned 'verbal word frame' multiplication facts are retrieved, and that additions are solved by the use of the central executive to calculate or count addition totals.

The existence of task x problem interactions supports the encoding complex – interactive model of Campbell and Clarke (1998), but brings into question the abstract modular model of McCloskey, Caramazza and Basili (1985).

Secondary tasks cannot be assumed to load WM simply in virtue of their phonological, visuo-spatial and *executive* associations, and different secondary *random generation* tasks are not equivalent in terms of their ability to load the *executive* components of WM.

Response times and error rates demonstrated different patterns of main effects and interactions. This suggests that they measure different aspects of cognitive load. Response times possibly being more indicative of encoding processes, and error rates more indicative of calculation processes. Error scores should therefore be the preferred measure in future studies that investigate WM involvement in arithmetic processes.

Analysis of strategy use supports and extends the findings of Hecht (2002), in that relatively few participants change their solution strategy when faced with problems at different experimental levels, but different strategies can demonstrate significantly different response times. Consequently, strategy use, if not controlled and analysed, has the ability to undermine the logic of dual-task studies.

Acknowledgements

The support, co-operation and facilities offered by staff and students at Malbank School, Sandbach School, South Cheshire College, the Open University and Nottingham Trent University is gratefully acknowledged, as is the helpful advice of Professor Andre Vandierendonck on the interpretation of the RIGANAL program. Special thanks to my supervisors at Nottingham Trent, Professor Jean Underwood, Dr. Susannah Lamb and Dr. Keith Laws for their comments, support and guidance over the duration of this research.

Contents

List of tables

List of figures

Introduction

Rationale for the study – answering the 'why' question

This study is based on the desire to obtain an answer to the apparently simple question "How is working memory used in arithmetic?" The origins of this desire emerged partly through reflection on the experience of studying and teaching both mathematics and psychology, and partly through the growing realisation that the literature detailing previous studies of working memory and cognitive arithmetic gave inconsistent and contradictory accounts of working memory involvement.

Working memory (WM) is defined in different ways by different theorists, but here the working memory model of Baddeley and Hitch (1974, 2000) serves as the theoretical framework. The research described in this study is designed not only to indicate when the components of this WM model are implicated in the solution of arithmetic problems, but also to discover which particular aspects of these problems lead to such WM involvement. The three established components of the Baddeley and Hitch WM model; the central executive, phonological loop and visuo-spatial sketch pad, have all been variously implicated in arithmetic processing, but with broadly similar studies frequently leading to directly contradictory results.

The literature review, forming the initial three chapters of this study, investigates the reasons for this lack of consensus, and demonstrates that this is due to several factors. Not only do different studies generally use quite different types of arithmetic tasks and memory load tasks, but also from study to study there are wide differences in presentational format, number of carries, and the potential impact of strategy use on response times and error rates. This situation is exacerbated by various researchers advocating different theoretical positions, and by the different cognitive models of arithmetic described in the literature review. As a result, there is no simple answer to the question posed earlier, for an answer based on a review of the relevant literature is surprisingly but necessarily vague, "It depends upon how you define and measure working memory, how you define and measure arithmetic, the individual differences of those doing the arithmetic, and the strategies that they use when doing it." This naturally invites the response "Well how should you define and measure working memory, how should you define and measure arithmetic, and how should you control for individual differences in ability and strategy use?"

The chapters following the literature review provide an account of two series of dual-task experimental investigations designed to answer these questions, and to give a more satisfactory account of the role of WM in cognitive arithmetic. In brief, this study investigates particular sets of multi-digit arithmetic problems from the four arithmetic operations, and carefully manipulates these problems in terms of the number of carry operations and the visual format of presentation. There are two reasons for the inclusion of problems from all four arithmetic operations. The first is that no previous studies have been able to provide a comparison of problems from all four arithmetic operations, especially when performed by the same participants, and this should allow differences between operations to be clearly indicated. The second is that any differences in WM involvement across these four operations should provide an indication of support for one or more of the cognitive models of arithmetic. These cognitive models emphasise working memory involvement in distinctly different ways, and lead to different predictions concerning WM involvement in arithmetic processes, varying from the prediction of no WM involvement at all, to the selective involvement of particular WM components dependent upon the type of arithmetic operation being investigated.

Strategy use is also investigated, for if particular strategies can remove the need for WM involvement, or decrease response times and error rates, then the logic of experimental studies is potentially undermined. The logic of dual-task methodology is also explored, for previous studies have relied upon significant main effects to demonstrate WM involvement, and it will be shown that this is insufficient, for significant main effects can be explained in terms of *cost of concurrence* rather than dual-task interference. Evidence for WM involvement should be found in the pattern of task x problem interactions, and such interactions can also provide an additional source of support for particular cognitive models. If dual-task interference is to be accepted, then there is also a need to demonstrate a performance decrement to both primary arithmetic task and secondary interference task. Where this is not the case, it is possible that attention or cognitive effort is switched between tasks dependent upon motivation and difficulty, thereby undermining any claim of dual-task interference. The choice of secondary tasks is also discussed, and the relative equivalence of three *random generation* tasks is investigated. The value of response times and errors as differential indicators of cognitive load is also addressed, as are the problems of individual differences in arithmetical competence and maths anxiety.

Chapter One - Working Memory

1.1 Which Model?

This study is concerned with working memory (WM) involvement in multi-digit arithmetic, but what is working memory? Miyake and Shah (1999) report that there is a general consensus, in that working memory is composed of the processes or mechanisms used to regulate, maintain or control, information relevant to a given cognitive task. There is little in this account with which to disagree, but as usual, 'the devil is in the detail', and there are several alternative theoretical positions supporting models of particular processing, storage and control mechanisms.

It might be more accurate to argue that the consensus relates to what working memory does rather than how it does it, for there is considerable disagreement over a number of questions relating to WM: Is it a unitary or multi-component system? (e.g. Conway & Engle 2001, Jones 1999, Anderson, Reder & Lebiere, 1996). Is it distinct from either short-term memory or long-term memory? (e.g. Kail & Hall 2001, Ericcson & Kintsch, 1995). How does the control mechanism allocate resources? (e.g. Norman & Shallice, 1980, Shallice & Burgess, 1996, Baddeley, 1996). How does WM develop over time, and to what extent are the various models complementary or conflicting? (e.g. Kemps, De Rammelaere & Desmet 2000, Pascual-Leone, 2000, Baddeley & Hitch, 2000). What predictions can be derived from the models, and what tasks can be used to explore these predictions? (e.g. Lehto, 1996). The choice of theoretical framework for this study is based on revisions of the original Baddeley and Hitch WM model (1974), but here their model is set in context through a review of some of the competing models.

Gathercole (1999) offers brief definitions for three alternatives to the Baddeley and Hitch model:

> "...(1) a system fuelled by a limited capacity resource that can be flexibly deployed to support either processing or storage, (2) activated portions of long-term memory controlled by an attentional resource with inhibitory capabilities, or (3) a short-term memory mechanism providing cue-based access to long-term working memory systems which are organized around special retrieval structures." (p.410).

The first of these three alternatives refers to a model originally proposed by Daneman and Carpenter (1980), which suggested that WM involved a trade-off between processing and storage needs, and that a measure of WM *span* – the ability to recall the final word from each of an increasing number of sentences - indicated WM storage and processing capacity, with span being a measure of processing resources after task demands are met (Miyake, 2001). Hitch, Towse and Hutton (2001) provide evidence against this *resource-sharing model*, as it has become known. The second of these three alternatives is known as the *controlled attention* account of WM (Engle, Kane & Tuholski, 1999, Tuholski, Engle & Bayliss, 2001, Kane, Bleckley, Conway & Engle, 2001). Controlled attention is the individual's ability to inhibit intrusions from irrelevant material during the processing of a task, and this combines with the relative efficiency of the individual's short-term memory storage, to determine overall WM capacity – as measured by WM span. The third of the alternatives referred to by Gathercole is the *long-term working memory* (LTWM) model offered by Ericcson and Kintsch (1995), who state:

> "To account for the large demands on working memory during text comprehension and expert performance, the traditional models of working memory involving temporary storage must be extended to include working memory based on storage in long-term memory". (p.211).

LTWM is not a general resource *per se*, but refers to the domain specific patterns of knowledge underlying skilled performance, such as the knowledge of openings and endgames possessed by an expert chess player, or the use of mnemonic strategies to aid recall. These patterns of knowledge or *retrieval structures,* as they are known in the model, are activated by relevant cues in short-term memory, and they then act as additional support for WM by providing storage in long-term memory, with the consequence that WM capacity appears to increase beyond the level assumed appropriate for a temporary store. This lead Ericcson and Kintsch to question the measurement of WM capacity in terms of such things as reading span, for this could involve storage in both LTWM and WM. However, Gobet (2000) claims that the concept of a *general retrieval structure* conflates three different types of retrieval structure. These are *generic retrieval structures*, such as mnemonic strategies, *episodic text structures*, as used in text comprehension, and finally, *domain specific knowledge*. The involvement of these different retrieval structures is dependent both upon the type of task studied, and upon the relative level of expertise of those performing the task. It is therefore appropriate to regard LTWM as an explanation of individual differences in WM ability, relevant to particular tasks, rather than as a model of WM processes.

Destefano and Lefevre (2004) in their review of WM and arithmetic, briefly discuss models of WM, and identify the Baddeley and Hitch model as underpinning the 'vast majority' of studies of working memory and arithmetic. The same reasoning underlies the choice of theoretical framework for this study, and although this does not necessarily imply that the Baddeley and Hitch model is regarded as the definitive account of WM, it does allow meaningful comparison with other studies within this tradition.

1.2 A brief history and description of the Baddeley and Hitch WM model

Baddeley's (2000) *Trends in Cognitive Sciences* paper describes the development of the multi-component WM model of Baddeley and Hitch (1974) from the earlier *modal model* of unitary short-term memory (Atkinson & Shiffrin, 1968). The initial three components of WM were given as the *central executive* (CE) – an attentional controller allocating processing resources, and two subsidiary or *slave* systems, the *visuospatial sketchpad* (VSSP) and the *phonological loop* (PL) – each slave system having both storage and rehearsal mechanisms, the VSSP maintaining both images and spatial representations, and the PL maintaining sound based representations. There is currently debate about whether the phonological loop should be viewed as a general processing resource for complex cognition, or as fulfilling a special function. Baddeley, Gathercole and Papagno (1998), proposed that the function of the PL is primarily as a *language-learning device*, which evolved to allow the storage of unfamiliar sound patterns during the construction of more permanent memories. Its use in the retention of word sequences is considered incidental. From this perspective, any role that the PL might have in the performance of arithmetic could also be incidental, but given the need for the temporary retention of partial sums in multi-digit arithmetic, and the potential for phonological involvement in such retention, it is necessary to investigate this possibility.

Logie (1995) provided evidence that the VSSP has a *passive visual cache* refreshed by an *active spatial rehearsal mechanism*, and although the VSSP has not been researched as extensively as the PL, these and other distinctions within WM are supported by further experimental evidence, e.g. Bruyer and Scailquin (1998), McConnell and Quinn (2000), and by evidence from neuroimaging techniques, e.g. Smith and Jonides (1997). The Smith and Jonides study used positron emission tomography (PET) scans, and reported different WM systems for spatial and verbal information, the former in the right hemisphere, and the latter in the left hemisphere. They also supported the existence of separate processing areas, for passive storage and active manipulation, in both the visuospatial and verbal systems, storage occurring in areas at the back of the brain, and active maintenance in areas at the front of the brain, with central executive functioning being tentatively identified with areas of prefrontal cortex. The original Baddeley and Hitch model has been revised a number of times, Baddeley (1986, 1990, and 2000) and the latest version now contains a new fourth component – the *episodic buffer* (EB). This new component is given for several theoretical reasons, partly in response to criticisms of the three-component model that require a more explicit linkage of WM and long-term memory, and partly as a solution to the *binding problem* – how is information linked together to allow consistency and conscious processing? A diagram adapted from Baddeley (2000), but also incorporating Logie's (1995) revision of visuospatial WM, is given below:

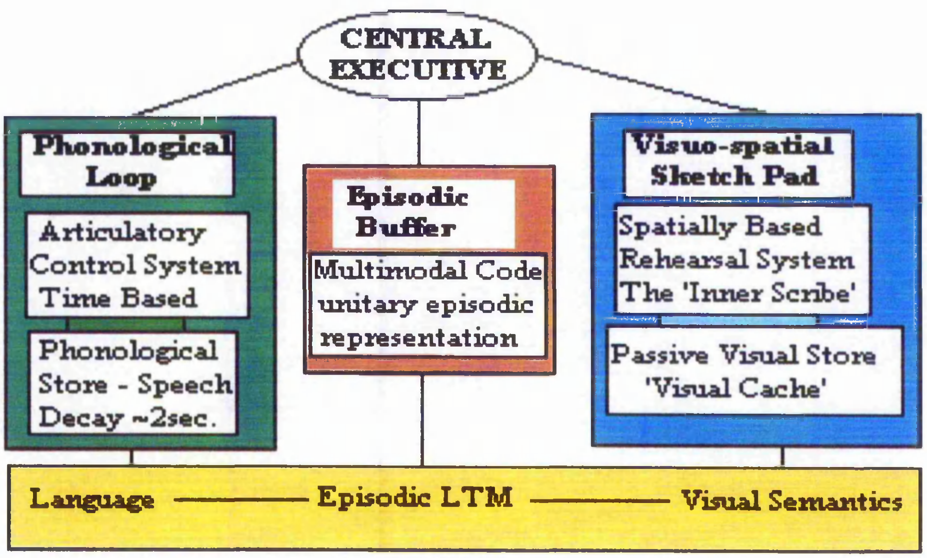

Figure 1. The revised WM model.

There is evidence that questions the earlier versions of the WM model. An example of a study that undermines the three-component model is provided by Butterworth, Cipolotti and

Warrington (1996), reporting the case of patient (MRF), who demonstrated short-term memory span impairment, but normal ability to solve auditorily presented multi-digit addition and subtraction problems. Butterworth et al. claimed that this brings into question both the role of the phonological loop component of the model, and the value of span measures in the investigation of mental arithmetic. Another example, provided by Macken Tremblay, Alford and Jones (1999), in a discussion of the evidence underlying the Object-Oriented Episodic Record (O-OER) model of short-term memory (Jones, 1993), reported that articulatory suppression and irrelevant speech are equally efficient at impairing both verbal and visual serial recall. The O-OER model has two components; *objects*, which are abstract non-modality specific, and hence *object-oriented,* representations of all sensory aspects of real world events, and *episodic pointers*, which give the probability of sequential position, and encode the temporal order of these object-oriented representations. In this model, dual-task interference arises not as a result of the two tasks being in the same sensory modality, but because both tasks require sequential ordering irrespective of their sensory modality. The model's *changing-state hypothesis* states that only secondary tasks that involve sequential ordering can cause interference. In discussing the effect of *irrelevant sound* as a secondary task, Jones (1999) stated:

> "Crucially, because the primary and irrelevant tasks are presented in different sensory modalities, the effect cannot be attributed to some kind of interference at the sensory level; instead, it must due to a confluence of processing from the eye and the ear at some level beyond the sensory organs." (p.169).

Clearly, the newly proposed *episodic buffer* in the revised Baddeley and Hitch WM model goes some way towards meeting the criticisms raised by the above studies. The EB provides a temporary store that has access to long-term memory independently of the slave systems, thereby offering a possible route for the preserved mathematical abilities of patient MRF in the Butterworth et al. study. It also has the purpose of binding together information from different sources into *coherent episodes,* utilising a multi-dimensional amodal code, under the direction of the *central executive*, and this is seemingly compatible with the O-OER model if the seriation process occurs in the EB rather than in the slave systems. However, Baddeley and Larsen (2003) in a defence of the PL component of WM, argue that articulatory suppression removed the *phonological similarity effect* – sequences such as *CDGPTV* leading to poorer recall than sequences such as *BFHJQR*, whereas irrelevant speech and syncopated tapping did not remove this effect, thereby questioning the equivalence of articulatory suppression and irrelevant speech. It might appear that the new episodic buffer would aid the explanation of how WM is used in the solution of multi-digit arithmetic problems, but there are reasons to expect that, at least initially, it will make interpretation more difficult.

Baddeley (2000) states that the EB can be seen either as a new component of WM, or as a further fractionation of the central executive. The central executive in the Baddeley and Hitch model is essentially similar to the *Supervisory Attentional System* (SAS) in the Norman and Shallice model (1980), and as Shallice and Burgess (1996) offered evidence that the SAS is unlikely to be a unitary whole, so Baddeley (1996) identified four possible fractionations of the CE. These are; co-ordination of performance on two separate tasks, capacity to switch retrieval strategies, selective attention to one stimulus while inhibiting the effect of others, and the ability to hold and manipulate information in long-term memory. It is not clear whether these fractionations are separate from the proposed role of the episodic buffer.

Andrade (2001) offers an initial appraisal of the episodic buffer, and suggested two related reasons for caution in interpreting the involvement of this component. The first arises because there are as yet no experimental manipulations that can be unambiguously linked to the episodic buffer, and the second, from the real possibility that the contribution of the central executive and the two slave systems could be under or over-estimated.

May (2001) stated:

> "...detecting working memory involvement in complex cognition will no longer be a relatively simple matter of looking for modality-specific effects of concurrent tasks on the phenomenon of interest. Null results may reflect involvement of the episodic buffer, or the central executive, or no working memory involvement at all." (p.306).

As the current study utilised a variety of dual-tasks designed to disrupt generally non-specific central executive performance, the episodic buffer was regarded as a component of the central executive, and results interpreted accordingly.

1.3 Dual-task techniques to disrupt the various components of WM

Dual-task techniques have been used extensively in explorations of the WM model. It is possible to measure the capacity of WM through the use of techniques such as digit span or word span, but this does not necessarily indicate whether the WM components are involved in particular cognitive tasks. The logic of dual-task techniques is simple – if a primary task and a secondary task, performed concurrently, are competing for the same limited processing resources, then one or both of the tasks should suffer some performance decrement. If there is no performance decrement, then arguably, the tasks utilise different processing resources, or the cognitive load associated with the secondary task is insufficient to disrupt performance on the primary task. The Baddeley and Hitch model is indeed partly based upon dual-task studies which indicated separate phonological and visuospatial resources, and while it might appear to be something of a circular argument to use dual-tasks to disrupt components identified by dual-task studies, this is the methodology identified by Destefano and Lefevre (2004), as underlying the majority of studies involving WM and arithmetic.

The technique of *articulatory suppression is* well established in the literature (e.g. Baddeley, 1990) and can be used as a secondary task to disrupt the PL. At its simplest, this involves the continual repetition of a word, such as 'the', to overload the capacity of the rehearsal mechanism and thereby prevent the sub-vocal rehearsal of information necessary to the primary task. *Canonical articulation* – the continual repetition of a phrase or sequence of letters, such as 'A, B, C', is thought to provide a slightly higher cognitive load in virtue of the maintenance of sequential order in addition to articulation. The disruption of the VSSP is complicated by the fact that it is difficult to be certain whether the storage or rehearsal component is affected by a particular secondary task, or indeed, whether image or spatial relationships are jointly or separately affected. Logie, Gilhooly and Wynn (1994) projected *irrelevant pictures* into the visual field of participants in order to disrupt the visual cache, and used *hand movement* – tapping a sequence of four unseen buttons – to disrupt the spatial rehearsal mechanism. Heathcote (1994) presented randomly generated matrix patterns to disrupt the visual cache, and 'forward and reverse' spatial tapping, along the rows of a 5 x 5 array of unseen keys, to disrupt the spatial rehearsal mechanism. Quinn and McConnell (1999) and McConnell and Quinn (2000), provide evidence that a static visual field, as used in the above studies, does not have obligatory access to the visual cache. This suggests that although these studies clearly have surface similarities, it is at least possible that the tasks differed not only in terms of their relative complexity, but also in their ability to disrupt the separate components of the VSSP.

There are a variety of tasks that can be used to disrupt the central executive. Lehto (1996), provided evidence that 'frontal lobe function' tasks such as the Wisconsin Card Sorting Task – used to measure *perseveration* through the relative inability to switch card sorting strategies – are associated with *complex span* measures of WM, but that such tests may be measuring different executive functions. Baddeley, Emslie, Kolodny and Duncan

(1998), investigated *random generation* tasks as potential disruptors of executive function, and in particular the equivalence of a verbal random generation task *random number generation* (RNG) – articulation of numbers from 0-9 - with a 'motor' random generation task *random key-pressing* (RKP) – tapping on an array of 10 keys. They concluded that random generation tasks require participants to both inhibit repetition and to switch retrieval plans – two of the executive functions described in Baddeley (1996). Miyake, Friedman, Emerson and Witzki (2000) also linked RNG performance to 'inhibition' and 'updating', but Towse (1998) disputed the equivalence of RNG and RKP by demonstrating that the randomness produced in the oral generation task differs in relation to response speed and number of items in the set of responses, which is not the case for RKP, and he cautions against the straightforward adoption of random key-pressing as an executive loading task. The potential for the involvement of the two slave systems must also be considered in relation to random generation tasks. The articulation of random numbers or letters will load the phonological loop, and is therefore considered not only an executive task, but also an articulatory suppression task. Similarly, the tapping of keys at random might involve spatial processing – participants generating and following their own sets of spatial sequences – and RKP can therefore be seen as a task which loads the VSSP and the CE.

In an attempt to provide a relatively 'pure' executive task, a *random interval generation* (RIG) task has been proposed (Vandierendonck, De Vooght & Van der Goten 1998, Vandierendonck 2000). This task involves the pressing of a single key to produce a series of time intervals as 'random and unpredictable' as possible. This task does not involve articulation, and only requires minimal hand movement without the need to follow any pattern. The RIG task has been shown to disrupt executive processing, but Vandierendonck et al. (1998) reported that the effects of random letter generation were much larger than the disruption produced by the RIG task.

1.4 Methodological issues relating to the use of dual-tasks

It is possible that dual-task methodology is undermined in certain circumstances. Hegarty, Shah and Miyake (2000) examined the logic of dual-task performance, especially when dual-tasks are used to disrupt the central executive, and identified two 'inherently related' factors, a *response selection bottleneck* and a *strategic trade-off* between primary and secondary tasks. A strategic trade-off arises when there is a large difference in perceived difficulty between the primary and secondary tasks, and processing resources are diverted to the more difficult of the two tasks. If the more difficult task is the secondary task, then any decrement in primary task performance might not indicate competition for the same processing resources, but result from this strategic trade-off. The reverse conditions, where the primary task is perceived as the more difficult, might result in resources being diverted from the secondary task to the primary task, leading to a smaller performance decrement on the primary task. To increase confidence that two tasks are in fact competing for the same processing resources, both tasks should suffer some level of decrement compared to their performance under single-task conditions. Therefore, dual-task studies need to measure and analyse *'baseline'* performance of primary and secondary tasks as single-tasks, and compare this with dual-task performance. However, a negligible, or relatively small, decrement to a secondary task could indicate a number of things; that there was a strategic trade-off, that the secondary task was not sufficiently difficult to load the WM component involved in processing the primary task, that the secondary task loads a different WM component, or even that it loads none of the components of WM. Andrade (2001) considered the assumption that both tasks in a dual-task study do in fact tap the same WM component, in terms of the analogy provided by Gregory (1961):

> "The removal of any of several widely spaced resistors may cause a radio set to emit howls, but it does not follow that howls are immediately associated with these resistors, or indeed that the causal relation is anything but the most indirect. In

> particular, we should not say that the function of the resistors in the normal circuit is to inhibit howling." (1961, p.323).

Clearly, disruption to a particular WM component, under dual-task conditions, is not of itself a sufficient condition for identification of either of the tasks with that component. A mixture of theoretical explanation and a body of experimental evidence are needed before such identification can gain acceptance.

The response selection bottleneck identified by Hegarty et al. (2000) provides more difficulties for those interpreting dual-task studies, and they follow the argument of Pashler (1994) when they state:

> "...that a bottleneck can occur in dual-task situations at the response selection phase (i.e., two responses cannot be selected at the same time) even when other perceptual and motor processes can co-occur." (Hegarty et al. p.377).

If both primary and secondary tasks involve rapid response selection, then this might be the cause of any decrement in performance, rather than competition for the same WM component. The reason that this problem is 'inherently related' to the strategic trade-off issue, is that as task complexity, and solution time, increases, the need for rapid response selection is likely to decrease. The solution responses to multi-digit arithmetic problems are given relatively slowly, but it is possible that the processes requiring the retention of partial solutions involve more rapid response selection, and accordingly, results need to be interpreted with caution. Ehrenstein, Schweikert, Sangster and Proctor (1997) also considered the problems of interpreting dual-task studies given a *response selection bottleneck* of the type already described. Their study reported two experiments that involved the concurrent presentation of two digits, one of the digits serving as the basis for a subtraction task, and the other as the basis for a memory search task. In the first experiment, participants were instructed to carry out the subtraction task before the memory search task, and in the second experiment to do the reverse. Ehrenstein and his colleagues interpreted results in terms of alternative models of sequential processing in WM, the *serial processing* and *grouping* models both assuming that one of the tasks is completed before the other, but the grouping model also assumes that the response for the first task is withheld until the second task is completed. On the basis of interactions between the two tasks and, through the technique of *critical path analysis*, they largely rejected these models in favour of a *double-bottleneck* model of concurrent processing, where switching between *central processes* forms the first bottleneck, and *response selection* forms the second bottleneck. The important point here is that significant main effects of the experimental factors, resulting from an increase in the measurements of response times and error rates, cannot be taken as direct evidence of dual-task interference. Such increases in response times and error rates might simply be attributed to *cost of concurrence* – which is that when doing two things at once more time is required because each task is performed sequentially.

In order to increase confidence that interference is demonstrated in dual-task studies, there should be significant patterns of primary and secondary task interaction, and not only should these interactions be apparent in previous studies of WM and arithmetic, they should also be expected in future experimental studies. Oberauer, Demmrich, Mayr and Kliegel (2001) provided further support for this view when they considered the lack of task x problem interaction in some previous studies of WM and arithmetic, and stated:

> "The absence of such interactions is consistent with the hypothesis that the load on working memory did not affect the arithmetic process itself, but a reaction time component that was constant over all difficulty levels." (p.19).

In other words, an increase in reaction time due to cost of concurrence and not competition for working memory resources. In summary, evidence for WM involvement needs to be sought in the pattern of task x problem interactions, and not in the pattern of main effects. Having considered WM and the use of dual-task methodology to identify WM involvement, the next chapter considers the field of cognitive arithmetic.

Chapter Two - Cognitive Arithmetic

2.1 What is Cognitive Arithmetic?

Cognitive arithmetic is not simply an alternative term for mental arithmetic; it is the study of the mental representations and cognitive processes underlying the manipulation of number in children and adults, and in particular, those representations and processes associated with performance on a variety of arithmetic tasks, whether simple or complex. Theories and models in the field of cognitive arithmetic have to consider not only the more general processes required for encoding, storing and maintaining verbal and visual information in short-term memory, but also those processes involving the storage and retrieval of arithmetic facts and problem solutions in long-term declarative and procedural memory. Working memory can be seen as providing the interface between short-term and long-term memory, and consequently, the study of cognitive arithmetic often involves the study of working memory. As arithmetic facts are not innate, and solution processes depend upon the possession of learned algorithms or heuristic strategies, studies in cognitive arithmetic must also take into account the development of numerical cognition and individual differences in performance on arithmetic tasks. Ashcraft (1992, 1995) offered two reviews of cognitive arithmetic, and reported that the field effectively began with the publication of a paper by Groen and Parkman (1972), which offered the first cognitive model of addition, the *min* model. The *min* model is a basic 'counting-on' model of simple addition where the smaller of two addends, the *min* (minimum) is added to the larger addend in single digit steps e.g. 5 + 2 = 5+1 (6) + 1 (7). Early models, such as the *min* model, tried to provide an initial account for some of the now well established effects identified in studies of arithmetic processing: *Problem size/difficulty effect* – response times and errors increase with the magnitude of the operands, although this is modified by the *tie* *effect* – e.g. 5 + 5 processed faster than 7 + 3. *Error effects* – the errors made in arithmetic problems are not always simple mistakes, they often reflect intrusions from the same or other arithmetic operations – e.g. an incorrect addition total being a correct subtraction total, or *table errors* (6 x 3 = 24), and these have increased response times compared to simple errors. Increased error rates generally reflect increased response times, so that overall response times are not an accurate indicator of differences across conditions if error rates vary considerably over these conditions. To overcome this, studies using response times tend to analyse correct and incorrect problems separately. *Split effect* – in (true/false) arithmetic verification tasks, response times for false trials tend to decrease with the increasing distance between the correct and incorrect answers, suggesting that *plausibility* judgements are involved, which leads to a more general consideration of individual differences.

2.2 Maths anxiety

Individual differences in the ability to perform arithmetic tasks, from whatever cause, provide further difficulties for the design and interpretation of dual-task studies. Ashcraft (1995), Kellogg, Hopko and Ashcraft (1999) and Ashcraft and Kirk (2001) all provide evidence that individuals with a high level of math anxiety perform less well on arithmetic tasks, and have lower WM spans, than individuals with low levels of maths anxiety. Ashcraft and Kirk did not consider this performance decrement to be a direct consequence of simple differences in mathematical competence, as the decrement appeared when arithmetic tasks were limited to simple single digit operations for both high and low maths anxiety groups. However, on more complex multi-digit addition problems high maths anxiety participants did show less mathematical competence than low maths anxiety participants, and they also demonstrated increased response times for problems involving the carry operation. Ashcraft and Kirk consider that high levels of math anxiety provide an additional WM load, possibly as a result of the central executive diverting attentional resources from the arithmetic task to a 'preoccupation with worrying thoughts'. It is not clear whether attentional resources are used in the representation of such thoughts, or in the attempt to inhibit them, but if attentional resources are required, then maths anxiety becomes an interference task in its own right, thereby making a dual-task study into a triple-task study. Kellogg et al. (1999) investigated whether these 'worrying thoughts' involved concerns related to time pressure, but found no greater difference between high and low maths anxiety groups in timed and untimed conditions, although both groups demonstrated poorer performance in the timed condition.

2.3 Acquired competence

Competence in arithmetic is the ability to correctly solve arithmetic problems through the application of appropriate knowledge and procedures, but what factors underlie the acquisition of competence, and how might these factors impact upon studies in cognitive arithmetic? Mathematics teaching relies largely upon theories of cognitive development to guide educational practice (e.g. Liebeck, 1988, Nunes & Bryant, 1996), especially Piagetian theory (e.g. Bryant, Christie & Rendu, 1999), and consequently, mathematics educators seek to provide teaching materials and strategies appropriate to particular ages/stages of cognitive development. Typically, as Liebeck describes, mathematics teaching begins with activities designed to provide the *conceptual formation* of number and numerals, then counting procedures as an initial basis for addition and subtraction, and finally to strategies for multiplication and division based on retrieval. Fuson, Wearne, Hiebert, Murray, Kuman, Oliver, Carpenter and Fenema (1997) reported that children from a variety of cultures do have a common framework of *conceptual structures* for dealing with such things as multi-digit numbers, but that the acquisition of these conceptual structures is negatively impacted, especially in European and American cultures, by inconsistent activities, strategies and language use. An example of a conceptual structure is the mental representation of the place-value system for thousands, hundreds, tens and units for a number such as 1513, which also provides an example of inconsistent language use – 'one thousand five hundred and thirteen' – when consistency between language and conceptual structure would require – 'one thousand five hundreds one ten and three'. Miura, Okamota, Vlahovic-Stetic, Chungsoon and Jong Hye (1999) give an example of conceptual structure and language consistency – the Korean expression for the fraction one-third is *sam bun ui il*, which literally means 'of three parts one'. The importance of this for studies in cognitive arithmetic is that the cognitive load generated by a particular problem is likely to be dependent upon how that problem is represented (e.g. Kotovsky, Hayes & Simon, 1985). Differences between individuals, in terms of WM span or dual-task performance, might be a measure of the efficiency of representation rather than a clear indicator of WM involvement in a particular arithmetic processes or operations. Campbell and Xue (2001) also reported cross-cultural differences in mathematical competence, which they explained partly in terms of the differential use of representations based on procedural and retrieval strategies. Their sample of Chinese, Chinese Canadian, and non-Asian Canadian undergraduates all used procedural strategies to some extent, depending upon the type of problem presented. This questions the traditional view that adult simple arithmetic problem solving is predominantly, if not exclusively, based on retrieval (see also Baroody, 1999, Roussel, Fayol & Barroulet, 2002). In the Campbell and Xue study, the superior performance of undergraduates educated in China is partly explained by significant differences in the reported use of electronic calculators during early education, the implication being that as calculator use in China is less than that found in Western educational systems, children need to rely on mental representation to a greater extent. Brenner, Herman, Ho and Zimmer (1999) reported that the superior mathematical competence of Asian students results from a more flexible use of multiple representations, and that this is in part due to differences in the presentation of materials in mathematics textbooks. They found that Japanese textbooks integrate verbal, pictorial and symbolic representations in problem explanations, whereas typically, American textbooks do not integrate these representations. Graham (1999) reported the importance of *gestures* – such as finger pointing, in cognitive arithmetic studies of younger children, and Ashcraft (1995) provided anecdotal evidence that some participants used *finger writing* in the re-representation of division problems into conventional format. This suggests that for some individuals, external representations might be commonly used to reduce cognitive load, and that for these individuals, a visuospatial interference task could have a disproportionate effect. Further complications resulting from other individual differences also need to be considered: Geary, Saults, Liu and Hoard (2000) reported sex differences in arithmetical reasoning, spatial cognition and computational fluency in favour of males, despite no overall differences in IQ scores. Age-related WM differences are also reported – e.g. Vecchi and Cornoldi (1999), Oberauer et al (2001), such that WM ability is reduced in elderly adults, but there is a lack of information to show whether this is a gradual process across adulthood or only evident after a particular age.

2.4 Strategy use

The term *strategy* can refer to several different things in studies of cognitive arithmetic, but usually implies some choice of process made by the problem solver. Choice implies attention, and the use of *top-down,* conscious, slow, serial and capacity-limited processing, whereas retrieval suggests *bottom-up,* automatic, sub-conscious, fast, parallel and relatively unlimited-capacity processing, although this is not to say that top-down and bottom-up processes do not interact. In the material that follows, retrieval is usually described as a strategy, although whether retrieval can properly be said to be based on 'choice' is questionable. Ashcraft (1995) suggested a 'broad consensus' that arithmetic facts are stored in long-term memory as associative or network representations of varying strength, with the strength of each association being dependent upon prior learning, experience and frequency of use, and describes a number of *retrieval-based* models based on this assumption, such as that of Siegler and Jenkins (1989). The Siegler and Jenkins model is unusual in that it not only has an associative network for each arithmetic fact or problem, but also for the solution strategy associated with each problem, so that when a problem is experienced again, the solution strategy for that problem is directly activated. Dixon, Deets and Bangert (2001) reported that samples of college students were sensitive to violations of a number of *principles* of arithmetic, such as the *relation to operands* principle – additions and multiplications always give totals greater than the largest operand, while subtractions and divisions always give totals smaller than the largest operand. Conceivably, this is the sort of information that could also be represented in an associative network of arithmetic facts. The Siegler and Jenkins model explained poor problem performance in terms of weak associations, both of arithmetic facts and links to suitable strategies, but has been reformulated as the Adaptive Strategy Choice Model (ASCM) (Siegler & Shipley, 1995), and as the Strategy Choice and Adaptivity Simulation (SCADS) model, (Shrager & Siegler, 1998), in order to enhance the explanation of strategy choice. In the ASCM, each strategy has information associated with it concerning speed, novelty and choice, and several strategies may be available for any given task, individuals choosing the best strategy in relation to the situation experienced at the time. Strategy choice in the ASCM is not necessarily seen as dependent upon the availability of WM resources, but as dependent upon the relative efficiency with which each strategy can be executed, and on the solution efficiency associated with each strategy. The SCADS model incorporates the ASCM, but also models *metacognitive* processes (see also, Schunn, Lovett & Reder, 2001), to allow for the discovery of new strategies. These models are generally described as models of the strategies used by children, but predictions derived from the ASCM have been successfully tested with adults in a study by Siegler and Lemaire (1997). They used the choice / no-choice method, to compare the choices of twenty-year old and seventy-year old participants when allowed, or not allowed, to use different strategies including, a calculator, mental calculation or pencil and paper in the solution of multidigit multiplication problems. Choice of strategy improved performance for both groups, suggesting that strategy preference and choice is normal in the solution of such problems.

Fuson (1997) and Thompson (2000) both provide guides to arithmetic strategies for multidigit arithmetic, examples include; *counting on* – see Groen and Parkman above, *overshooting/compensating* – e.g. 86 + 47 as 86 +50 –3, *partitioning* – e.g. 63 + 56 as (60 + 50) + (3 + 6), and *decomposition/regrouping/mental carrying* – e.g. 38 +26 as 14 units and 5 tens, 4units + (5 + 1) tens. It is clear, that appropriate strategy choice, from these or other examples, can reduce cognitive load, as this is demonstrated in the following problem solution. The problem 29 x 12 if solved conventionally, might involve calculation processes leading to 8 + (40 + 10) + 290, whereas restating the problem as (30 x12) –12 = (300 + 60 – 10) – 2, appears to entirely remove the need for the carry operation, which might also be achieved by direct retrieval of the 'table fact' 12 x 9 = 108, giving 108 + 240. Strategies such as the ones described here can obviously be taught, and McNamara and Scott (2001) provided evidence that a taught mnemonic *chaining* strategy allowed participants to improve on measures of WM span. Lucangeli, Tressoldi, Bendotti, Bonanomi and Siegel (2003) reported different strategy use across arithmetic operations, and for written and mental arithmetic, with children reporting the formation of a mental image, which they operated on from right to left, referred to as the *mental algorithm* (MA) strategy. This MA strategy suggests that the image formed is columnar, as it is difficult to see why any other image format would be operated on right to left, when normal reading processes, in this culture, are left to right.

Shanahan and LeFevre (2003) investigated strategy choice in an adult sample for multi-digit additions (two-digit plus one-digit) presented in vertical (columnar) or horizontal (linear) format, with either one carry or no carry, and with either the single-digit operand presented first, or the two-digit operand presented first. Results indicated a preference for processing the two-digit operand first, and differences in *retrieval* and *decomposition* strategies for carry and no carry problems, decomposition being preferred for carry problems, and retrieval for no carry problems. They concluded that a preliminary decision is made about strategy choice depending upon the need for the carry operation, and that this might explain why problems with carries have longer response times.

Reichle, Carpenter and Just (2000) used functional magnetic resonance imaging (fMRI) to establish the neural bases of strategy selection in a sentence – picture verification task, and reported that participants demonstrated a preference either for verbal or visual strategies, and that the use of these preferred strategies reduced cognitive load, in comparison to non-preferred strategies. The response time measurements in studies of cognitive arithmetic might therefore represent strategy choice differences for each problem rather than WM involvement. This could potentially undermine the logic of dual-task studies, for if strategy choice is dependent upon which dual-task is presented with any particular problem, then problem response times could simply indicate differences due to this strategy choice. Hecht (2002) in a dual-task study of the verification of simple additions addressed this point by having participants report their strategy choice for every problem presented. Results indicated that adults used a variety of solution strategies including counting, decomposition and retrieval, but that the choice of these strategies was not dependent upon the availability of WM resources – which means that participants did not tend to alter their choice of strategy for different dual-tasks such as articulatory suppression and random letter generation. Strategies other than retrieval had longer response times, although neither AS nor RLG affected retrieval response times, suggesting that retrieval does not load WM.

2.5 Cognitive models of arithmetic

Several cognitive models of arithmetic have been advanced to explain the processes involved in the encoding, calculation and production stages of arithmetic problem solving. Some of the models rely on general processing resources, while others postulate processing resources specific to mathematics, but however specified, these models have to be able to deal with arithmetic problems from any of the arithmetic operations, whether these do or do not involve the carry operation, whether presented in different sensory modalities – verbal, visual, - or whether involving within-modality differences - Arabic digits, number words, horizontal or linear format. The Baddeley and Hitch WM model allows for phonological codes, visuo-spatial codes, and an amodal code for the newly postulated episodic buffer, so it is informative to see which representational codes are required, and how this might signify WM involvement, in some of these cognitive models. It is also important to see if and how the different models explain individual differences - can tasks be represented, or indeed solved, in different ways by different individuals, or is there one, and only one way of reaching problem solution?

2.5.1 The triple-code model

The triple-code model (Dehaene, 1992, Dehaene & Cohen, 1995), as its name suggests, is based upon three different representational codes for the manipulation of numbers, as follows; the *visual Arabic number form,* where numbers are represented on an internal visuo-spatial scratchpad as strings of digits, preserving the spatial relationship between the digits – the *verbal word frame* representation, where numbers are given as *syntactically organised* sequences of words, e.g. 23 as Tens {2} Ones {3}, or 13 as Teens {3}, and the *analogical magnitude* representation of quantity.

Figure 2 presents a diagram of the model, based on Dehaene (1992) and Dehaene and Cohen (1995) mapped to the brain's right and left hemispheres.

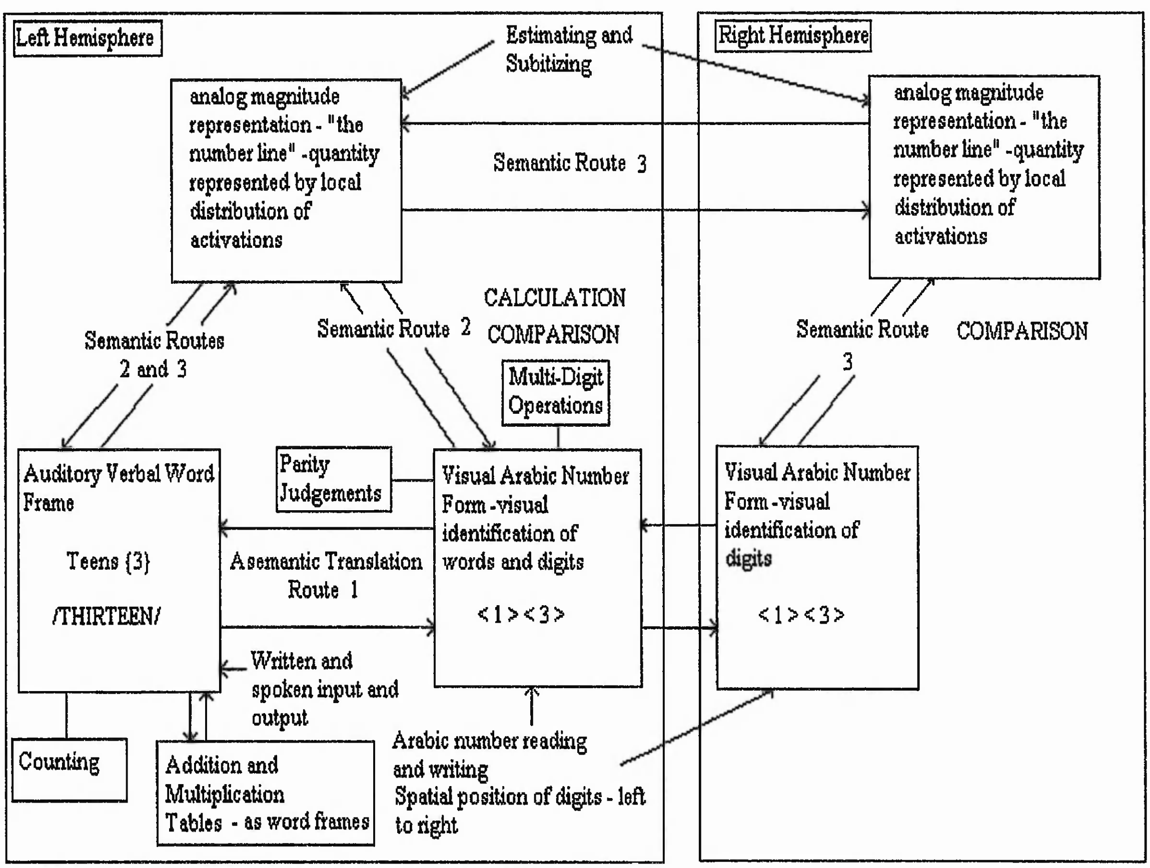

Figure 2. The triple-code model.

Verbal word frame representations are not strictly phonological, in the sense of being based purely on sound, but are more abstract verbal representations, which can be used to retrieve the graphemes or phonemes, associated with a particular number, while arithmetic facts are considered to be rote-memorised associations between such verbal word frame representations. Visual Arabic number forms and verbal word frames do not contain semantic information about the magnitude of the number represented, or about the relationship of that number to other numbers. The provision of semantic information is the function of the *analogical magnitude* representation, which is described as a *number line*, with numbers as patterns of activation spaced at different distances from left to right along this line. The patterns of activation so formed obey the Weber-Fechner law, which states that perception is based on the ratio of two magnitudes, so that while small numbers have relatively precise locations, large numbers are imprecisely located, e.g. the ratio of the numbers 3 to 6 is 1:2, whereas 1003 to 1006 is almost unity - relationships between numbers consisting of the relative distance or overlap between activations.

Figure 2 provides greater detail of the difference in components between the right and left hemispheres, the main differences being that the right hemisphere visual Arabic number form has no direct link to language processing, the visual system in the left hemisphere can recognise all single and multi-digit numerals, and printed words, whereas the visual system in the right hemisphere is possibly better for quantity judgements, but is generally more limited, especially for words. As a result the right and left hemispheres both have arithmetic processing capability, but only the left hemisphere contains all the components necessary for a stand-alone fully functioning arithmetic processing system.

There are three routes, or *translation paths*, within the model, see figure 2, which are described as:

(1) - The asemantic translation route, where word sequences are identified in the verbal word frame for digits represented in the visual Arabic number form, and vice versa, without any reference to semantic information in the analog magnitude representation – this is equivalent to the *reading without meaning* route in models of reading comprehension, e.g. (Ellis, 1993).

(2) - The left hemispheric semantic route, left visual Arabic number form to left analog magnitude representation to verbal word frame.

(3) –The right hemisphere semantic route, right visual Arabic number form to right analog magnitude representation to left analog magnitude representation to verbal word frame.

The route used depends upon the format in which numbers are presented, and the model proposes that the initial task is to form an appropriate representation in one code, which can then be transcoded into either of the other codes. The overall assumption being that different numerical tasks require specific input and output codes. The right hemisphere can carry out numerical tasks such as numerical comparison or approximation, but only the left hemisphere has access to arithmetic facts. Tasks involving multi-digit calculations are considered to require complex interaction between all representations, but especially between verbal and visuo-spatial representations, the verbal system having to retrieve arithmetic facts and the visual system maintaining the *on-line spatial layout* of digit identities due to *the spatial organisation of calculation algorithms*. These calculation algorithms may be routinised, and although Dehaene and Cohen did not explicitly state this, it would appear that calculations are preferentially presented in a learned, presumably columnar format. Simple arithmetic facts can be retrieved directly without recourse to semantic information, but the analog magnitude representation is thought to be involved when strategies such as partitioning are used instead of direct retrieval, e.g. 9 + 7 = 10 + (7 – 1), when quantities have to be compared on the number line, which is described as the use of *semantic elaboration.* The analog magnitude representation is described as a preverbal system of arithmetic reasoning, as in *subitizing*, the rapid enumeration, evident in humans and other animals, of small sets of objects without counting, whereas the auditory verbal word frame is considered to be part of general purpose modules for language processing, and this was explicitly linked to the Baddeley and Hitch WM model by Dehaene and Cohen (1995) when they stated:

> "Operands and intermediate results may have to be stored in working memory via the articulatory loop or via the visual number form and other areas forming the "visuo-spatial scratchpad." Dorsolateral prefrontal circuits contribute to the planning, sequencing and controlling of successive operations." (p.104).

2.5.2 The encoding complex hypothesis – interactive model

The *encoding complex* hypothesis, (Campbell & Clark, 1988) can be thought of both as a set of principles to modify other models, or as a separate model based on these principles, Campbell (1994). Figure 3 presents a diagram of the interactive model, adapted from Dehaene (1992).

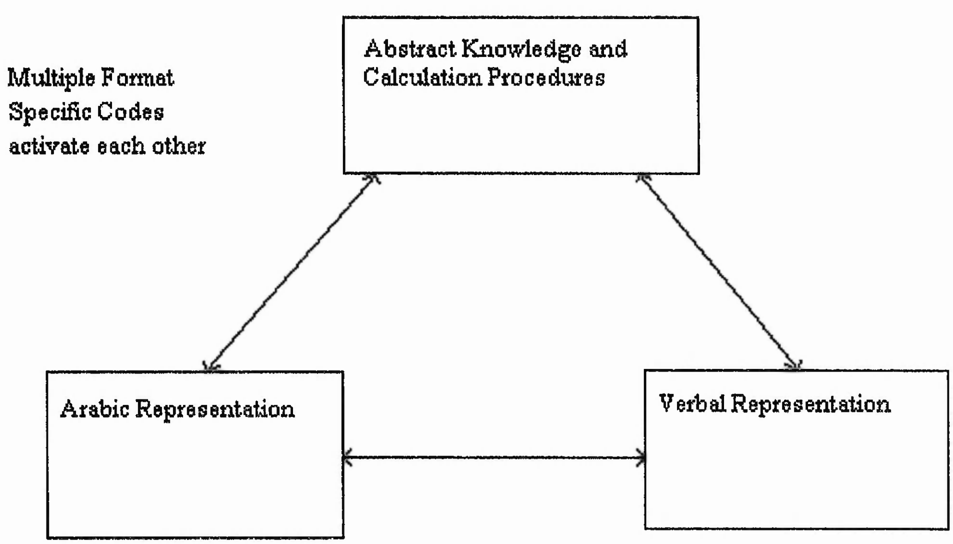

Figure 3. The encoding complex hypothesis – interactive model

The three principles of the model are firstly, that modality specific processes, structures and codes have the primary role in number processing, rather than any abstract code. Secondly, the surface form or notation used to present numerical problems has a direct influence on the processes, codes and strategies used to solve the problems. Finally, any function in numerical cognition might involve the use of several different codes depending upon the type of arithmetic operation and the format of presentation. This means that in contrast to the triple-code model, which specifies a phonological/ verbal word frame access, arithmetic facts in the interactive model can be accessed by both visual and verbal representations. Consequently, there could be differences in the relative efficiency of these representations depending on the frequency of occurrence of particular problem formats, for example, seven x eight encountered less frequently than 7 x 8. It is interesting to speculate whether the relative frequency of occurrence of horizontal (linear) and vertical (columnar) problem presentations might lead to similar differences. Noel and Seron's (1992) *preferred entry code hypothesis* offers a variation on the interactive model by suggesting that individuals are divided into auditory and visual types, so that those who experience greater levels of imagery will use Arabic digit visual presentation to access number facts, whereas those who use predominantly verbal coding will use verbal representations to access number facts, and indeed, Campbell (1994) questioned the applicability of all models specifying a unique access code:

> "...it is unlikely that a single specific architecture applies generally because the codes and processes involved in number processing may vary with individual differences in culture, pedagogy, strategy, and other idiosyncratic factors, including brain injury." (p.37).

2.5.3 The abstract-modular model

The abstract-modular model of, McCloskey, Caramazza and Basili (1985) is illustrated in figure 4, adapted from McCloskey (1992):

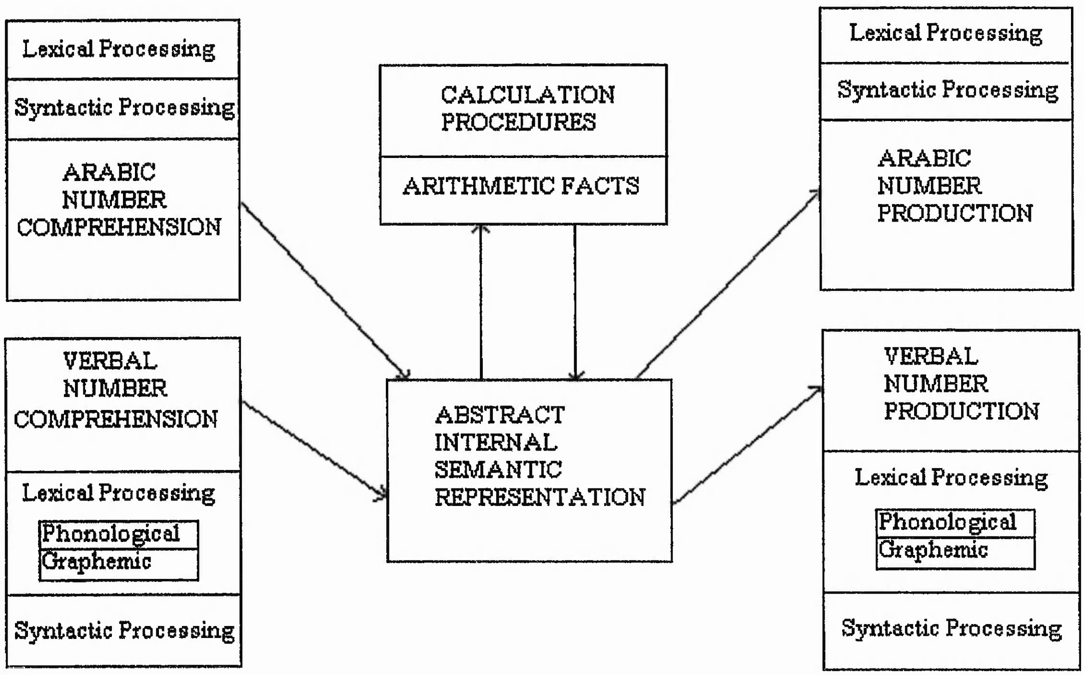

Figure 4. The abstract-modular model

This model differs from the previous models in that there are no routes, asemantic or otherwise, between verbal and visual comprehension systems, or between verbal and visual production systems. The modules within the model are independent rather than interactive, so that the abstract semantic representation used in the calculation process can neither guide input processes nor modify output processes. This means that the same problem, whether presented verbally or visually, will lead to the same abstract internal semantic representation, and that any differences in the format of presentation can only have an impact on the encoding process during the comprehension stage. The model works by converting verbal or visual input into a *syntactic frame*, as in the following example, derived from the discussion of the calculation of the same multiplication problem in McCloskey (1992):

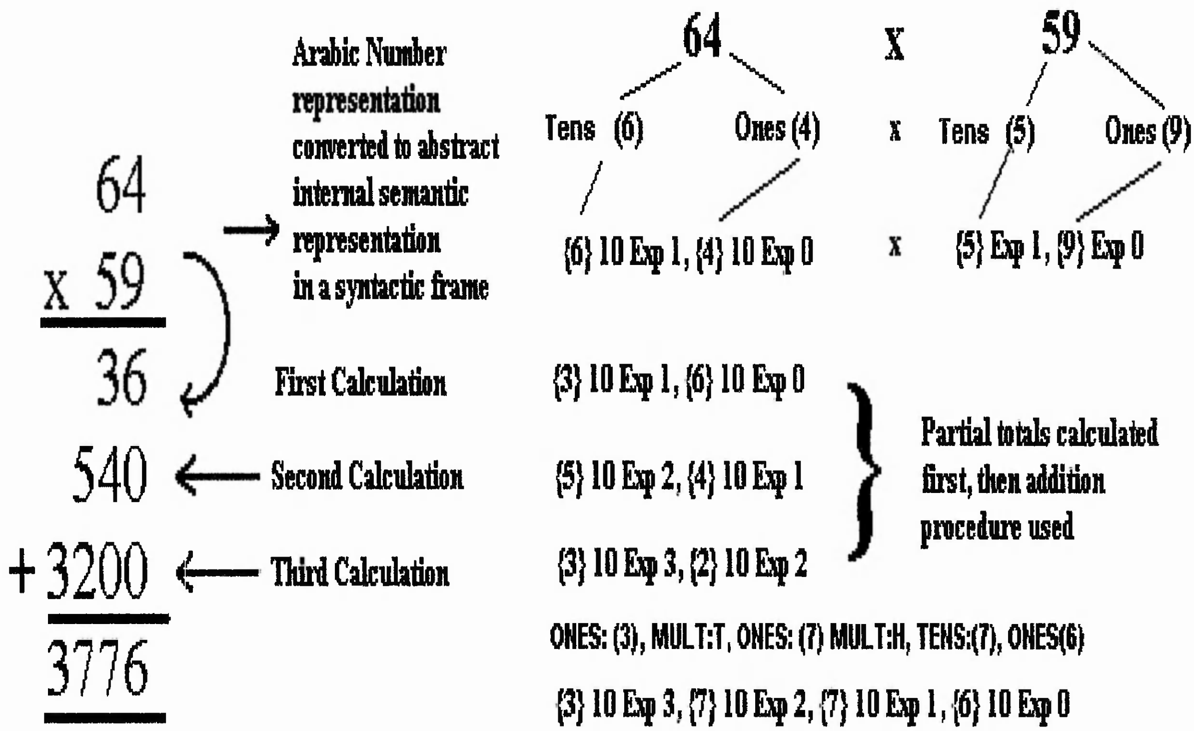

Figure 5. The use of an abstract internal semantic code in multi-digit multiplication

The *filled syntactic frame* holds not only the number as powers of ten, but also the information necessary to produce the written or verbal form of the number, as in Figure 5, where the number 3776 is also given the information 3 ones x thousand (three thousand), 7 ones x hundred (seven hundred), 7 tens (seventy), 6 ones (six). The columnar representation in Figure 5 is not arbitrary, but is based on the following statements, concerning the same multiplication problem, in McCloskey (1992):

> "This procedure, which provides an ordered plan for the solution of multiplication problems, would call first for the processing of the digits in the rightmost column (i.e. 4 and 9)...The multiplication procedure would then call for the ones portion of the product to be written in Arabic form beneath the rightmost column of the problem...Processing would continue in this fashion until all partial products had been computed. At this point the addition procedure would be called, and the partial products would be summed and written under control of this procedure." (p.115).

It is not clear whether McCloskey intended to imply that all multi-digit multiplication requires obligatory representation in columnar format, or whether this is how the model would deal with problems already presented in columnar format. It is clear, however, that the filled syntactic frame *reads* from left to right, and the above quote suggests columnar format for the presentation of partial totals, although it is unclear where columnar information is, or can be, stored without recourse to a visuo-spatial store. If partial totals are 'written in Arabic form', is this done in the Arabic Number Production module, and if so, how is this information brought back into the abstract internal semantic representation required for the addition process?

2.5.4 The independent number module and the COMP model

Butterworth (1999) presented the case for a mathematical processing system that is independent of general cognitive processes, and dependent primarily on a 'number module' thought to be contained within the inferior lobule of the left parietal lobe of the brain – possibly in relation to the representation and control of our fingers, as indicators of magnitude. Dehaene (1997) also identified the importance of the inferior parietal lobe in representing magnitude and in calculation, but made the point that to claim that calculation rests on the inferior parietal region would be to introduce a 'neo-phrenologic' framework of cognitive functions, and in relation to the function of the inferior parietal lobe stated:

> "...this region probably contributes to a narrow process: the transformation of numerical symbols into quantities, and the representation of relative number magnitudes. It does not play a generic role in arithmetic since damage to it does not necessarily affect the rote retrieval of simple arithmetic facts ...nor the rules of algebra...nor the encyclopaedic knowledge of numbers." (p.217).

Butterworth presents evidence from a variety of neurological case studies supporting the view that mathematical ability can remain essentially intact when reasoning, short-term memory, long-term memory and language abilities are selectively or collectively impaired. Figure 5, adapted from Butterworth (1999) gives the 'functional architecture' of this mathematical system:

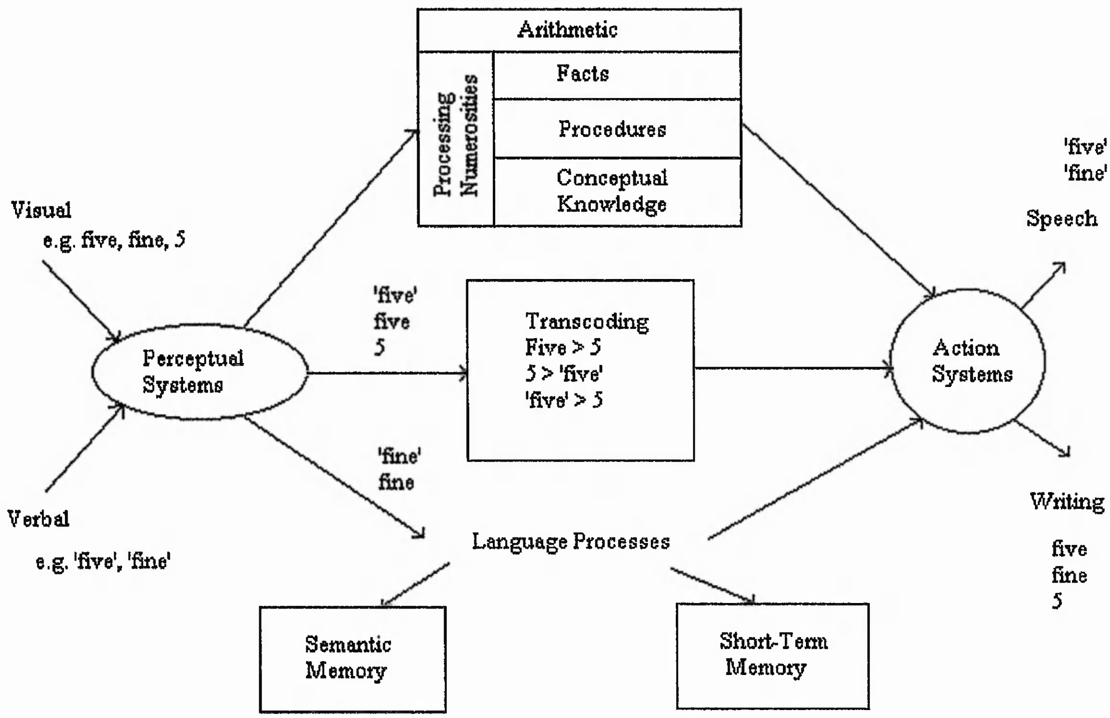

Figure 6. Butterworth's 'functional architecture' of arithmetic.

This model clearly has an asemantic transcoding route, and separate routes for arithmetic processing and language processing, but interestingly, short-term memory or working memory (Butterworth makes no distinction between the two) is not involved in arithmetic processing, and this is made explicit in the discussion of the abilities of Mr. Morris, the same patient (MRF), described in Butterworth et al. (1996):

> "Of course, Mr. Morris had to remember the problem in order to work out the answer, but we think that rather than use the general-purpose, verbally coded short-term memory, arithmetic uses a special-purpose memory." (Butterworth, 1999, p.178).

The nature of this 'special-purpose' memory is not described, but presumably, as well as holding problem information, it also acts as a long-term store for conceptual knowledge, arithmetic facts and procedures, and thereby possesses some of the characteristics normally ascribed both to short-term or working memory, and to long-term memory. Butterworth, Zorzi, Girelli and Jonckheere (2001) provided evidence that arithmetic addition facts are organised in terms of the cardinal magnitude of the addends, as *half-tables* of *max* and *min* values, e.g. 5 + 4 = 9, but not 4 + 5 = 9, which they state as consistent with educational practice in China, where children are only provided with half-tables. The COMP model is proposed to explain addition fact retrieval, as illustrated in the diagram below (adapted from Butterworth et al., 2001).

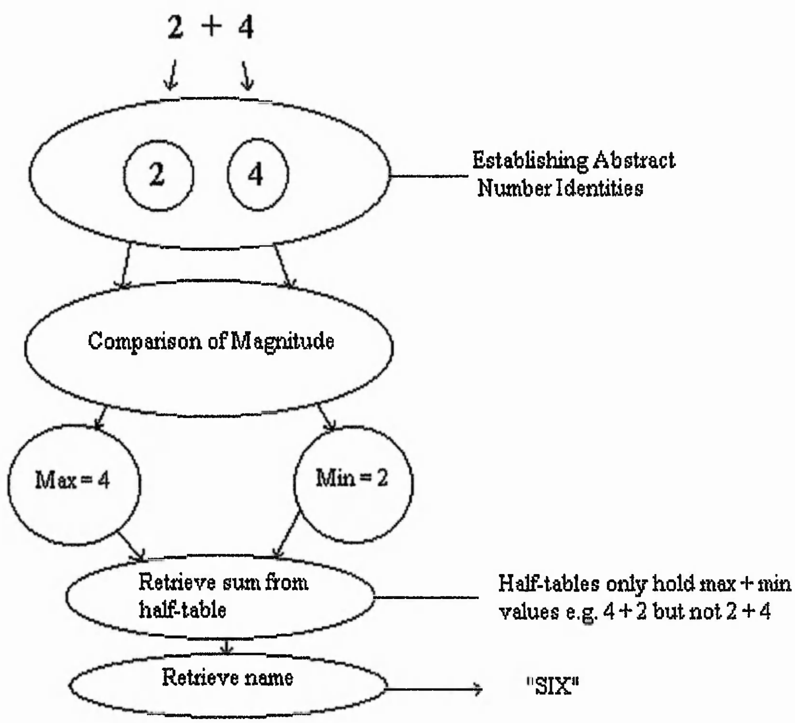

Figure 7. The COMP (comparison) model of arithmetic addition fact retrieval

Butterworth et al. make the point that if arithmetic facts are stored in terms of max and min cardinal magnitudes, then this is consistent with domain-specific organization rather than the domain-independent organization of arithmetic facts as association strengths or verbal associations, and this is taken as evidence against such models as that of Dehaene and Cohen (1995), and as evidence for a separate independent cognitive module for arithmetic.

Chapter Three - Evidence for the cognitive models and WM involvement in cognitive arithmetic

Studies in cognitive arithmetic can encompass a number of perspectives, but one broad classification stems from grouping them as neuropsychological (fMRI and PET scans) or experimental (response times, error rates, spans). Studies with a neuropsychological basis often complement those with a more traditional experimental basis, but individual studies can sometimes reflect a particular focus – some studies are primarily concerned with providing evidence for or against particular cognitive models, while others are concerned with the factors underlying the performance or developmental differences of various groups. The identification of WM involvement can either be a central aim, as in various dual-task or span experiments, or something of a peripheral afterthought, as when the results are consistent or inconsistent with a particular interpretation of WM. In addition to the normal concern with representative sampling, the ability to generalise from these studies is heavily dependent on a number of other factors, including the type of arithmetic task chosen, and how this task is presented, manipulated and measured. The domain of arithmetic is relatively large, and the results for any particular arithmetic task are unlikely to reflect the results for all possible arithmetic tasks – typically, the arithmetic task used in any given study might be simple or complex, single-digit or multi-digit, easy or hard, with or without the carry operation, and from one or more of the four arithmetic operations – addition, subtraction, multiplication and division. The arithmetic task might also be presented as a production task (answer required but not given) or verification task (answer given, true or false decision required), in visual or verbal input modality, as Arabic digits or as number words, continuously or for variable durations, in various spatial orientations, and with variations in phonological or visual similarity, with performance measured, either in verbal or visual output modality, and in terms of differences in WM span, response times and error rates. Consequently, in the review of previous studies that follows, these and other variations will be identified when relevant, and studies will be grouped to reflect the importance of particular factors.

3.1 Neurological studies

The neurological studies reported here have two primary objectives, the provision of evidence, both for and against arithmetic facts being stored in a particular code, and for and against different neural routes for the different arithmetic processes. The models reviewed in chapter three suggested four possibilities for the storage of arithmetic facts; the rote learned verbal word frames of the triple-code model, the multiple code representations of the encoding complex – interactive model, the abstract semantic codes of the abstract-modular model, and the max-min cardinal magnitudes of the COMP model. The COMP model refers only to addition facts rather than to all arithmetic facts, and the encoding complex – interactive model already proposes access via different representational codes, so the neurological studies tend to be based on specific predictions about representational codes derived from the other two models.

A central claim of the triple-code model, Cohen and Dehaene (2000) is that multiplication facts are retrieved as rote learned verbal word frames. Addition facts are also normally retrieved as rote learned verbal word frames, but addition could also be achieved in a similar manner to subtractions, which, because they have no store of rote learned facts, have to be solved by *'strategical quantity manipulations'*. The verbal word frame involves a different neural route to the quantity route, multiplication causing greater left hemisphere activation and subtraction causing activation in both hemispheres, Dehaene et al (1996), Chochon and Cohen (1999), Cohen et al (2000). This leads Cohen and Dehaene (2000) to the specific prediction:

> "It should never be possible to find a patient with impaired multiplication *and* subtraction, yet with relatively preserved addition; nor should it be possible to have a selective impairment of addition relative to multiplication *and* subtraction " (p.576-7).

The abstract-modular model postulates potentially dissociable stores of arithmetic facts/procedures, so that there is no reason to suppose that addition could not be selectively impaired in relation to multiplication and subtraction, and the selective impairment of addition is reported in patient FS by Van Harskamp and Cipolotti (2001), which is taken as support for the separation of arithmetic fact stores for each operation, see also Whalen et al (2002). Several studies have shown varied patterns of brain activation during mental calculation, with general but not exclusive support for the triple-code model, e.g. Zago et al (2001), Rickard et al (2000) Pesenti et al (2000), but although Dehaene et al (2004) acknowledge that there are studies that challenge the triple-code model, they suggest that a recent meta-analysis of neuroimaging studies provides stronger support for the triple-code model through the clear identification of the horizontal segment of the bilateral intraparietal sulcus (HIPS) as the site for quantity representation in numerical processing:

> "In the simplest experiments, which involve number detection or comparison rather than more complex calculation, the HIPS is the only region specifically engaged. This suggests that the HIPS region plays a central role in basic quantity representation and manipulation, whereas other prefrontal areas might serve a more supportive role in the management of successive operations in working memory." (p.218).

The HIPS region is automatically activated in response to numerical quantity and is considered by Dehaene to be amodal and language-independent, in that the same activation is seen for spoken numerals, and for written numerals in Arabic digit or word format, and to be accessible through a variety of symbolic or non-symbolic codes. Dehaene also reports that activation in the HIPS region has to be distinguished from activation in the Angular

Gyrus, suggesting that the latter area is activated by language tasks such as digit naming, and this gives rise to a new schematic formulation of the triple-code model, which identifies the neuroanatomical locations and pathways used for various arithmetical tasks, as is shown in figure 8, (adapted from Dehaene, 2004):

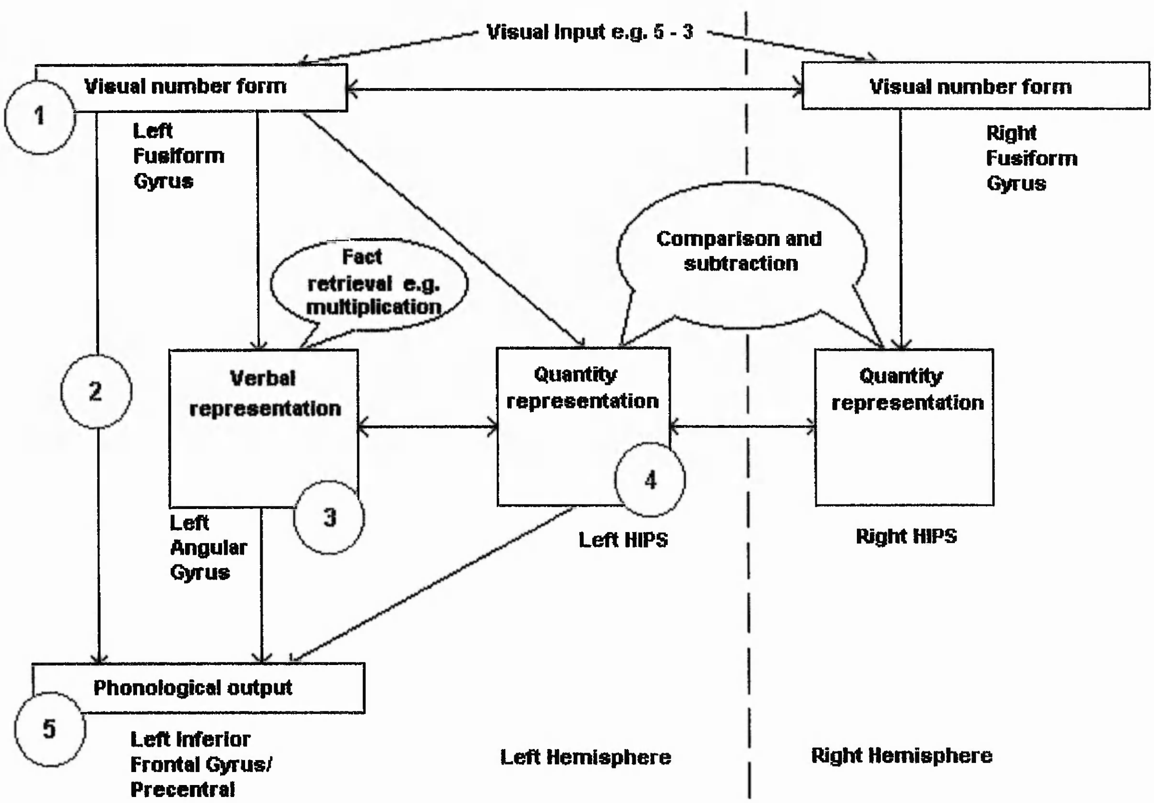

Figure 8. Neuroanatomical routes for arithmetic processing in the triple-code model

The numbered circles in the diagram refer to *functional lesion sites* associated with certain neuropsychological dissociations; a lesion at site (1) being associated with *Pure Alexia* – an inability to read, which in terms of arithmetic and of the model, would imply an inability to read numbers and to multiply, but would not prevent numerical comparison or the ability to subtract. A site (2) lesion leading to *Phonological Dyslexia* – which is taken here to imply an inability to read numbers, rather than the inability to read unfamiliar words, (Ellis, 1983), but with multiplication, subtraction and comparison being unaffected, while lesions at sites (3) and (4) are considered as a possible explanation for a double dissociation between subtraction and multiplication with preserved ability to read numbers. Finally, a site (5) lesion, removing phonological output, might offer an explanation of those demonstrating preserved ability to solve written calculation but failure to produce oral solutions.

Interestingly, Dehaene (2004) reports that the deficits in arithmetic processing found in individuals with Turner's syndrome are associated with the abnormal shape and depth of the intraparietal sulcus (IPS), with missing *grey matter* in the left or right IPS and with an inability to recruit the IPS as arithmetic task difficulty increases, which links with the experimental studies described in the next section.

3.2 Experimental studies

3.2.1 Studies of developmental difficulties and WM involvement

Temple and Sherwood (2002) measured short-term memory span, speed of counting and speech, and speed of access to lexical items, including number facts, in children with Turner's syndrome compared to normal controls. Speed of access providing the only differences, which Temple and Sherwood take as support both for modular theories of arithmetic cognition in general, and for Butterworth's (1999) proposal of a 'core module' of arithmetic skills underlying the development of arithmetic skills. Bull and Johnston (1997) in a study of two groups of high and low ability 7 year olds, found that poor arithmetic ability was best explained by a speed of processing deficit in the ability to automate basic arithmetic facts, and that measures of short-term memory did not differ between the two groups. A second follow-up study, Bull, Johnston and Roy (1999), using a sub-set of participants from the (1997) study, investigated CE and VSSP involvement in arithmetical skills through performance on the Wisconsin Card Sorting Test (WCST) and Corsi Blocks, and found no difference in visual sequential memory, but they did find a significant correlation between poor arithmetic performance and WCST perseveration measures, indicating an inability to switch solution strategy, which is one of the tasks of the CE.

Geary, Hoard and Hamson (1999) studied children (aged ~ 7 years) classified as at risk for a learning disability in reading (RD), mathematics (MD) or both reading and mathematics (RD/MD), in comparison to normal controls on tasks assessing number comprehension and production, counting, arithmetic, working memory and retrieval from long-term memory. The RD/MD group, and some, but not all, of the MD group had significantly poorer performance on backward digit span, and this was taken to indicate that the children had difficulty retaining information in the phonological loop while manipulating or attending to other information. Geary et al. suggested that one possible explanation to account for the differences between the MD and MD/RD groups is that the MD group might have CE impairment affecting the ability to hold and manipulate information, while the MD/RD group might have a phonological loop deficit.

In a follow up study, Geary et al. (2000), reported a significant difference in backward digit span for MD/RD and normal groups, but suggested that the *apparent working memory deficits* of children with learning difficulties might relate more to IQ than WM itself, and that simple digit span measures might be incapable of capturing the relationship between learning difficulties and WM.

McLean and Hitch (1999) investigated WM impairment in children with specific arithmetic learning difficulties in comparison to age-matched and ability-matched controls, and reported that in comparison to age-matched controls the poor arithmetic group had impairment to spatial working memory and executive processing, but not phonological WM, whereas in comparison to ability-matched controls, the poor arithmetic group only showed a deficit on the executive task designed to assess ability to hold and maintain information in long-term memory. In summary, these studies collectively identify the involvement of the CE, the PL and the VSSP in arithmetic learning difficulties, but not with any consistency between studies.

3.2.2 Studies supporting the cognitive models of arithmetic

As with the neurological studies, one of the aims of experimental studies has been to provide evidence identifying the access code for retrieving arithmetic facts. This has revolved around the identification and interpretation of task x format interactions because of the specific predictions arising from the cognitive models, only the encoding complex – interactive model predicting such interactions. Campbell (1994) has provided evidence of such task x format interaction. He has shown slower response times and greater error rates for problems such as (four x five) compared to the equivalent Arabic digit problems (5 x 4), with frequencies of *operation* and *operand-naming* errors varying with arithmetic operation and presentation format, operation errors (e.g. 2 + 9 = 18, 4 x 8 = 12) more frequent for additions than multiplications and greater in Arabic digit than word format, whereas operand-naming errors (e.g. 2 + 9 = 9, 4 x 8 = 48) were also more frequent in additions than multiplications, but greater in word than Arabic digit format. This latter point being explained as the possible interaction of number fact retrieval processes and number reading processes, e.g. 6 x 9 *read* as a two-digit number 'sixty-nine'. The interpretation offered by Campbell is that such task x format interaction provides evidence that the access code for arithmetic facts is neither abstract nor solely verbal, thereby undermining both the abstract-modular model and the triple-code model. Noel, Fias and Brysbaert (1997) reported the same format effects in a multiplication task as found by Campbell, but considered that they were due to encoding time differences between word and digit numerals. Furthermore, they found no difference in the error types for French and Dutch participants, which led them to question the hypothesis that reading processes interfere with arithmetic-fact retrieval processes, and to suggest that arithmetic facts are represented in a single medium. Campbell (1998) suggested that an alternative analysis of the error data in the Noel et al. (1997) study did in fact show support for the *language-specificity* of number fact memory, although Noel, Robert and Brysbaert (1998) question Campbell's re-analysis of the Noel et al. (1997) study, and confirm their view that there is no evidence for a modality-specific arithmetical-fact network.

Campbell (1999) investigated the format x *problem size effect* interaction (problem size effect is greater when arithmetic problems are presented as words) by introducing a *display* factor, operands either appeared simultaneously or with an 800-millisecond delay between first and second operand. The rationale being that if format effects operate solely at the encoding stage, then there should be a format x problem size effect x display interaction, because the 800-millisecond delay would allow the first operand to be entirely encoded before the presentation of the second operand. However, no such interaction was found although the experiment was considered to have sufficient power to identify the interaction if it existed. Campbell and Fugelsang (2001) reported a format x solution strategy interaction such that word problems involved greater use of calculation strategies than digit problems, which is taken as additional support that format has its effect after the encoding stage. Blanken, Dorn and Sinn (1997) reported *inversion errors* in patient AT, stemming from the German Arabic number system. For example, they found that although AT should have read 26 as *sechsundzwanzig* (six and twenty), he actually read 26 as *zweiundsechzig* (two and sixty). They claim that this would not be possible given the abstract representation of 26, (2) Exp 1 (6) Exp 0, in McCloskey's abstract-modular model, for this could only be inverted as *zwanzigsechs* (twenty and six) and never *zweiundsechzig* (two and sixty).

Chincotta, Underwood, Abd Ghani, Papadopoulou and Wresinski (1999) investigated the role of WM in a dual-task study of the *numeral advantage effect* – memory span for digit numerals is greater than memory span for digit words. They found that spatial tapping was the only task to remove the numeral advantage effect, which led them to suggest that Arabic digits are maintained more efficiently in the visual store of the VSSP. Although this could explain the mechanism underlying task x format interaction, it does not indicate whether this occurs at encoding or a later stage.

Brandimonte, Hitch and Bishop (1992) reported that articulatory suppression during item-learning significantly improved performance on a later imagery task for easily named items. They explained this as being due to the availability and preferential use of verbal codes, so that when AS reduced the availability of verbal codes, participants were forced to pay more attention to the visual characteristics of the stimulus items. It is interesting to speculate whether this might also be true of arithmetic processing. If Arabic digits can have direct access to the store of rote-learned, and presumably easily-named, arithmetic facts, then articulatory suppression might be assumed to have little if any disruptive effect, possibly even improving performance, on arithmetic tasks.

Berch, Foley, Hill and Ryan (1999) demonstrated support for the 'number line', analog magnitude representation in the triple-code model. They showed that children as young as Grade 3 (~9 years old) demonstrate the SNARC (Spatial-Numerical Association of Response Codes) effect – small numbers responded to faster with the left hand than the right hand. Although the SNARC effect is evident in adults, the effect was attenuated at Grades 6 and 8 by a MARC (Markedness Association of Response Codes) effect – odd numbers associated with faster left-key than right-key responses, and the reverse for even numbers, which is explained as the linguistic association of 'left' and 'odd', and 'right' and 'even'.

Whetstone (1998) reported the case of patient MC, who demonstrated a numerical processing impairment only for multiplication fact retrieval, but subsequently relearned three sets of multiplication facts in different formats, spoken, written number and Arabic digit, before being tested for all of these multiplication facts in all formats. The rationale for the study was that if arithmetic facts are represented in a single code, then retraining will restore that code, and there should be no effect of test format, as all formats will access the same code, alternatively, if arithmetic facts are stored in multiple codes, and all of these codes are lost, then facts learned in one code should not be accessible via other codes, at least not without practice. Results showed no effect of input format, although response times were faster when test format matched training format, and this is taken as support for arithmetic fact representation in a single code. Whetstone gives as an example the problem "six times eight", retrained in spoken format, which MC could not answer prior to retraining regardless of input format, but following retraining could answer when tested in any format. However, it is unclear whether learning "six times eight" in spoken format also re-activates the associated visual and written representations, or whether MC did in fact practice multiplication problems in different formats.

To summarise, it seems that the interpretation of format x task interactions is still controversial, and although they seem to provide clear evidence of multiple code access to arithmetic facts, it might be possible to give explanations of these interactions solely in terms of encoding processes, and this is reflected in DeStefano and LeFevre's (2003) review, which offers a qualified statement that the abstract code model is *probably* not adequate for the full range of arithmetic processing.

3.2.3. Studies supporting WM involvement in arithmetic processing

There are two central studies, Heathcote (1994), and Logie, Gilhooly and Wynn (1994), which, individually or jointly, serve as a focus in most investigations and reviews of WM and cognitive arithmetic, (see Ashcraft, 1995, Noel et al., 2001, DeStefano & LeFevre, 2003), and as such, are deserving of more detailed description and analysis.

Heathcote (1994) reported a series of three experiments. The first, a dual-task experiment utilising articulatory suppression, spatial tapping and visual interference as secondary tasks, investigated the role of visuo-spatial WM in the mental addition of pairs of computer generated three-digit numbers requiring either zero or two carries, and with discontinuous visual or auditory presentation, recording or not recording partial totals. Solution latencies demonstrated significant main effects for presentation modality, secondary task, number of carries and partial result recording, a carry x secondary task interaction and a partial result x secondary task interaction – the recording of partial results giving a greater reduction of solution latencies under articulatory suppression compared to either spatial tapping or visual interference. All three secondary tasks significantly increased solution latencies, but this was modified by a significant visual interference x carry x presentation modality interaction - visual interference increasing latencies in carry problems for both auditory and visual presentation, but only increasing latencies for no-carry problems presented auditorily. A significant spatial task x carry interaction, with spatial task interference being greater for carry problems was also reported. Further analysis showed that the effect of the carry operation was greatest under spatial tapping and least under articulatory suppression. Results for error rates provided significant main effects of secondary task, partial result recording and carry requirement, with no interactions, and further analysis showed that overall, all three secondary tasks led to greater error rates, although visual interference only produced greater error rates for carry problems. On the basis of these results, Heathcote suggested that sub-vocal rehearsal is involved in the maintenance of partial results and in the retention of problem information, and that spatial WM is involved in the retention of carries.

The role of spatial WM was further investigated in a second, single-task experiment, manipulating horizontal (lateral) and vertical (columnar) presentation of pairs of three-digit addition problems, and recording partial totals for all participants, results indicating that horizontal format increased both latencies and error rates. Heathcote suggested that these results, and the results of a previous unreported experiment, provided evidence that problem information is retained in columnar format in the VSSP, and that this might be the case even if problems were not originally presented in columnar format.

To explore the possibility that problem information might be retained in the VSSP in a spatial but non-visual form, a third dual-task experiment, using articulatory suppression as a secondary task, investigated the effect of the visual similarity of digits, in two-carry problems presented either auditorily or in visual columnar format, with partial totals reported by all participants. The purpose of the articulatory suppression task here was to prevent participants from using phonological coding, thereby forcing them to rely more on visual coding. The predicted outcome was that visually similar digits would have increased latencies and error rates. The purpose of presentation in visual and auditory format was to test the possibility that visual confusion occurs only at encoding – if visual similarity effects occur for both presentation formats, then this would suggest that the effect is not at the encoding stage. Solution latencies demonstrated no effect of visual similarity, but error rates were significantly greater for visually similar digits, but only under articulatory suppression, with no differences due to presentation modality. On the basis of these results, Heathcote suggested that digits are visually presented in the VSSP, and that the visual similarity effect is not entirely due to visual confusions at encoding.

Heathcote's (1994) study has attracted criticism for a number of reasons; Noel, Desert, Aubrun and Seron (2001) point out that results were collapsed for correct and incorrect trials, the power of the analysis was low due to small sample size, and participants would have experienced difficulty maintaining articulatory suppression while reporting partial results. DeStefano and LeFevre (2003) point out that the period for visual presentations was 1070 milliseconds in comparison to the period for auditory presentation of 4000 milliseconds, and although Heathcote measured solution latencies from *offset* of problem presentation, the brief visual presentation might have forced participants to use phonological coding rather than visuo-spatial coding.

Logie, Gilhooly and Wynn (1994) reported two dual-task experiments investigating the role of working memory in mental addition. The first of these experiments presented each of four groups of 6 participants with a total of 40 problem sequences. Each sequence consisted of a series of two-digit numbers, presented via headphones for a period of 20 seconds, and required participants to add the sequence of numbers and maintain a running total for presentation at the end of the sequence. The amount of numbers in the 20-second sequence varied in accordance with participant ability as measured by arithmetic span, so that different participants might be presented with sequences of 3, 4, 5 or 6 numbers. Participants in each of four different groups solved 20 problem sequences without a secondary task, and 20 problem sequences concurrently with one of four secondary tasks; articulatory suppression, random letter generation, presentation of irrelevant pictures, and hand movement (spatial tapping). The 40 problem sequences were arranged into two subsets, each subset of 20 problem sequences had 10 sequences with each addition in the series requiring a carry – one of the additions in the sequence requiring a carry in the *tens* rather than the *units* column, and the other ten problem sequences normally requiring only one carry addition over the complete sequence – if the sequence was less than six numbers.

Performance was measured both in terms of the number of errors - incorrect totals given by each participant, and in terms of a mathematical derivation of error magnitude described as percentage error size. Results for number of errors producing significant group differences, a significant effect of dual-task, a significant effect of carries, and a significant interaction of dual-task and group, further analysis showing that only articulatory suppression and random letter generation provided significantly greater number of errors. In contrast, results for error size demonstrated significance only for random letter generation, although articulatory suppression demonstrated marginal significance. Interestingly, analyses of secondary task decrement showed no single versus dual-task differences for the spatial task, but highly significant differences for the articulatory suppression task. The random letter generation task could not be analysed because although performance on RLG as a single-task produced sufficient responses to analyse, performance on RLG as a dual-task did not. Logie et al. interpret the results of this first experiment as providing support for the role of the PL and CE in this verbally presented mental arithmetic task, but with no role for the VSSP.

The second experiment only differed from the first in that each of the numbers for addition was presented visually on a computer screen for a period of 1 second, and the visual interference task was replaced with *irrelevant speech* – two-digit numbers spoken at a rate of one per second, which participants were instructed to ignore. Results for number of errors demonstrated significant main effects for group, dual task and carries, with a significant dual task x group interaction, with further analyses showing that all four secondary tasks produced significant increases in number of errors, with RLG producing the greatest effect, and the other three secondary tasks producing similar effects. The error size results demonstrated a significant main effect of group and of dual-task, and a group x dual-task interaction, with further analyses indicating that RLG was the only secondary task producing a significant effect, and that multiple carry problem sequences produced larger errors than single carry problem sequences. Analyses of secondary task decrement found a significant difference in AS and RLG, but no difference for hand movement (spatial tapping). Further analysis compared single-task performance in the first and second experiments, and demonstrated that visual presentation led to a significantly smaller number of errors and lower error sizes. Logie et al. interpreted these results as consistent with a role for the CE, PL and VSSP in visually presented additions, and overall, suggest that the results of experiments 1 and 2 are inconsistent with McCloskey's abstract-modular model, but consistent with the separate mechanisms for accuracy and approximation proposed by Dehaene and Cohen.

The Logie et al (1994) study can also be criticised in terms of low group size, but it has attracted other criticism; Noel et al. (2001) pointed out the lack of task x carry interactions, which are necessary to demonstrate genuine dual-task interference rather than simple cost of concurrence. De Rammelaere, Stuyven and Vandierendonck (2001) suggested that although the sequential 'running addition' task might involve the CE and PL, this still leaves the question of whether single additions involve the CE and PL, and this involves a further distinction in studies of cognitive arithmetic and WM, that between *complex* and *simple* arithmetic problems. The Heathcote (1994) and Logie et al. (1994) studies used complex multi-digit addition, so further studies using relatively complex arithmetic operations will be explored before considering those using more simple arithmetic.

Adams and Hitch (1997) reported two experiments, one with English children of various ages, and one with German children of similar ages, which measured both addition span and addition speed, for additions which progressively increased in complexity over eight levels, from one digit + one digit to four digit + four digit problems, each level having an easy, hard and carry problem, and which were presented in either verbal or visual format. Results were similar for both experiments, and demonstrated that visual presentation gave significantly greater spans, and that variation in span associated with age and problem difficulty could be explained as a linear function of the speed of adding integers. Adams and Hitch took this as evidence that WM is a *general-purpose resource* supporting children's mental arithmetic.

Furst and Hitch (2000) reported three dual-task experiments designed to *follow-up* on the Logie et al. (1994) study, by re-examining the role of the CE and PL in multi-digit addition. The first experiment presented participants with three-digit + three-digit additions involving one, two or three carry operations, presented briefly (4 seconds) or continuously on a computer screen, and then requiring written answers, with articulatory suppression as a secondary task to load the PL, and recitation of the alphabet – from a randomly chosen starting point, as a secondary CE task. Analysis of solution latencies provided no significant results, and analysis of errors provided a significant main effect of presentation duration, but overall no effect of secondary task, although further analyses revealed that errors were greater under dual-task conditions for brief presentation but not for continuous presentation, and that errors increased with number of carries. Furst and Hitch indicated the surprising nature of these results, in that contrary to Logie et al. (1994), there was no indication that the CE was involved in calculation at all, but they explained this by suggesting that the alphabet recitation task failed to load the CE, and furthermore, they took the significant effect of presentation duration in the error data as indicating support that problem information is stored in the PL, and the lack of dual-task interference with continuous presentation as indicative that the PL is *not highly loaded* when arithmetic facts are retrieved. The second experiment was similar to the first experiment, but the simple recitation of the alphabet was replaced with the more difficult spoken *Trails* task, which required participants to alternate between recitation of the alphabet and naming days of the week, e.g. A - Tuesday, B – Wednesday, C – Thursday, and all problems were presented continuously. Analysis of solution latencies and errors demonstrated significant differences for the Trails task, but no differences between AS and the single-task control, further analysis of errors showed a Trails task x Carries interaction – errors greater for one carry than zero carry problems, and two carry problems showing greater errors than one carry problems. Furst and Hitch identified this as clear support for CE involvement in multi-digit mental addition, and suggesting that the CE is implicated in the carry operation, but as solution latencies and error rates could possibly be explained through carry additions taking longer to solve, and in order to investigate CE involvement in arithmetic fact retrieval, they carried out a third experiment.

The third experiment partially replicated the second experiment, but without AS as a secondary task, and included a measure of errors when the Trails task was performed as a single-task over different time durations, matching the range of solution times in the second experiment. Results for solution latencies and errors were similar to those in experiment two. Further analyses revealed a secondary task decrement when the Trails task was combined with arithmetic, but only for one and two-carry problems not for zero-carry problems, and a minimal effect of duration on Trails task performance. Furst and Hitch suggest that this indicates that the effect of carries on Trails performance is not due to the duration for which the Trails task is performed, and that the lack of interference with zero-carry problems indicates that the CE is not involved when there is no carrying, and therefore is not involved with the retrieval of arithmetic facts.

Seitz and Schumann-Hengsteler (2000) reported two dual-task experiments, each with 12 participants, which investigated WM involvement in the mental multiplication of visually presented *easy* (both operands 5 or smaller, e.g. 3 x 4) and *hard* (one-digit x two-digit, e.g. 8 x 17) problems, with answers produced verbally. The first within-participants experimental design, used irrelevant speech, articulatory suppression and visuo-spatial tapping as secondary tasks in comparison to a *neutral tapping* control condition, and provided participants with prior training both in strategies for the mental calculation of multiplication problems, and in the performance of the secondary tasks. Analysis of results (MANOVA), for latencies, indicated significant effects of secondary task, difficulty level, and a secondary task x difficulty interaction, with further analysis showing that latencies for easy problems were not affected by any secondary tasks, but that articulatory suppression did increase latencies for difficult problems. Analysis of errors was restricted to difficult problems as easy problems demonstrated too few errors to analyse. Articulatory suppression provided a significantly greater error rate than neutral tapping, but interestingly, error rates for spatial tapping and irrelevant speech were both lower than error rate for neutral tapping, although not significantly so. Seitz and Schumann-Hengsteler suggested that these results demonstrate that access to multiplication arithmetic facts does not involve the PL, although the PL is used for difficult calculations not based on retrieval, and that, at least with adults, the VSSP is not involved in complex multiplication at all. The second experimental design varied from the first only in the replacement of one secondary task, irrelevant speech, with another, random letter generation, which was included to explore the involvement of the CE. Results for solution latencies indicated that both AS and RLG had a significant impact on latencies for difficult problems, but that RLG also impacted easy problems, whereas results for error rates indicted that RLG increased error rates for difficult problems, but AS did not. These results lead Seitz and Schumann-Hengsteler to conclude that access to arithmetic facts is under CE control, and that the CE is also involved in other processes in complex multiplication.

One possible criticism of this study stems from the similarity of results for irrelevant speech and visuo-spatial tapping in comparison to results for neutral tapping, the neutral tapping (control) condition sometimes providing greater latencies and error rates than either irrelevant speech or spatial tapping. It is unclear whether multiplication performed with neutral tapping is equivalent to multiplication performed alone, and it is therefore possible that neutral tapping is itself a secondary task, which either provided an additional cognitive load for WM, or provided a simple cost of concurrence through the serial scheduling of two tasks, in which case, results based on comparisons of other secondary tasks with neutral tapping might be compromised.

Noel et al. (2001) reported an experiment which manipulated either the phonological similarity (Task 1) or the visual similarity (Task 2) of addends in 96 (48 +48) pairs of three-digit + three digit problems, and which measured not only latencies and errors, but four measures of span for each participant, oral digit span, visual digit span, visual digit span with concurrent articulatory suppression, and visuo-spatial span for non-verbal material. Problems were presented visually on a computer screen, the first three-digit number appeared for 1500 milliseconds, and the second, appearing 500 milliseconds after the disappearance of the first, also remained visible for 1500 milliseconds. Problems were also presented horizontally rather than in columnar format to discourage an association with, or preference for, visual coding, and participants were allowed to state their answers from left to right, e.g. 'three hundred and forty five', or as the equivalent spoken digits from right to left, e.g. 5 4 3. Each of the 48 addition problems in each task were further sub-divided into two sets of 24 (high versus low similarity), but the two tasks differed in terms of the carry operation – in the phonological similarity task, each set contained 12 problems with no carries and 12 problems requiring a carry on both units and decades, whereas in the visual similarity task each set contained 12 problems requiring a carry on both units and decades, and 12 problems requiring only one carry – 6 on decades and 6 on units.

The results for the phonological similarity task demonstrated a main effect of similarity modified by a significant task x carry interaction, such that high similarity only increased solution times for two-carry problems, whereas the results for visual similarity provided only a significant main effect of carries – one carry response times faster than two-carry response times. Error analysis of the two tasks gave the same pattern of results as response times. Analysis of span measures demonstrated high correlations between the three measures of digit-span, but no correlation of visuo-spatial span with measures of digit-span, and no correlations between memory span and response times, but negative correlations between all span measures and errors – poorer spans leading to a greater number of errors. Further analysis demonstrating that poorer short-term memory performance, as measured by span, produced more errors in the high phonological similarity condition, but had no effect with visual similarity. Noel et al. conclude that the lack of a visual similarity effect replicates the findings from the Heathcote (1994) study, but in contrast to Heathcote, they suggest that the VSSP does not seem to be involved in mental addition, whereas they consider that the PL plays the major role in storing problem information because phonological similarity clearly impacts response times and errors, although this impact depends upon memory span.

Trbovitch and LeFevre (2003) reported a dual-task experiment investigating WM involvement in the addition of 72, (36) forward, and the same (36) reversed, two-digit + one-digit problems, e.g. (52 +3, 3 + 52). Half of the additions contained one carry and half had no carry. The problems were given continuous visual presentation in vertical (columnar) or horizontal (lateral) format, either as a single-task, or with an *easy* or *hard* concurrent phonological load – one or three consonant-vowel-consonant (CVC) nonwords, or an *easy* or *hard* visual memory load – random pattern of four or eight asterisks, giving an overall (2 x 2 x 2 x 3) design. Analysis of solution latencies indicated a main effect of format; problems in vertical format were solved faster than those in horizontal format. A main effect of operand order was also recorded. Double-digit + single-digit problems were solved faster than single + double-digit problems. Finally, an operand order x format interaction was present, such that double + single-digit problems were solved more quickly only in horizontal format. Further analyses, of latency interactions and error interactions, demonstrated that problems presented horizontally involved greater phonological coding, whereas problems presented vertically involved greater visual coding. Trbovitch and LeFevre suggest that these results are consistent with a role for the VSSP in representing problems in columnar format, and that the PL involvement when problems are presented horizontally might reflect the choice of solution strategies – possibly due to the need to hold partial totals, or due to the use of counting.

One possible criticism of the Trbovitch and LeFevre study is that the phonological load and visual load secondary tasks could be considered to involve more, or at least different, processing than either articulatory suppression or spatial tapping, in that the maintenance of order/ position is required, and if this involves attention, then the tasks might load the CE in addition to the PL and VSSP.

Clearly, studies of more complex mental arithmetic collectively implicate the CE, the VSSP and PL components of WM, but individually they can be directly contradictory. The CE is both implicated and not implicated in the retrieval of arithmetic facts, and the VSSP is both implicated and not implicated in the representation of problems, whereas the PL is implicated in several different ways – holding partial totals, holding problem information, possibly accessing solution strategies, and possibly, but probably not accessing arithmetic facts. Is the situation for simple arithmetic any clearer?

Lemaire, Abdi and Fayol (1996) reported two dual-task experiments investigating WM involvement in the verification of visual, continuously presented, true and false single-digit additions and multiplications. The problems were classified as *easy* or *difficult*, and performed alone or concurrently with one of three secondary tasks, articulatory suppression, canonical suppression and random letter generation. False problems included *confusion* problems, e.g., 3 x 5 = 8 and 3 + 5 = 15 and *non-confusion* problems, e.g. 3 + 5 = 7. Solution latencies for true problems demonstrated; a significant difference between RLG and other secondary tasks, increased latencies for addition problems in comparison to multiplication problems for the AS and control conditions, and increased latencies for multiplications compared to additions in the RLG and canonical articulation conditions. Easy problems were verified faster than difficult problems for all tasks, with differences most noticeable under RLG. Analysis of latencies for false problems indicated; that RLG was significantly different from other secondary tasks, and that easy problems were faster than difficult problems. Difficulty effects were smaller for RLG and canonical articulation, and confusion problems took longer in all conditions except RLG. Analysis of errors showed that errors increased only under RLG, and that easy problems produced fewer errors than difficult problems. Lemaire et al. suggested that the results demonstrate that in verifying true problems, the PL and CE are involved, although the CE is the critical WM component in simple mental arithmetic as a whole. In order to check that differences in the first experiment were not due to participants in the RLG condition *double-checking* the correct operation encoding sign, they replicated this experiment with additions and multiplications as between-participants factors, and found similar results.

DeStefano and LeFevre (2003) identified a possible problem in the Lemaire et al. study, which is that the articulation tasks were performed to a time schedule, and this might have involved CE resources, while Noel et al. (2001) again question the lack of interaction in this and other studies, pointing out that the CE task simply added a constant amount of processing time, but did not appear to affect performance on the arithmetic task.

De Rammelaere et al. (2001) reported two experiments that investigated the role of the CE and PL in the verification of pairs of single-digit additions and multiplications. As in the Lemaire et al. (1996) study, they had the specific aim of clarifying the role of the PL in the verification of true problems, and to this end, they carried out a number of modifications to the materials used in the Lemaire et al. study. The *split* size for false confusion addition problems was restricted to a either a simple + or – 1, or to + or – 9, because those in the Lemaire et al. study were considered *extreme* and too easy to falsify through a simple *implausibility* strategy rather than a *retrieval* strategy, e.g. product + or – 1, such as 7 + 8 = 55, whereas De Rammelaere et al. used problems such as 7 + 8 = 14, or 7 + 8 = 24. The RLG task was replaced with random interval generation (RIG), because this was considered to be a relatively pure CE task that did not load the PL. The first experiment considered addition alone, and included a size factor, problems with sums less than 10 classed as *small*, giving a (3 load – control, AS, RIG x 2 size –small, large x 3 sum – true, small split, large split) design, with results supporting CE involvement, but demonstrating no role for the PL in true or false problems.

The second experiment used the same 30 participants, and was essentially similar to the first experiment, but considered multiplication problems alone, with confusion errors modified by introducing splits based on multiplying the true product by four values, 0.1,0.9, 1.1 or 1.9, and then rounding to an appropriate opposite odd/even integer which was not a multiple of either operand or the sum of the two operands, e.g. 9 x 3 = 27 (odd), 27 x 1.9 = 51.3 (52, even). Results demonstrated a role for the CE in supporting simple multiplication but no evidence was found for the role of the PL in simple multiplication.

De Rammelaere et al. explain why this is an important issue, for the results of Lemaire et al. (1996) were directly supportive of the triple-code model (Dehaene 1992, Dehaene and Cohen 1995) in that the triple-code model requires that true rote-learned arithmetic facts are accessed phonologically/ verbally, whereas false arithmetic facts would not be learned at all. De Rammelaere et al. consider that the CE has a general rather than a specific processing role in supporting arithmetic, due to the lack of load x size interaction in their study, but that their problem size x split interactions provide further evidence that large-split problems are falsified on the basis of implausibility, whereas small-split problems require the use of retrieval.

Hecht's (2002) dual-task study of the verification of simple additions has already been mentioned with regard to strategy use and WM, but is also important here, because the study essentially analysed true and false verification times for additions under AS and RLG in terms of five different strategy groups. Raw latencies demonstrated an effect of RLG but not AS, but a regression analysis of strategy solution times indicated that AS increased response times for those using the *counting* strategy, both for true and false trials, and that RLG increased response times for those using the counting strategy, but only for true trials. This lead Hecht to the conclusion that the CE and PL are only involved in simple addition when participants are counting, and are not used at all in the retrieval of arithmetic facts.

Lee and Kang (2002) reported a dual-task study, involving 10 participants, which investigated the effect of phonological suppression (repetition following the alternating presentation – every 5 seconds, of a non-word string) and visuo-spatial suppression (holding shape and location details following the alternating presentation – every 5 seconds, of an image) on pairs of single-digit simple subtractions and multiplications, with continuous visual presentation. Results for response times demonstrated that phonological suppression led to significantly longer response times for multiplication but not subtraction problems, whereas visuo-spatial suppression led to significantly longer response times for subtraction but not multiplication problems. Lee and Kang took this as evidence against McCloskey's abstract-modular model, because multiplication and subtraction were both performed in the same visual modality, and they consider that this means that disruption is unlikely to be due to encoding or output stage problems. They argue that these findings support the Dehaene and Cohen triple-code model, because multiplication facts in the triple-code model are accessed phonologically, whereas subtraction is considered to involve a non-phonological analog magnitude representation route. Clearly, the situation in simple single-digit arithmetic studies is no less contradictory than in studies of complex multi-digit arithmetic.

Chapter Four - Rationale for experimental study

The first three chapters have demonstrated that studies indicating WM involvement in cognitive arithmetic are not only largely contradictory, but also potentially biased towards particular cognitive models. This is, at least in part, due to the following factors; the large variety of secondary tasks and arithmetic tasks used in these studies, the way that these tasks are presented, performed and measured, and individual differences in participant age, ability and strategy use. While it is not practical to experimentally investigate every aspect of these factors, it is possible to explore the efficacy of a variety of secondary tasks in terms of their ability to disrupt cognitive arithmetic, thereby indicating WM involvement, and to be more specific not only about the type of arithmetic task used and the way in which the arithmetic task is presented, but also to control or investigate age, ability and strategy use, and to this end, a first series of experiments were designed.

4.1 Rationale and general experimental design for first series of experiments

The first series of five dual-task experiments, carried out in 2001, was exploratory, in that the primary purpose was to integrate material from various sources to partially replicate and extend previous studies, in particular those of Logie et al. (1994), Heathcote (1994) and Furst and Hitch (2000). The experiments were designed to establish the efficacy of five different concurrent secondary tasks in the disruption of a particular set of multi-digit additions, with additional analysis of dual-task disruption and overall strategy use, and with participants of limited age-range, low in math anxiety and demonstrating acceptable arithmetic ability. Articulatory Suppression (AS) or more precisely, canonical articulation, in the form of repetitions of (A, B, C) was used to load the phonological loop component of WM, while Spatial Tapping (ST), repetitions of a figure-of-eight pattern on a numerical keyboard, was used to load the visuo-spatial sketch pad. Three putative *executive tasks* were used to load the central executive - Random Letter Generation (RLG), the *random* articulation of letters from a subset of the alphabet (A, B, C, D, E, F, G, H, I), Random Key Pressing (RKP), pressing one of nine keys at *random*, and Random Interval Generation (RIG), pressing one key at *random* time intervals. Multi-digit (three-digit + two-digit) additions, were presented visually and continuously on a computer screen, with the additions themselves being manipulated both in terms of the number of carry operations required, and in terms of the format of presentation (linear/horizontal versus columnar/vertical). The different patterns of interaction between format, task and number of carries offered not only a potentially more detailed analysis of WM involvement, but also the clarification of the relationships between different secondary tasks designed to affect the central executive component of WM.

To establish whether a decrement to both primary and secondary tasks did occur, the measurement of arithmetic task performance under dual task conditions was compared with performance on equivalent tasks under single task conditions, and similar comparisons were made for secondary tasks performed alone or under dual-task conditions. All five experiments had the same repeated-measures design: Task (2 levels; additions alone vs. additions under dual-task conditions) x Carries (3 levels; zero, one and two carries) x Format (2 levels; linear presentation vs. columnar presentation). Response times and error rates constituted the two dependent variables, and these were analysed separately for each experiment, with additional analyses for different error types and participant strategies. A total of forty seven sixth-form and undergraduate students participated in these initial experiments, 27 males and 20 females, aged between seventeen and twenty-four years (modal age 17 years). All were volunteers, and all received a small fee for their participation. To avoid recruiting participants suffering from maths anxiety (Ashcraft and

Kirk, 2001), the recruitment procedure informed potential participants that basic mathematical competence was required, and that anyone who experienced anxiety when doing arithmetic should not take part in the study. Table 1 shows the number of participants in each experiment.

Experiment	Total (N)	Males	Females	Age	(SD)
Articulatory Suppression	10	5	5	18.10	(2.28)
Spatial Tapping	9	6	3	18.33	(2.24)
Random Letter Generation	8	4	4	17.50	(1.07)
Random Key-pressing	10	7	3	17.60	(1.58)
Random Interval Generation	10	5	5	18.30	(1.95)

Table 1. Number and mean age (in years) of participants in each experiment

4.2 The arithmetic task

The arithmetic task was common to all five experiments, and consisted of a total of 360 multi-digit additions. Each addition had a similar combination of three digit plus two digit addends with totals ranging from 157 to 235 (see appendix A), and were presented as twelve separate but equivalent sets, each of thirty additions, arranged in six paired categories for single versus dual-task use. The categories reflected the combination of the carry and format variables as follows; zero carries linear format (zcl), zero carries columnar format (zcc), one carry linear format (ocl), one carry columnar format (occ), two carries linear format (tcl), and finally, two carries columnar format (tcc). Figure 9 provides an example of the six paired categories of additions.

1zcl	2zcl	1zcc	2zcc	1ocl	2ocl	1occ	2occ	1tcl	2tcl	1tcc	2tcc
123+34 (157)	125+32 (157)	126+ 42 (168)	125+ 43 (168)	129+62 (191)	139+52 (191)	134+ 57 (191)	135+ 56 (191)	123+78 (201)	136+65 (201)	127+ 74 (201)	132+ 69 (201)

Figure 9. Examples of the six categories of additions

The choice of additions reflects the need to control for the problems identified earlier, in Ashcraft's (1992, 1995) reviews of cognitive arithmetic, and consequently, the additions were subject to the following constraints: 1) Hundreds column of three digit addends restricted to 1 in order to limit the range of answers to between 157 and 235, and to thereby limit problem size effect, although in fact, problem size effect is largely controlled by using pairs of additions with the same totals in single and dual-task conditions. 2) Different numbers in tens and units of both addends to avoid *ties effect*, and associated decrease in response times. 3) No ones or zeroes in tens and units addends, for this also decreases response times. 4) Sum totals must be capable of being generated from at least two different pairs of addends. Each addition problem was produced as a 26 point black and white bitmap image using a PC running the Microsoft ® *Paint* program, see Figure 10.

$$125 + 68 = \_\_\_\_ \qquad\qquad \begin{array}{r} 129 \\ +\ \ 64 \\ \hline = \_\_\_\_ \end{array}$$

Figure 10. Examples of additions produced as bitmaps

These images were transferred to the *Superlab Pro for Windows* program version 1.05 – Cedrus Corporation. This program allowed the recording of response times (to nearest millisecond) from the onset of presentation to key press (spacebar), and also provided a timing delay of three seconds between key press and presentation of next trial, this interval being chosen in order to provide sufficient time not only for the answer to be given, but also sufficient time to recommence secondary task performance. Each set of trials was randomised using the program's random number generator with a *constant seed* setting, and each block of trials was prefaced with an instruction screen and supplemented with verbal instructions where necessary. The twelve blocks of trials were counter-balanced for order of presentation in accordance with a task order table (see appendix B). Problems were presented singly, in the centre of the computer screen, and were continually visible until terminated by key press, and separate response sheets were used to record participant answers, errors and omissions. The AS and RLG experiments required only a single computer, as articulations were recorded using a portable tape recorder, but the ST, RKP and RIG experiments all required the use of a second computer running separate Superlab programs to record secondary task responses. Protective boxes were used to prevent any other keys than the 1-9 keys on the numeric keypads from being pressed, and these also covered the keypads so that participants had to rely on spatial location and touch. All experiments were conducted in rooms provided by the host institutions, with only the experimenter and participant present. Participants were encouraged to arrange the seating, keyboard(s) and screen distance as desired, provided that this did not affect the requirements of the experiment. Prior to the experiment each participant completed a *verbal protocol* task, to identify their particular preference for solution strategies, as they solved three linear and three columnar arithmetic additions. The allocation of participants to experiments followed a simple quota system depending on the experiment that had been prepared at particular times and institutions. Each participant was given an instruction sheet for the secondary task involved, supplemented by further explanation if required. Participants were encouraged to mentally calculate the answer to the problem as quickly as possible, and then to press and release the spacebar to record their response times. They were asked to clearly state their answers immediately after pressing the spacebar.

4.3 Experiment one - articulatory suppression

If phonological representation is involved in the carrying operation, then there should be a significant task x carry interaction, whereas if phonological representation is involved differentially in the extraction of information from linear or columnar formats, then a significant task x format interaction should be found. If the carrying operation is supported by presentation format, then a significant carry x format interaction can be expected, but if this is dependent upon phonological representation, then there should be a three-way interaction such that the pattern of carry x format interaction is different over the two levels of the task variable.

The articulatory suppression task required participants to constantly repeat the three letters (A, B, C) at a steady self-determined rate, as variation in articulation rate was the measure of secondary task decrement, and in order to maintain the flow of articulations, participants were instructed to insert the answers to the additions into the AS task. A baseline measurement of AS as a single task was obtained for a period of two minutes, for later comparison with three 60-second samples taken from the beginning, middle and end of each of the six dual-task blocks, and the average articulation rate per second was calculated from each of these measurements.

The analysis of errors is common to all of the experiments, and an error index was obtained by dividing the number of errors in each category by the number of completed items in each category – items which were not attempted for any reason, or those which had response times greater than two standard deviations from the mean, were excluded from the analysis. The errors in each category were divided into five different types: incorrect units column, incorrect tens column, incorrect hundreds column, reversals (tens and units swapped) and multiple column errors, to give single and dual-task percentage rates for these five error types.

4.3.1 Articulatory suppression - response times

Initial exploratory analysis for response times in all experiments indicated that the data transformation (y = 10 log *x)* led to improvements in the normality of the distributions of scores, and to greater homogeneity of variance, thereby providing greater compliance with the underlying assumptions of ANOVA, (see Howell, 1997). However, it is possible that this logarithmic transformation leads to a multiplicative rather than additive instance of the General Linear Model (GLM) underlying ANOVA. This does not prevent the use of the GLM, (see Tabachnick and Fidell, 2001), but it does mean that interactions between the experimental factors in the transformed scores might indicate different effects to those indicated in the untransformed scores. Figure 11 provides an example of how a purely additive effect in the untransformed scores can appear as a sub-additive interaction in log transformed scores.

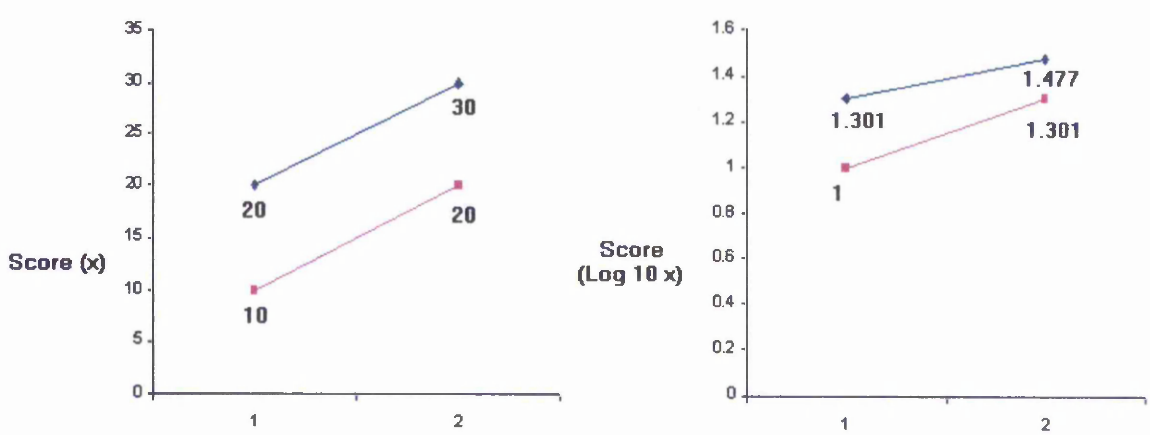

Figure 11. Apparent interaction when scores are plotted as log 10 values.

Subsequent analyses are all based on transformed scores, but where necessary, to aid interpretation, graphs of interactions will be plotted using both transformed and untransformed scores, although tables will continue to report untransformed mean response times in milliseconds.

Table 2 presents untransformed mean response times for task (dual, single) by carries (one, two or three) and format (linear, columnar).

Dual -Task	Mean (SD)	Single-Task	Mean (SD)
ZCL	5200 (2006)	ZCL	4230 (1327)
ZCC	4861 (2173)	ZCC	3846 (1357)
OCL	8068 (3288)	OCL	6058 (1758)
OCC	6941 (2542)	OCC	5898 (1900)
TCL	9491 (3522)	TCL	8450 (3341)
TCC	9032 (3199)	TCC	7941 (2596)

Key: (zcl) zero carries linear format (occ) one carry columnar format
(zcc) zero carries columnar format (tcl) two carry linear format
(ocl) one carry linear format (tcc) two carry columnar format

Table 2. Untransformed mean response times and standard deviations in milliseconds (Articulatory suppression)

NB. Greenhouse-Geisser probability corrections are routinely reported in this and subsequent analyses, and (*M*) refers to mean response times re-calculated from log-transformed scores. A (2 x 3 x 2) repeated-measures ANOVA of the transformed AS response time data demonstrated a significant difference between dual-task (M = 6516 msec) and single-task (M = 5483 msec), $F_{(1, 9)} = 14.91, p < 0.005$.

A significant difference between zero-carry ($M = 4198$ msec), one-carry ($M = 6281$ msec) and two-carries ($M = 8110$ msec), $F_{(2, 18)} = 54.13$, $p < 0.001$, and a significant difference between linear format ($M = 6209$ msec) and columnar format ($M = 5754$ msec), $F_{(1, 9)} = 8.70$, $p < 0.001$. A significant task x carry interaction, $F_{(2, 18)} = 4.25$, $p < 0.001$, and a significant task x carry x format interaction, $F_{(2, 18)} = 4.11$, $p < 0.05$. No other interactions were significant. Tests of within-subjects contrasts revealed that the task x carry interaction was due to a greater effect of articulatory suppression on one-carry problems compared to two-carry problems, $F_{(1, 9)} = 5.60$, $p < 0.05$, and this was modified by the task x carry x format interaction which demonstrated that articulatory suppression had a significantly greater impact on one-carry problems in linear format, $F_{(1, 9)} = 6.53$, $p < 0.05$.

4.3.2 Articulatory suppression – error rates

Error rates, and resultant error index – based on number of errors divided by number of items completed, varied considerably between participants, and between the categories of problem, with relatively few errors in the no-carry problems compared to one-carry and two-carry problems. The untransformed average error index per category is given in table 3, but initial exploratory analysis indicated that the data transformation [$y = \text{sqrt}(x + 0.5)$] gave improvements to the shape of the distributions, and provided greater homogeneity of variance, thereby more closely meeting the ANOVA assumptions, and this transformation was used here and in all subsequent error analyses and graphs, although tables will continue to give untransformed error index values.

Dual -Task	Mean (SD)	Single-Task	Mean (SD)
ZCL	0.081 (0.091)	ZCL	0.065 (0.062)
ZCC	0.048 (0.051)	ZCC	0.030 (0.043)
OCL	0.093 (0.133)	OCL	0.081 (0.050)
OCC	0.072 (0.111)	OCC	0.101 (0.069)
TCL	0.148 (0.073)	TCL	0.142 (0.102)
TCC	0.189 (0.122)	TCC	0.160 (0.111)

Key: (zcl) zero carries linear format
(zcc) zero carries columnar format
(ocl) one carry linear format
(occ) one carry columnar format
(tcl) two carry linear format
(tcc) two carry columnar format

Table 3. Untransformed mean error index and standard deviations (Articulatory suppression)

NB. In the analysis of errors (*M*) refers to the mean transformed error index.

A (2 x 3 x 2) repeated-measures ANOVA demonstrated a significant difference between zero-carries (M = 0.745), one-carry (M = 0.764) and two-carries (M = 0.810), $F_{(2, 18)} = 15.11$, $p < 0.001$, but no other significant main effects or interactions. Tests of within-subjects contrasts revealed that two-carry problems had significantly more errors than one-carry problems, $F_{(1, 9)} = 14.16$, $p < 0.005$, but that one-carry problems did not have greater error rates than zero-carry problems.

Analysis of errors showed that only one of the error types (multiple errors) demonstrated more errors in dual-task conditions, $t = 2.886$, $p < 0.05$, and this has to be interpreted with caution given the multiple use of t-tests and the somewhat general nature of this error type. Articulation rate – baseline (AS alone) vs. dual-task (AS with maths) – was also analysed using a t-test, and was found not to be significantly different, which means that the experiment failed to meet the previously stated criterion regarding evidence of resource competition in dual-task studies. There was no evidence of a speed/accuracy trade-off, as no significant negative, or indeed positive, correlations were obtained, between each group of category response times and respective category errors, in single or dual-task conditions.

4.3.3 Articulatory suppression - interpretation and discussion of results

For response times, the main effects of task, carries and format were all significant, as was the task x carry interaction, and the task x carry x format interaction. Overall, response times increase under AS, increase with number of carries, and increase when problems are presented in linear rather than columnar format. As figures 12 and 13 demonstrate, the interactions show AS as having a differentially greater effect only on one-carry problems, and then, only when these were presented in linear format.

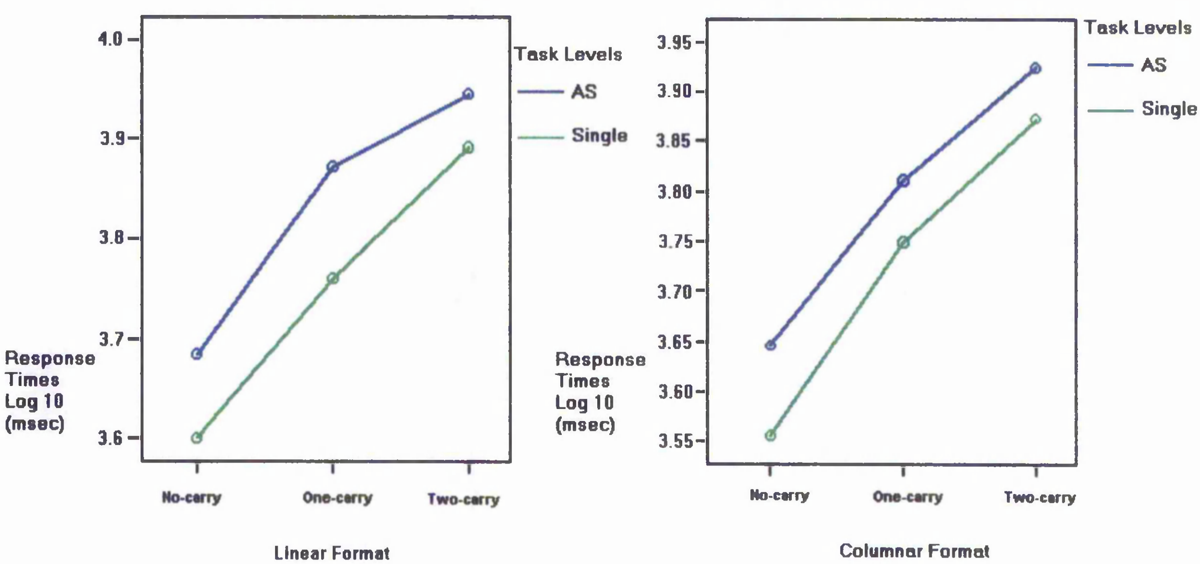

Figure 12. Graph of task x carry x format interaction (AS transformed response times)

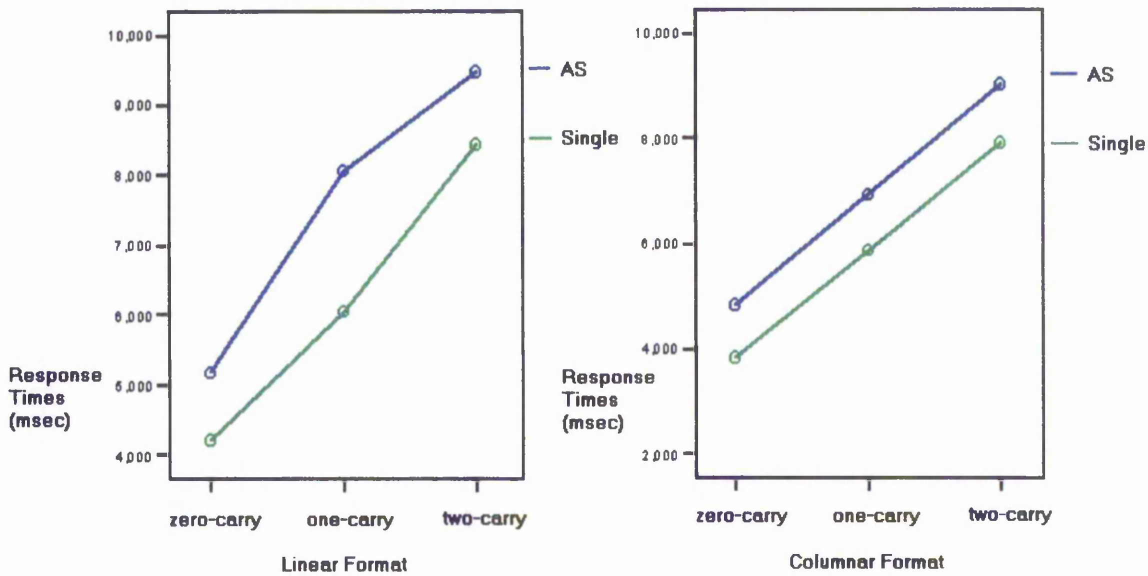

Figure 13. Graph of task x carry x format interaction (AS untransformed response times)

The interaction is present in both the transformed and untransformed data, but this can only be consistent with a role for phonological coding, mediated both by the number of carries and the format of presentation, if the two-carry problems are ignored. A linear relationship might be expected, with two-carry problems showing greater rather than lesser impact. If the lack of effect of AS at the two-carry level is ignored, we are still left with the question of why phonological coding would be needed for visually presented linear format single-carry additions?

A possible, but speculative explanation might be found by re-considering the *triple code* model of Dehaene (1992) and the conclusions of Heathcote's (1994) study. Dehaene claims that multi-digit operations involve the manipulation of spatial images, presumably in columnar format, and Heathcote certainly considers the possibility that linear material might have to be re-represented in columnar format. Could it be that verbal recoding of linear additions is used to achieve columnar format in the VSSP? Trbovitch and LeFevre (2003) reported that problems presented horizontally seem to access phonological codes, whereas vertical columnar presentation appeared to engage visual codes, does this mean that linear presentation engages reading rather than calculation processes? If arithmetic facts are learned by rote, then could representations also be learned by rote, giving a preference for columnar format? The additional time that might be required to recode linear additions is certainly consistent with the above interactions, but unfortunately, it is also consistent with other explanations, such as the extra encoding time needed to move the eyes back and forth across linearly presented problems?

No definitive conclusions can be drawn from the analysis of response times here, but the results do provide sufficient justification for further experimentation in the second series of experiments. The analysis of error rates provided no evidence for phonological coding - no significant interactions involving task, and no significant main effect of task. If errors do more directly reflect calculation processes rather than encoding processes, then this could indicate that phonological coding does not play a part in the calculation process, but to argue that the lack of interaction is important, is to encounter the problem of using negative results to demonstrate a negative – I haven't found it, therefore it doesn't exist. The lack of interaction could be due to other reasons – such as the small numbers in this study, and this is another reason to partially replicate this experiment in the second series of experiments. The failure to demonstrate a secondary task decrement could indicate a *strategic trade-off* as detailed by Hegarty et al. (2000), but as the arithmetic task is considerably harder than the AS task, it is difficult to see why participants would wish to divert resources to the AS task. The "A, B, C" task used here has just the three responses and these seem almost automatic given the 'over-learned' status of this triplet. It is perhaps because of this near automaticity that the rate of articulation remained the same in single and dual-task conditions, the AS task simply failing to load phonological WM, but it could also indicate that phonological coding is only minimally involved, or not involved at all, in this arithmetic task.

4.4 Experiment two – spatial tapping

The spatial tapping task required participants to tap a 'figure of eight' pattern on the numeric keypad keys 1-9. The rate of tapping was again self-determined, but variation in tapping rate was not considered critical here, as the measure of secondary task decrement was the number of sequences tapped correctly from selected blocks of twenty sequences. Participants were asked to use their non-dominant hand to tap the spatial pattern, and their dominant hand to press the spacebar of the computer running the arithmetic program. The correct 'figure of eight' pattern key sequence was as follows: (1, 2, 3, 6, 5, 4, 7, 8, 9, 6, 5, 4...). A baseline measurement of spatial tapping was obtained, with participants carrying out spatial tapping as a single task for a period of 3 minutes. A series of twenty sequences involved 240 individual taps, and each correct sequence of twelve taps achieved a score of 0.05, giving a range from 0-1 overall. Each of the six blocks of dual-task trials was sampled and scored separately and the scores were averaged over all blocks. In other respects, the spatial tapping experiment was the same as the first experiment.

4.4.1 Spatial tapping – response times

A (2 x 3 x 2) repeated-measures ANOVA of the ST transformed response time data demonstrated a significant difference between dual-task (M = 6427 msec) and single-task (M = 5284 msec), $F_{(1, 8)} = 28.95, p < 0.001$. A significant difference between zero-carries (M = 4169 msec), one-carry (M = 6281 msec) and two-carries, $F_{(2, 16)} = 80.93, p < 0.001$. A significant difference between linear format (M = 6039 msec) and columnar format (M = 5623 msec), $F_{(1, 8)} = 9.82, p < 0.05$. None of the interactions were significant. Tests of within-subjects contrasts demonstrated that response times for one-carry problems were greater than zero-carry problems, and response times for two-carry problems were greater than one-carry problems. Table 4 presents mean response times for task (dual, single) by carries (one, two or three) and format (linear, columnar).

Dual -Task	Mean (SD)	Single-Task	Mean (SD)
ZCL	5510 (2381)	ZCL	4426 (1972)
ZCC	4796 (2375)	ZCC	3914 (1712)
OCL	7675 (2942)	OCL	6185 (2380)
OCC	7511 (2972)	OCC	6311 (2721)
TCL	9419 (3827)	TCL	7925 (3453)
TCC	8866 (3797)	TCC	7455 (3178)

Key: (zcl) zero carries linear format
(zcc) zero carries columnar format
(ocl) one carry linear format
(occ) one carry columnar format
(tcl) two carry linear format
(tcc) two carry columnar format

Table 4. Mean response times and standard deviations in milliseconds (Spatial tapping).

4.4.2 Spatial tapping –error rates

A (2 x 3 x 2) repeated-measures ANOVA of the ST transformed error data demonstrated a significant difference between zero-carry (M = 0.727), one-carry (M = 0.759) and two-carries (M = 0.778), $F_{(2, 16)}$ =12.35, $p < 0.001$, and a significant task x carry interaction, $F_{(2, 16)} = 4.87$, $p < 0.05$. No other effects or interactions were significant. Tests of within-subjects contrasts indicated that two-carry errors were not significantly different from one-carry errors, and one-carry errors were not significantly different from no-carry errors, but no-carry errors were significantly less than two-carry errors, $F_{(1, 8)} = 13.81$, $p < 0.01$. This was modified by the task x carry interaction which is, at least in part, a crossover interaction, such that concurrent ST produced more errors on no-carry problems, but fewer errors on two-carry problems, $F_{(1, 8)} = 7.00$, $p < 0.05$ (see figure 14). The untransformed mean error index scores are given in table 5.

Dual -Task	Mean (SD)	Single-Task	Mean (SD)
ZCL	0.041 (0.055)	ZCL	0.019 (0.025)
ZCC	0.038 (0.042)	ZCC	0.019 (0.025)
OCL	0.090 (0.063)	OCL	0.077 (0.061)
OCC	0.069 (0.046)	OCC	0.072 (0.053)
TCL	0.073 (0.053)	TCL	0.122 (0.095)
TCC	0.123 (0.085)	TCC	0.113 (0.077)

Key: (zcl) zero carries linear format
(zcc) zero carries columnar format
(ocl) one carry linear format
(occ) one carry columnar format
(tcl) two carry linear format
(tcc) two carry columnar format

Table 5. Mean error index and standard deviations (Spatial tapping)

There were no significant differences in the five error types between single and dual-task conditions, and no significant differences were found in the number of correctly tapped sequences in baseline versus dual-task conditions, so that as in the first experiment, the criterion for the acceptance of dual-task interference was not met. An additional separate analysis of tapping-rate showed no difference between single and dual-task conditions, so accuracy was not maintained at the cost of speed. There was also no evidence of a speed/accuracy trade-off, in terms of significant negative correlations, between response times and errors, but there were significant positive correlations between response times and errors for two-carry columnar problems, both in single-task, $r = 0.687$, $p < 0.05$, and dual-task conditions, $r = 0.840$, $p < 0.01$. The unsurprising interpretation of these findings is that as processing time and complexity increases so do errors.

4.4.3 Spatial tapping – interpretation and discussion of results

Given the lack of interactions in the response time data, the significant main effects provided no evidence that carrying has a spatial component, or that either of the presentational formats involved spatial processing to a different extent. The main effect of format irrespective of task again suggested the speculative possibility that linearly presented additions require additional processing time as they are recoded into conventional format, but as before, this could be due to linear format engaging phonological or reading processes, or to increased eye movement. The main effect of task might simply indicate *cost of concurrence* rather than any effect that ST might have on WM, while the main effect of carries simply indicates that response time increases as the number of operations increases.

The significance of the error data task x carry interaction, shown in figure 14, initially appeared to support the role of spatial processing in the carry operation, but further

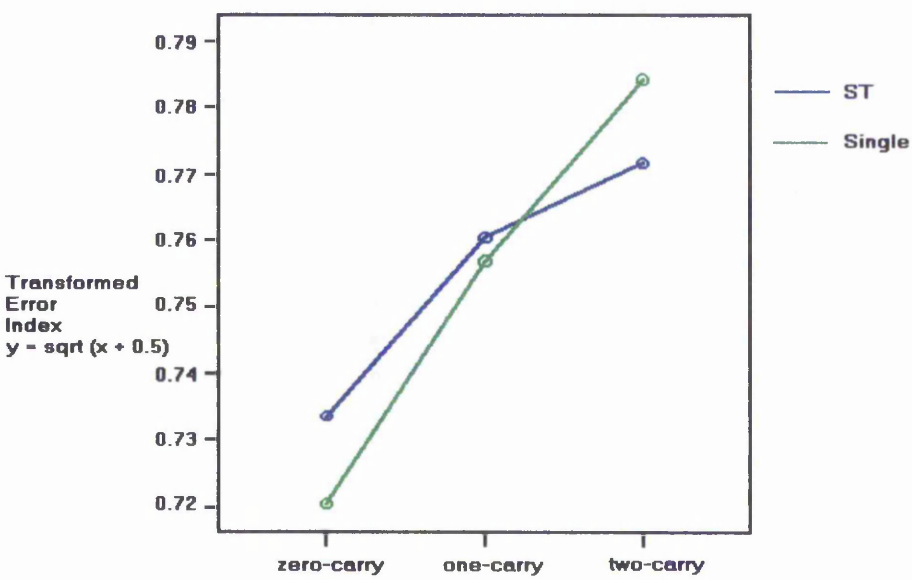

Figure 14. The spatial tapping task x carry interaction (transformed error index)

analysis of this interaction showed no support for the role of a spatial rehearsal mechanism in the processing of these multi-digit additions, for the interaction stemmed from ST producing fewer errors on two-carry problems than when they were performed alone. The absence of task x format interactions suggests that if spatial interference affects linear and conventional formats, it does so equally, and this could mean that the superiority of the conventional columnar format is not due to spatial processing, although the lack of significant difference between the number of correctly tapped sequences in baseline and dual-task conditions provided no evidence that the spatial tapping task did in fact interfere with visuo-spatial processes. However, the comments made by two participants during debriefing suggest another possibility, for they claimed to have found the arithmetic task easier to perform under dual task conditions because the spatial tapping task provided a consistent rhythm that aided their calculation, a fact which might begin to explain the counter-intuitive task x carry interaction described above. Whether this is in any way related to the prior experience of participants is unknown, but rhythmic chanting has been used to rote learn additions and multiplications in some schools. It may simply be that tapping a known rhythm is relatively automatic, allowing greater allocation of resources to calculation.

4.5 Experiment three – random letter generation

If the central executive is involved in the carrying operation in arithmetic, then once again, this should lead to significant task x carry interactions, and insofar as RLG also constitutes an AS task in addition to a CE task, it will be informative to compare the pattern of results obtained in this experiment with that found in experiment one. Similarly, if the central executive is involved in the processing of the format of presentation, then this should also be evident in the pattern of interactions.

This experiment followed the procedure outlined for the earlier experiments except for the nature of the secondary task. Participants were asked to articulate letters from a sub-set of the alphabet (A, B, C, D, E, F, G, H, I) as randomly as possible, and were allowed to respond at a self-determined rate. A baseline measurement of RLG, as a single task, was obtained from tape recordings made before the commencement of trials, while tape recordings made during the trials were later sampled to provide sequences of 81 articulations from each of the six dual-task trials. The sequences were taken from the approximate mid-point of the duration of each trial, and were used in the construction of a 9 x 9 numerical matrix - a modification of the procedure described in Evans (1978), which provides an index of randomness between 0 and 1, higher scores indicating increasing departure from randomness.

4.5.1 Random letter generation – response times

The untransformed mean response times for each task x carry x format level are given in table 6. A (2 x 3 x 2) repeated measures ANOVA of the RLG transformed response time data showed a significant difference between single-task (M = 4325 msec) and dual-task (M = 6281 msec), $F_{(1, 7)} = 49.75$, $p < 0.001$. A significant difference between zero-carry (M = 3631 msec), one-carry (M = 5598 msec) and two-carries (M = 6966 msec), $F_{(2, 14)} = 34.24$, $p < 0.001$. A significant difference between linear format (M = 5445 msec) and columnar format (M = 4989 msec), $F_{(1, 7)} = 27.83$, $p < 0.001$. None of the interactions achieved significance. Tests of within-subjects contrasts showed that no-carry problems were solved significantly faster than one-carry problems, $F_{(1, 7)} = 33.55$, $p < 0.001$, and that one-carry problems were solved significantly faster than two-carry problems, $F_{(1, 7)} = 30.64$, $p < 0.001$.

Dual -Task	Mean (SD)	Single-Task	Mean (SD)
ZCL	4948 (1961)	ZCL	3394 (1000)
ZCC	4341 (1398)	ZCC	2948 (859)
OCL	8031 (4226)	OCL	5182 (2087)
OCC	7599 (3720)	OCC	4592 (1762)
TCL	10135 (6201)	TCL	6847 (3479)
TCC	9611 (6091)	TCC	6361 (3016)

Key: (zcl) zero carries linear format
(zcc) zero carries columnar format
(ocl) one carry linear format
(occ) one carry columnar format
(tcl) two carry linear format
(tcc) two carry columnar format

Table 6. Mean response times and standard deviations in milliseconds (Random letter generation)

4.5.2 Random letter generation – error rates

The untransformed mean error index for each task x carry x format level is given in table 7. A (2 x 3 x 2) repeated-measures ANOVA of the RLG transformed error data showed a significant difference between single-task (M = 0.799) and dual-task (M = 0.827), $F_{(1, 7)} = 21.19$, $p < 0.01$. A significant difference between zero-carry (M = 0.733), one-carry (M = 0.812) and two-carries (M = 0.894), $F_{(2, 14)} = 17.05$, $p < 0.001$, and a significant task x carries interaction, $F_{(2, 14)} = 4.167$, $p < 0.05$. No other effects or interactions were significant.

Tests of within-subjects contrasts indicated that no-carry problems produced significantly fewer errors than one-carry problems, $F_{(1, 7)} = 9.00$, $p < 0.05$, and significantly fewer errors than two-carry problems, $F_{(1, 7)} = 7.67$, $p < 0.05$, whereas one and two-carry problems were not significantly different. This appears to be the pattern of interaction expected if RLG, as an executive task, has an effect only on the carry operation, see figure 15. However, it is possible that the lack of significant difference at the no-carry level represents a 'ceiling effect' - error scores cannot be less than zero, so that where the number of errors is small, as is the case for the no-carry conditions shown here, the scores do not have the range to demonstrate significant differences.

Dual -Task	Mean (SD)	Single-Task	Mean (SD)
ZCL	0.050 (0.040)	ZCL	0.048 (0.057)
ZCC	0.022 (0.036)	ZCC	0.034 (0.036)
OCL	0.197 (0.128)	OCL	0.100 (0.122)
OCC	0.201 (0.096)	OCC	0.161 (0.122)
TCL	0.333 (0.218)	TCL	0.263 (0.198)
TCC	0.375 (0.245)	TCC	0.281 (0.207)

Key: (zcl) zero carries linear format
(zcc) zero carries columnar format
(ocl) one carry linear format
(occ) one carry columnar format
(tcl) two carry linear format
(tcc) two carry columnar format

Table 7.Mean error index and standard deviations (Random letter generation)

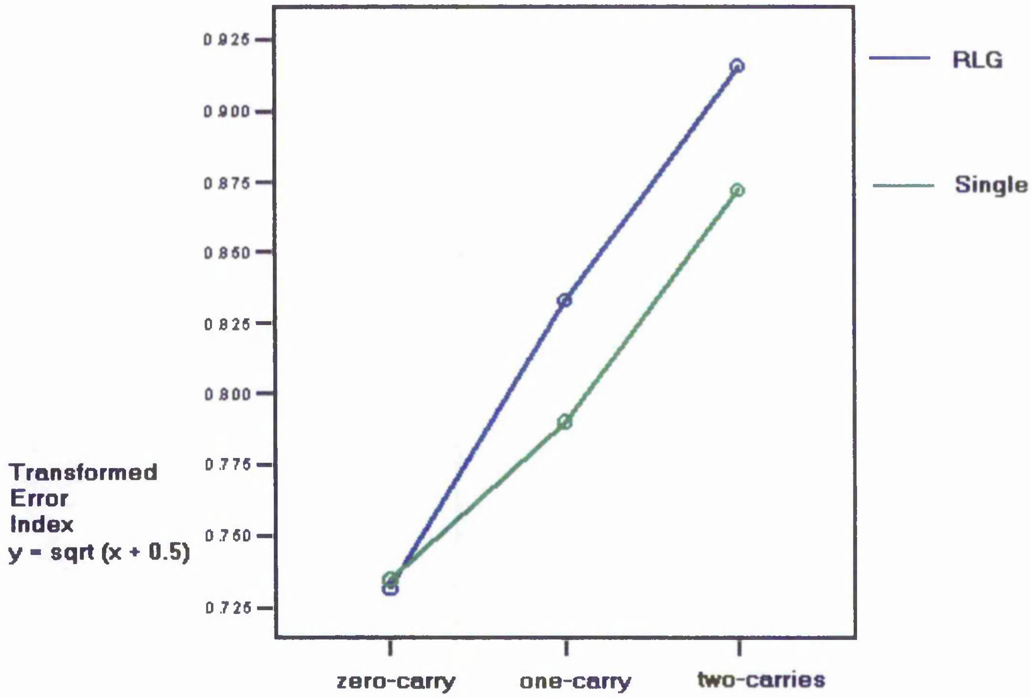

Figure 15. Graph of task x carry interaction (RLG transformed error data)

There were no significant differences in the five error types between single and dual task conditions, and no significant negative, or positive, correlations between each category of response times and associated errors, providing no evidence of a speed/accuracy trade-off. The comparison of baseline and dual task RNG did show a significant decrease in randomness under dual task conditions (t =3.32, p <0.01), and this does fulfil the criterion for evidence of resource competition between primary and secondary tasks.

4.5.3 Random letter generation – interpretation and discussion of results

Without significant interactions there was no evidence from the response time data that the carry operation was disrupted by RLG, or that RLG disrupted format, and insofar as RLG is a central executive task, no evidence of the involvement of the central executive. The task x carry interaction in the error data initially appeared to provide support for the involvement of central executive processing in the carrying operation. However, as figure 15 demonstrates, this could be interpreted as a 'ceiling effect', and the lack of significant difference between the one-carry and two-carry conditions supports this interpretation. Overall, the pattern of results suggests that response times and error rates are measuring different aspects of cognitive load, and this distinction is supported by comparison of the RLG experiment with the articulatory suppression experiment. The error and response time results are markedly different between AS and RLG, with no main effect of task in the AS error data, and no significant task x carry interaction, and whereas the AS response time interactions indicate, at least speculatively, some disruption to phonological coding, there are no such interactions in the RLG response times. This suggests that the articulatory component of the RLG task is not directly relevant to the results.

4.6 Experiment four - random key pressing

Random key pressing potentially disrupts both the CE and the VSSP, insofar as randomly tapping keys can be considered to be a spatial task, but if the RKP task loads the central executive in the same way as RLG, then a similar pattern of results to those found for the RLG experiment might be predicted. This experiment differed from the previous experiments only with regard to the secondary task used. Instructions to participants were to tap the nine keys of a numeric keypad as randomly as possible, at a self-determined rate, with their non-dominant hand, while using their dominant hand to press the spacebar of a separate keyboard to record response times to the addition problems. A baseline measurement of RKP as a single task was obtained prior to the commencement of experimental trials, and this was compared with RKP under dual-task conditions. The same modified procedure used for the RLG experiment, based on Evans (1978), was used here, the numbers 1 – 9 being substituted for the letters A – I, in the calculation of the randomness index.

4.6.1 Random key pressing – response times

The untransformed mean response times for each task x carry x format level are given in table 8.

Dual -Task	Mean (SD)	Single-Task	Mean (SD)
ZCL	5439 (1520)	ZCL	4154 (1232)
ZCC	4833 (1550)	ZCC	3590 (1149)
OCL	7017 (1794)	OCL	5627 (1362)
OCC	6945 (2118)	OCC	5481 (1599)
TCL	8678 (2591)	TCL	7090 (2147)
TCC	8197 (2575)	TCC	6975 (2064)

Key: (zcl) zero carries linear format
(zcc) zero carries columnar format
(ocl) one carry linear format
(occ) one carry columnar format
(tcl) two carry linear format
(tcc) two carry columnar format

Table 8. Mean response times and standard deviations in milliseconds (Random key pressing)

A (2 x 3 x 2) repeated-measures ANOVA of the RKP response time data showed a significant difference between single-task (*M* = 5093 msec) and dual-task (*M* = 6427 msec), $F_{(1, 9)} = 48.93$, $p < 0.001$. A significant difference between zero-carry (*M* = 4266 msec), one-carry (*M* = 5984 msec) and two-carries (*M* = 7328 msec), $F_{(2,18)} = 64.14$, $p < 0.001$. A significant difference between linear format (*M* = 5929 msec) and columnar format (*M* = 5521 msec), $F_{(1, 9)} = 19.68$, $p < 0.01$. A significant task x carry interaction, $F_{(2, 18)} = 6.21$, $p < 0.05$, and a significant carry x format interaction, $F_{(2, 18)} = 4.67$, $p < 0.05$. No other interactions were significant.

Tests of within-subjects contrasts showed that RKP had a significantly greater effect on no-carry problems than two-carry problems, $F_{(1, 9)} = 7.97, p < 0.05$, and a significantly greater effect on one-carry problems than two-carry problems, $F_{(1, 9)} = 5.31, p < 0.05$. In other words, the impact of RKP decreased with increased number of carries. Tests of within-subjects contrasts demonstrated that the carry x format interaction resulted from significantly greater differences in linear versus columnar response times for no-carry problems than for both one-carry problems, $F_{(1, 9)} = 6.80, p < 0.05$, and two-carry problems, $F_{(1, 9)} = 4.90, p < 0.06$ (marginal). This suggests that the carry operation reduced the benefit of presentation in columnar format. Both interactions appear to indicate effects in the opposite direction to that expected, but this might be explained as an effect of the log transform of the response time data, and this possibility is discussed in section 4.6.3.

4.6.2 Random key pressing– error rates

A (2 x 3 x 2) repeated-measures ANOVA of the RKP transformed error index data showed a significant difference between zero-carry (M = 0.723), one-carry (M = 0.745) and two-carries (M = 0.768), $F_{(2, 18)} = 7.21, p < 0.05$. No other main effects or interactions were significant.

Tests of within-subjects contrasts indicated that no-carry problems had significantly fewer errors than both one-carry problems, $F_{(1, 9)} = 9.97, p < 0.05$, and two-carry problems, $F_{(1, 9)} = 10.51, p < 0.01$, but that one-carry and two-carry problems were not significantly different. Table 9 gives the untransformed mean error index values for each task x carry x format level.

Dual -Task	Mean (SD)	Single-Task	Mean (SD)
ZCL	0.011 (0.017)	ZCL	0.035 (0.043)
ZCC	0.021 (0.024)	ZCC	0.024 (0.036)
OCL	0.073 (0.078)	OCL	0.045 (0.036)
OCC	0.043 (0.063)	OCC	0.064 (0.058)
TCL	0.121 (0.144)	TCL	0.066 (0.059)
TCC	0.106 (0.075)	TCC	0.078 (0.082)

Key: (zcl) zero carries linear format
(zcc) zero carries columnar format
(ocl) one carry linear format
(occ) one carry columnar format
(tcl) two carry linear format
(tcc) two carry columnar format

Table 9. Mean error index and standard deviations (random key pressing)

The percentage error rates for the five different error types were not significantly different in single versus dual task conditions. There were no significant negative, or positive, correlations between each group of category response times and their associated error rates, providing no evidence for a speed/accuracy trade-off. All participants obtained a higher randomness index in dual task conditions (t = 6.97, p <0.001), and this met the stated criterion for accepting interference between primary and secondary tasks.

4.6.3 Random key pressing – interpretation and discussion of results

The pattern of response time data was similar, in terms of main effects, to the previous three experiments, but different in terms of the patterns of interaction. The task x carry interaction is plotted for transformed and untransformed response times in figure 16. The graph of the transformed data interaction appears to indicate that RKP had significantly less impact on carry than no-carry problems, which is not what could be expected if carrying involves central executive processing, and if RKP is a central executive task. However, the graph of the untransformed data does not support this interpretation, and the interaction in the transformed data might simply be the result of a multiplicative rather than an additive model – as discussed in section 4.3.1.

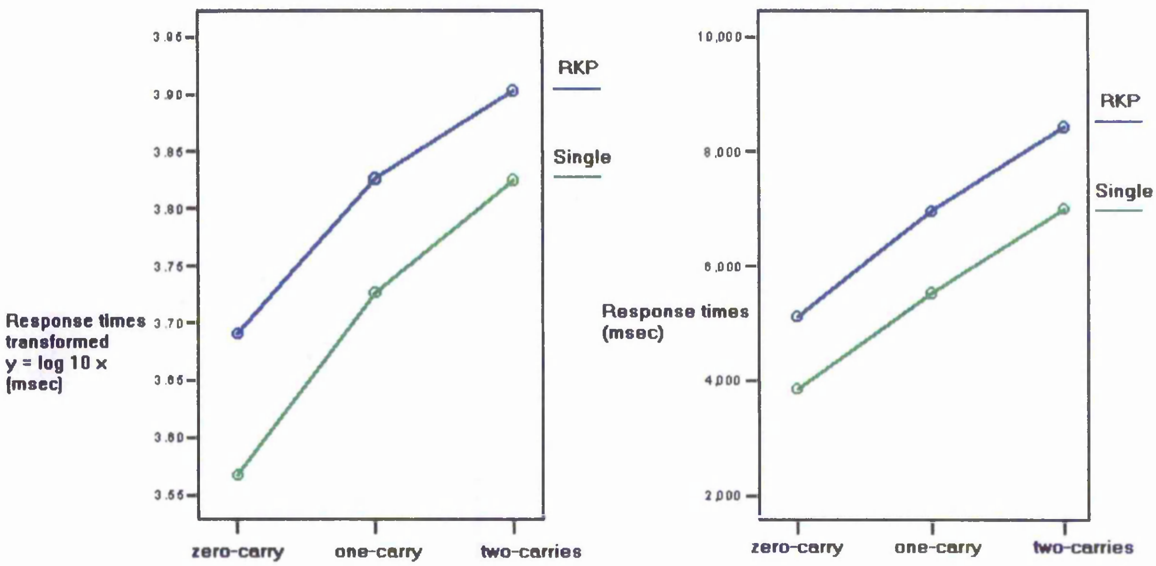

Figure 16. Graphs of the RKP task x carry interaction for transformed and untransformed response times

The format x carry interaction is plotted for transformed and untransformed response times in figure 17. The interaction in the transformed data appeared to indicate that over both single and dual-task conditions the superiority of columnar presentation was greater at the no-carry level than the carry levels. This, contrary to expectations, would suggest that columnar format does not provide better support for the carry operation. However, once again, the interaction in the transformed data is not evident in the untransformed data, and could simply stem from the log transformation of the response time scores.

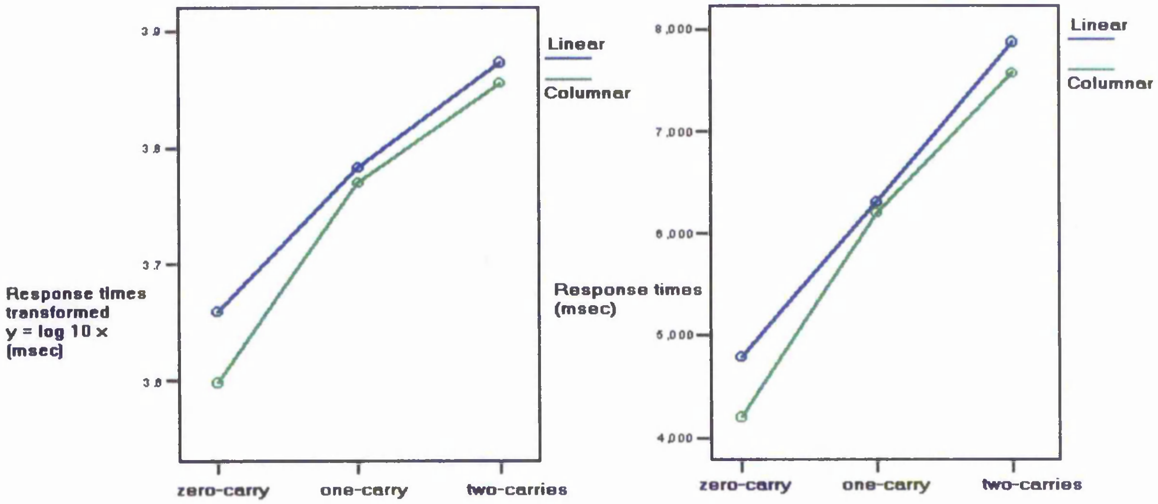

Figure 17. The RKP carry x format interaction for transformed and untransformed response times

The significant difference in the randomness index gives some support for RKP as an interference task, but this is not evident in the pattern of results for the error data, and if RKP does disrupt the central executive, it is seemingly a much less effective disruptor than RLG.

4.7. Experiment five – random interval generation

The Random Interval Generation technique (Vandierendonck et al., 1998) is proposed as a more *pure* measure of central executive involvement, as there is no associated articulation, and the only movement required of participants is to press a single key at *random* intervals. In this experiment participants pressed the number 1 key of a numeric keypad at self-determined *random* intervals, using the same apparatus and general procedure as in previous experiments. *Baseline* single-task measurement of RIG was carried out prior to the presentation of experimental trials, and the analysis of the randomness of the time sequences, in baseline and dual-task conditions, was achieved through the use of the *RIGANAL* program (Vandierendonck, 2000). This computer program calculates a variety of indices that provide an indication of the randomness of a sequence of time intervals.

4.7.1 Random interval generation– response times

Table 10 provides the untransformed mean response times for each task x carry x format level.

Dual -Task	Mean (SD)	Single-Task	Mean (SD)
ZCL	5904 (1691)	ZCL	5476 (1510)
ZCC	5312 (1899)	ZCC	4628 (1520)
OCL	8291 (1901)	OCL	8012 (1693)
OCC	7875 (1661)	OCC	7655 (1881)
TCL	10494 (2655)	TCL	9428 (2253)
TCC	9449 (2107)	TCC	9696 (2156)

Key: (zcl) zero carries linear format
(zcc) zero carries columnar format
(ocl) one carry linear format
(occ) one carry columnar format
(tcl) two carry linear format
(tcc) two carry columnar format

Table 10. Mean response times and standard deviations in milliseconds (Random interval generation)

A (2 x 3 x 2) repeated-measures ANOVA of transformed RIG response time data showed a significant difference between single-task (M = 6998 msec) and dual-task (M = 7430 msec), $F_{(1, 9)} = 8.99, p < 0.05$. A significant difference between zero-carry (M = 5093 msec), one-carry (M = 7762 msec) and two-carries (M = 9506 msec), $F_{(2, 18)} = 56.05, p < 0.001$. A significant difference between linear format (M = 7499 msec) and columnar format (M = 6934 msec), $F_{(1, 9)} = 17.81, p < 0.01$. A significant carry x format interaction, $F_{(2, 18)} = 6.77, p < 0.010$, and a significant task x carry x format interaction, $F_{(2, 18)} = 4.12, p < 0.05$.

Tests of within-subjects contrasts indicated that the carry x format interaction stemmed from significantly greater linear versus columnar format differences for no-carry problems compared to both one-carry problems, $F_{(1, 9)} = 7.97$, $p < 0.05$, and two-carry problems, $F_{(1, 9)} = 8.67$, $p < 0.05$.

As with the carry x format interaction in the RKP results, this again suggests that the carry operation reduces the benefit of columnar presentation, although here, the effect is modified by a significant task x carry x format interaction, which stemmed from RIG, in comparison to single-task performance, having a greater effect on two-carry problems in linear format, and a lesser effect on two-carry problems in columnar format, $F_{(1, 9)} = 5.82$, $p < 0.05$. The graphs in figures 18 and 19 provide a visual guide to this three-way interaction for transformed and untransformed scores.

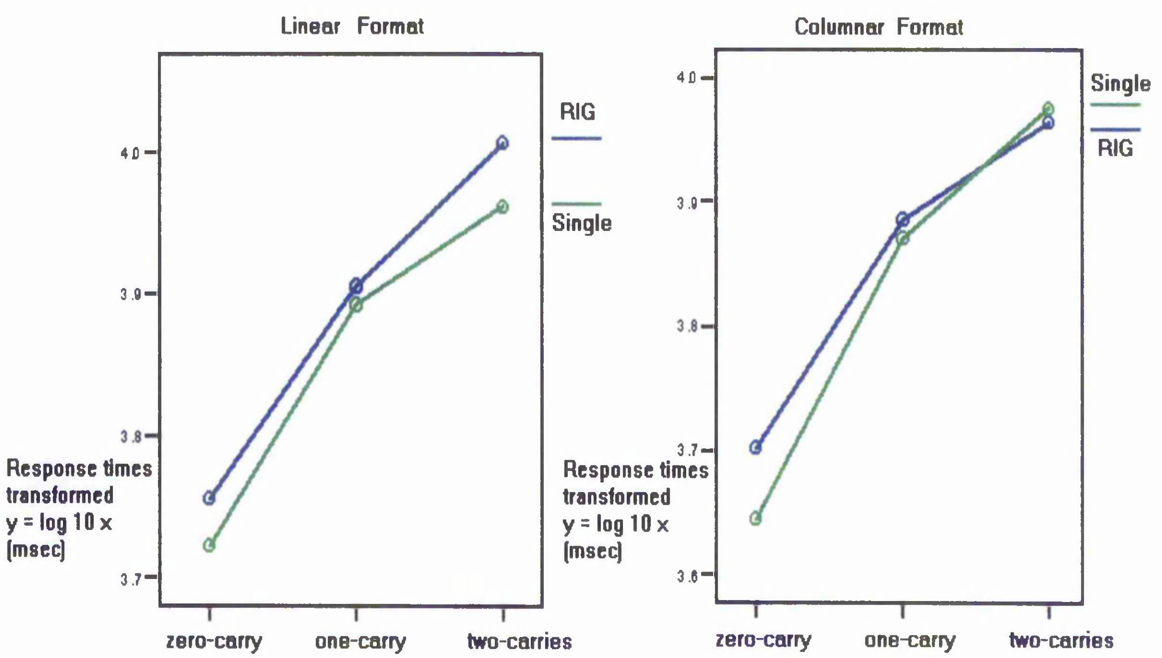

Figure 18. The RIG task x carry x format interaction (log transformed response times)

Looking across the two graphs in figure 18, the advantage of columnar presentation is evident in the RLG line plots, but although single-task zero-carry and one-carry columnar format problems also show this advantage, the single-task two-carry columnar format problems clearly have longer response times than the equivalent problems in linear format.

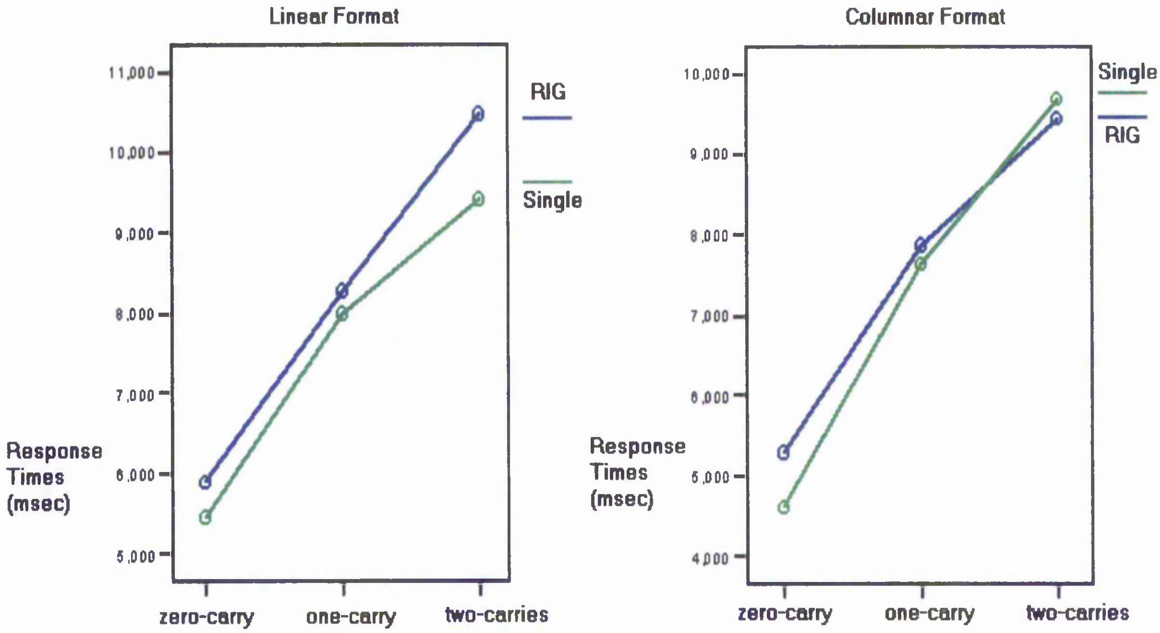

Figure 19. The RIG task x carry x format interaction (untransformed response times)

The interaction shown in figure 19 for the untransformed RIG response times is itself significant at $p < 0.05$, and is interpreted as for the interaction for transformed response times in figure 18. This task x carry x format interaction is not considered to be the result of the log transformation of the response time scores.

4.7.2 Random interval generation – error rates

Table 11 provides the untransformed mean error index values at each task x carry x format level. A (2 x 3 x 2) repeated-measures ANOVA of the transformed mean error index indicated a significant difference between zero-carry (M = 0.733), one-carry (M = 0.773) and two-carries (M = 0.813), $F_{(2,\ 18)} = 19.48$, $p < 0.001$. No other main effects or interactions were significant. Tests of within-subjects contrasts indicated that no-carry problems had significantly fewer errors than one-carry problems, $F_{(1, 9)} = 12.15$, $p < 0.01$, and two-carry problems had significantly fewer errors than one-carry problems, $F_{(1, 9)} = 13.29$, $p < 0.01$.

Dual -Task	Mean (SD)	Single-Task	Mean (SD)
ZCL	0.041 (0.031)	ZCL	0.055 (0.053)
ZCC	0.026 (0.025)	ZCC	0.028 (0.035)
OCL	0.126 (0.110)	OCL	0.108 (0.076)
OCC	0.084 (0.091)	OCC	0.082 (0.036)
TCL	0.153 (0.101)	TCL	0.152 (0.079)
TCC	0.166 (0.106)	TCC	0.192 (0.157)

Key: (zcl) zero carries linear format
(zcc) zero carries columnar format
(ocl) one carry linear format
(occ) one carry columnar format
(tcl) two carry linear format
(tcc) two carry columnar format

Table 11. Mean error index and standard deviations (Random interval generation)

There were no significant differences in the types of errors made under dual and single task conditions. There was no evidence, in the form of significant negative correlations, for a speed / accuracy trade-off between response times and errors, but one category pairing of response times and errors (single-task no-carry columnar) provided a significant positive correlation, $r = 0.648$, $p < 0.05$. The comparison of randomness used sequences of thirty time intervals entered into the *RIGANAL* program (Vandierendonck, 2000), with the Alternation Index output providing the simplest interpretation. This index varies between 0 and 1; values less than 0.5 indicate a tendency towards perseveration – similar time intervals, rates higher than 0.5 indicate a tendency towards alternation – patterns of longer then shorter time intervals. An index of ~ 0.50 indicating lack of bias either way. Vandierendonck (2000) states that human RIG is biased towards alternation and that this increases with increased primary task load. The calculated index for each participant in this study always showed perseveration or no bias, never rising above 0.54 for any participant, with an average baseline index value of 0.24, and no significant difference between baseline and dual-task performance. This could indicate that participants were not sufficiently motivated to produce random intervals, that they chose to allocate resources to the primary rather than the secondary task, or that RIG did not disrupt performance on the arithmetic task.

4.7.3 Random interval generation – interpretation and discussion of results

There is no direct support in the response time data for the disruptive effect of RIG on the carry operation overall, as there is no significant task x carry interaction, and the pattern of the task x carry x format interaction, in both transformed and untransformed scores, is consistent only with a disruptive effect of RIG on two-carry problems in linear format. The error data shows a significant main effect of the carry operation, and nothing else. If carrying does involve the central executive, then there is no evidence here for RIG as a central executive task. One possible interpretation is that in contrast to the RLG task articulations, participants were aware that they could not be directly monitored for their ability to demonstrate random intervals, and therefore did not allocate resources to the RIG task.

4.8 Analysis of strategy use

Prior to participating in the experiments, all participants completed a *verbal protocol* task to identify their use of particular addition strategies in the solution of six multi-digit additions. Participants were encouraged to negotiate their membership of particular strategy groups, and further strategy groups were constructed if participants offered strategies that were not consistent with the groups already identified. In total, five different strategy groups were found, and these are described as follows: *Conventional* – add units first, then tens, then hundreds, *Reversed* - add tens first, then units, and make adjustments as required, *Add units* – add units to larger sum, and then add tens, e.g. 135 + 43 = 138 + 40, *Add tens* – add tens to larger sum, and then add units, e.g. 135 + 43 = 175 + 3, *Mixture* – more than one of the above strategies, including estimating, rounding and adjustment. The number of participants assigning themselves to each group varied; *Conventional* (N = 15), *Reversed* (N = 8), *Add units* (N = 7), *Add tens* (N = 3), *Mixture* (N = 14), and although these groups could not be compared on the different secondary tasks, the single-task (arithmetic alone) was common to all 47 participants, and so the single-task mean response times and error rates for these groups could be compared. Given the different group sizes, Kruskal-Wallis tests were used, in place of one-way ANOVA, to examine each category of response times and error rates, but no significant difference between strategy groups emerged, suggesting either that all strategies were equally effective, or more probably, that most participants actually used a mixture of strategies.

4.9 Summary and general discussion of first series of experiments

The purpose of these five experiments was to establish which, if any, of the WM components are active when participants solve multi-digit additions with continuous visual presentation under single and dual-task conditions, and to explore a number of methodological issues relating both to the equivalence of various *executive* tasks and the general dual-task experimental rationale. The involvement of the phonological loop was addressed in the first experiment utilising articulatory suppression as the secondary task. The analysis of response times provided no direct evidence for phonological coding in the carry operation overall, but the three-way interaction showed a significant task x carry interaction for additions presented in linear format. This was tentatively interpreted as a possible role for phonological coding in the re-representation of linear additions in columnar format, which would provide indirect support for the triple-code model (Dehaene and Cohen, 1995), although the error rate data showed no supporting evidence of disruption due to concurrent articulatory suppression. How does this finding relate to previous studies?

Logie, Gilhooly and Wynn (1994) did find an effect of articulatory suppression on *running addition* with brief visual presentation of addends (1 second per addend), as did Heathcote (1994) also with brief visual presentation, whereas in the study of Furst and Hitch (2000) concurrent articulatory suppression did not disrupt performance on additions with continuous visual presentation. The articulatory suppression results in the present study therefore fit between these two positions, suggesting that under continuous visual presentation, the effects of articulatory suppression are mediated by format of presentation. The role of visuo-spatial processing was investigated in two ways, disruption of the spatial rehearsal mechanism through spatial tapping, and more generally, through the two levels of the format variable – linear versus columnar presentation. Heathcote's (1994) finding that format of presentation affected solution times was supported, but the response time data in the spatial tapping experiment did not provide any support for spatial processing in the carry operation, and the spatial tapping error data showed not only no main effect for task, but also a significant task x carry interaction that indicated results contrary to expectations. Again, this appears to provide no support for spatial processing in the carry operation.

However, the main effect of format was significant for response time data in all five experiments, see figure 20, but not significant for error rates in any of the experiments.

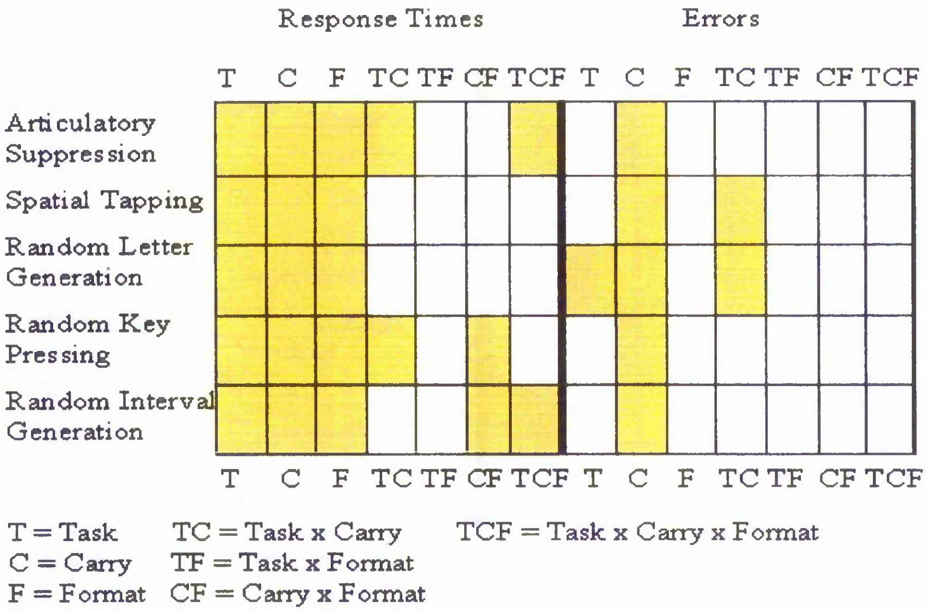

Figure 20. Summary of significant main effects and interactions (yellow segments) for all five experiments

This suggests that the visual format of presentation, and hence an aspect of visuo-spatial processing, is important in the time taken for the extraction or transcoding of problem information, but not in terms of the processes underlying calculation accuracy. Central executive involvement was investigated using three separate executive tasks, Random Letter Generation (RLG), Random Key-Pressing (RKP) and Random Interval Generation (RIG). The response time data for these three experiments provided no evidence for the disruption of the carry operation. Analysis of RLG error rates provided the only evidence of disruption to the carry operation, and this evidence was ambiguous given the potential 'ceiling effect' in the RLG error scores. If RLG does disrupt the carry operation, then this suggests that the central executive has a role in holding and manipulating numbers, totals or partial totals involving carries, but that this is reflected in increased error rates rather than increased response times. This is to some extent consistent with the suggestion of Logie et al. (1994) that there might be two separate cognitive systems; one dealing with accuracy and one dealing with keeping track of carries, and similarly, with the suggestion of Noel, Desert, Aubrun and Seron (2001) that the central executive might not be recruited for the calculation itself, but for the selection and comparison of responses.

The response time data is consistent over all five experiments where significant main effects of task, carries and format were observed in all cases, although in the general absence of appropriate interactions, the response times for each factor are assumed to be simply additive. Of the five secondary tasks, only RLG and RKP demonstrated significant secondary task decrement, thereby meeting the stated criterion for dual-task interference, and it is therefore questionable whether the forms of AS, ST and RIG used in these experiments had any appreciable interference effect on the multi-digit addition task, or more precisely, on the WM components involved in the processing of these particular multi-digit additions. If the increase in response times cannot be attributed to interference between task and problem factors competing for the same processing resources at the same time, then is it that the tasks are processed sequentially? If so, then this suggests that the main effect of the task variable can be explained as *cost of concurrence*; it takes more time to process two tasks than one task.

The increase in response time with the number of carries is similarly interpreted as increased processing time with increasing number of processing operations. The increase in response times, for additions presented in linear rather than columnar format, suggests a difficulty in the extraction of information from the linear format, with the possible explanation that linear format involves increased reading or encoding time, or more speculatively, that a recoding operation is necessary to re-represent the additions in columnar format. The main effect of carries is found in the error rate data for all five experiments, and clearly, the potential to make errors increases with the number of processing operations undertaken. There was no main effect of format for error rates in any of the five experiments, and no significant interactions involving the format variable, which suggests that it is unlikely that response times and error rates are measuring the same underlying process. Perhaps response times simply indicate the time taken to extract information from additions, while error rates provide a more direct measure of central executive involvement in post encoding processing.

4.10 Limitations of first series of experiments –statistical power, errors, strategy and maths anxiety

To argue that the lack of task x carry interactions in these experiments is meaningful, is to assume that the experiments had sufficient statistical power to detect such interactions if they do in fact exist, and this largely depends upon the effect size of the interaction and the number of participants in each experiment.

If effect size for the task x carry interactions is not large, then with small sample sizes, such as N = 10, power is low, and the lack of significant interactions in the experiments might not be a true reflection of the disruptive effect of the secondary tasks on the carry operation. However, across all five experiments, the main effects of task and carries, where significant, are relatively robust, and arguably, if a secondary task has the ability to disrupt the carry operation then this should produce a large difference in response times and error rates. Figure 21, gives a plot of power against effect size for the typical task x carry interactions reported in the first series of experiments.

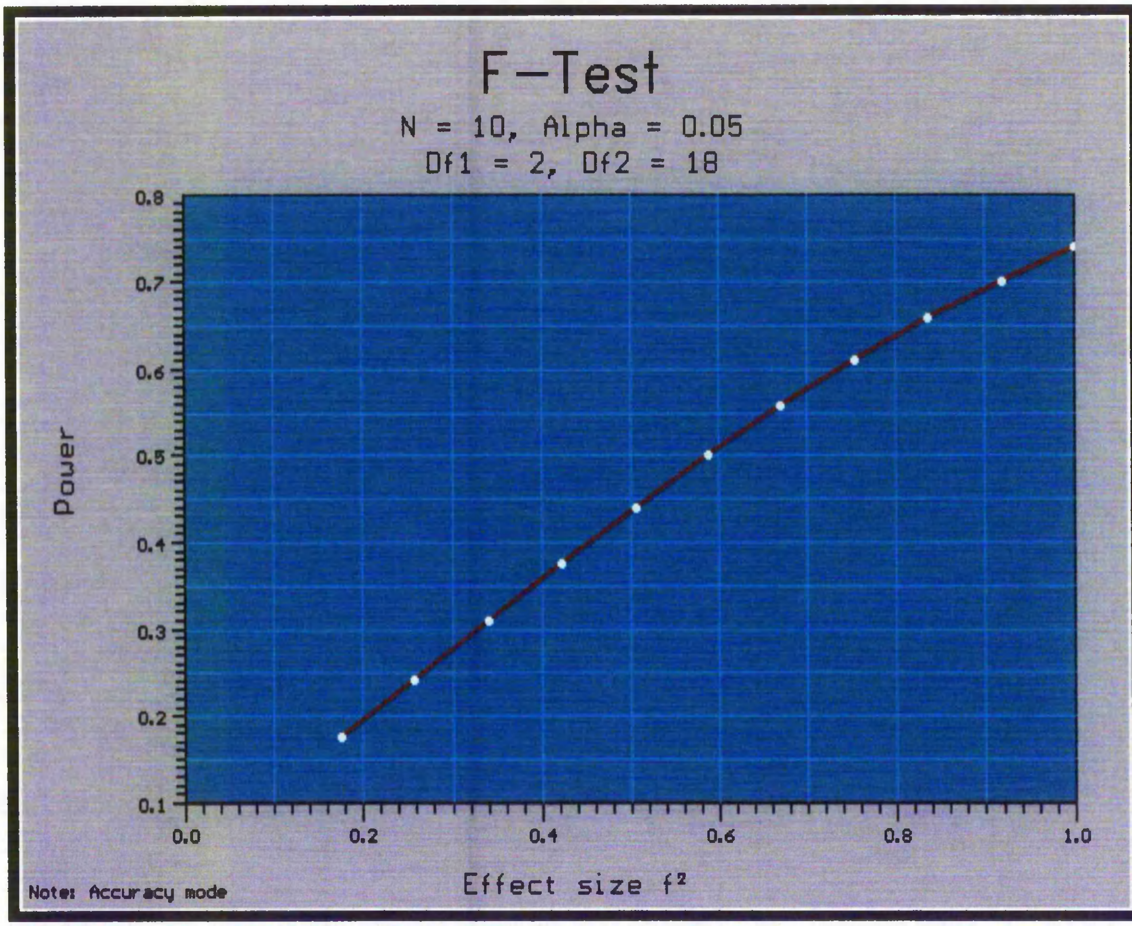

Figure 21. GPOWER (Faul and Erdfelder, 1992) plot of effect size against power, F (2, 18) N = 10

There is no reason to expect a small effect size, but nevertheless, it would be better to improve statistical power, although this has to reflect a compromise between the pragmatics of the research process and *acceptable* levels of statistical power, for to detect the smallest effect sizes would require sample sizes so large as to make research entirely impracticable. There is also the issue of what such small effect sizes demonstrate, for arguably, effects have to be reasonably large to be psychologically meaningful. Figure 22 demonstrates the changes to statistical power by doubling N from 10 to 20 when investigating the same task x carry interactions.

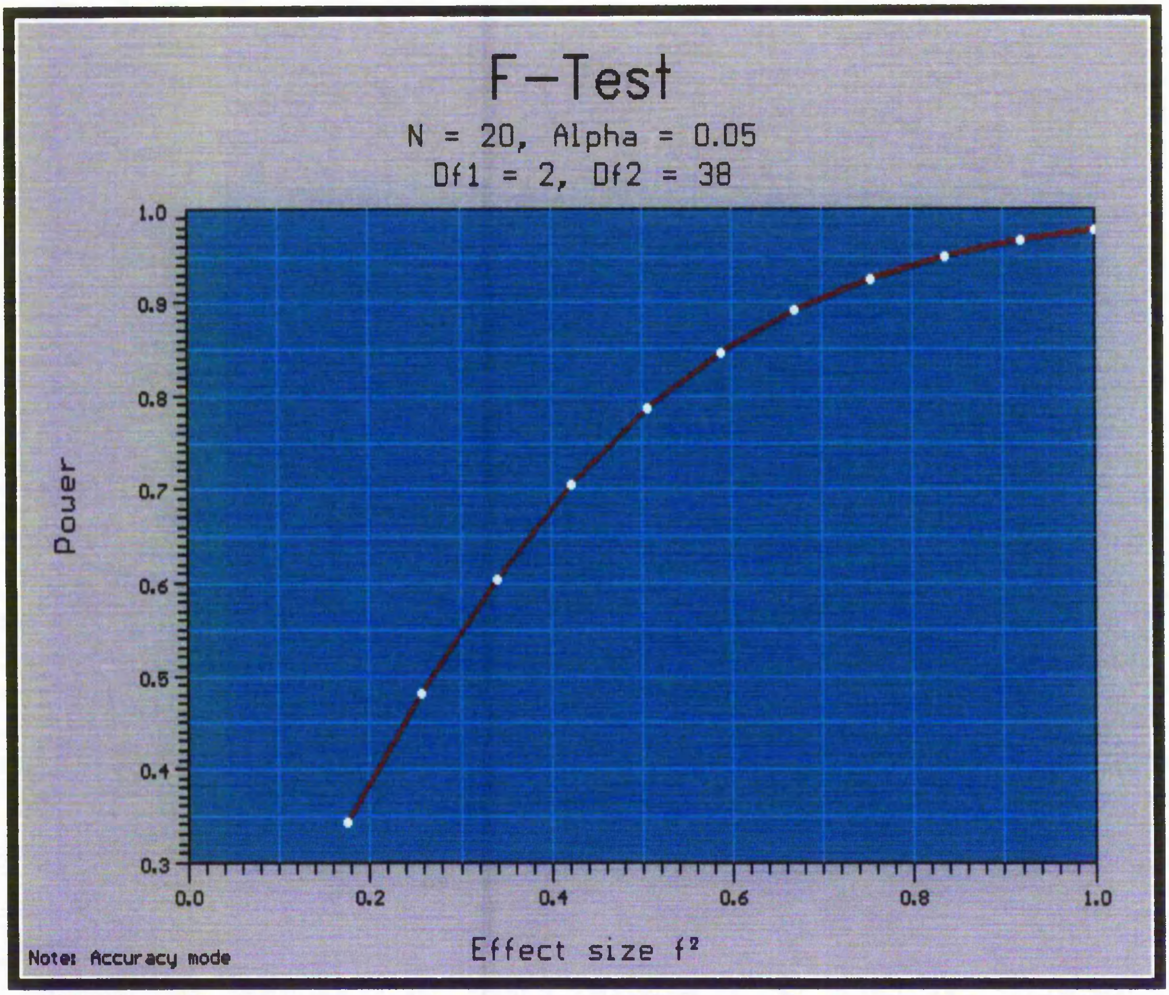

Figure 22. GPOWER (Faul and Erdfelder, 1992) plot of effect size against power, F (2, 38) N = 20

At N = 20, in figure 15, *large* effect sizes (0.35 +) have power levels in excess of 0.6, and this is considered to represent a reasonable compromise for the design of further experiments.

The analysis of the five error types across all five experiments demonstrated no significant differences, and this clearly is not an analysis worth repeating in further experiments, but beyond this, analysis of response times did not separate response times for error problems from response times for correct problems, and if error problems have greater response times, this could potentially increase mean response times for conditions with greater numbers of errors, and although this is not considered to be remotely sufficient to explain the differences in response times across conditions, it is something to incorporate into further experimental analysis.

The attempt to identify particular strategy use through analysis and negotiation of *verbal protocols* is clearly insufficient if strategy use changes with different problem types, and strategy use in further experiments will need to be assessed differently to take account of this possibility. Performance on the addition tasks varied considerably within the different experimental groups, and while individual differences in performance might be expected due to individual differences in speed of processing, it appeared that some participants were more confident of their arithmetic ability than justified by their error scores, perhaps demonstrating rather more maths anxiety than appropriate, and for this reason future experiments will require participants to possess either recognised qualifications in mathematics, or evidence of adequate mental arithmetic skills.

Chapter Five – rationale for further experimental study

5.1 General design of second series of experiments

A second series of experiments was designed and implemented. These experiments extended the investigation of WM involvement in cognitive arithmetic, and sought to overcome the limitations in the first series of experiments by taking advantage of theoretical and methodological advances identified in the literature during and after the period of the first experiments. The second series of experiments took place during 2002 – 2004, and involved an increase in participant numbers to improve statistical power, and a completely within-participants design, so that comparisons could be made within and between experimental conditions for the same participants. The number of secondary tasks was reduced to two; articulatory suppression to investigate phonological processing, and random letter generation, this time using a smaller subset of the alphabet (A, B, C, D, E), to investigate central executive processing, as these were the only secondary tasks from the first series of experiments that provided evidence of WM involvement. As in the first experiments, the secondary tasks were measured in both *baseline* and dual-task conditions to identify any secondary task decrement as an indicator of competition for processing resources.

Further multi-digit problems were constructed to investigate any differential WM involvement across the four arithmetic operations (addition, subtraction, multiplication and division), thereby providing an opportunity for comparisons with previous studies supporting the various cognitive models of arithmetic. The number of carry levels was reduced to two; either zero-carry or one-carry, to give a clear analysis of no-carry / carry conditions, and to reduce problem difficulty for participants, as pilot studies indicated that finding the solution of two-carry multiplications and divisions under dual-task conditions was often simply too difficult, or at least close to the upper limit of performance.

Item analysis was conducted on the group of arithmetic problems representing each level of the experimental designs, to check the possibility that particular problems might produce response times statistically different to the mean group response times, thereby providing the potential for differences not due to the experimental manipulations.

Following Hecht (2002), strategy was more directly assessed by asking participants to give their solution strategy, from a list of solution strategies, at the end of every problem, and this allowed the analysis of the effect of strategy choice on response times across the different combinations of factor levels in order to answer two questions; does strategy choice change with problems at different levels, e.g. do participants use different strategies for carry versus no-carry problems, and are some strategies faster than others for the same type of problem?

Format of presentation was considered to be sufficient to identify aspects of visuospatial processing, and this remained the same as in the first experiment for additions and subtractions (linear – columnar), but was further manipulated for divisions (linear – traditional - columnar), while another factor (difficulty) was introduced and manipulated for multiplication problems (easy - hard). The design for the addition and subtraction experiments became 3 x 2 x 2, Task – Single, AS, RLG, Format – linear, columnar, and Carry – 0, 1, allowing a partial replication of the addition experiments in the first series. The multiplication experiment had a 3 x 2 x 2 x 2 design, Task – Single, AS, RLG, Difficulty – easy, hard, Format – linear, columnar, and Carry – 0,1, and the division experiment had a 3 x 3 x 2 design, Task – Single, AS, RLG, Format – linear, traditional, and columnar, and Carry – 0,1. The three levels of the task variable allowed the analysis to be broken down into separate analyses of performance at the different combination of levels; single versus AS, single versus RLG, and AS versus RLG. Response times and error rates again constituted the dependent variables, and general procedure was as described for the first series of experiments.

Twenty participants, 10 males and 10 females, ranging in age from 17 – 42 years with an rounded average age of 25 years, were recruited from sixth form, undergraduate and postgraduate populations, following advertisements at various UK institutions, and in order to provide better control of maths anxiety, all were required to be involved in the continuing study of mathematics in some way, and to have at least a minimum GCSE pass in mathematics, although most possessed mathematics qualifications at higher levels. The same twenty participants completed all experimental trials across the four arithmetic operations, the total experimental duration per participant being between 3 – 4 hours, spread over two sessions each of ~ 2 hours duration, and participants were paid a set fee of £20 for their participation to reflect appreciation of their commitment to the experimental tasks over this time period.

5.2 Arithmetic tasks – second series of experiments

The presentation of arithmetic problems, in all four arithmetic operations, was achieved using the same programs and techniques as in the earlier addition experiments (see section 4.2), except that problem presentation was not automatic following pressing of the response key, and required a separate key to be pressed before onset of presentation 1500 ms later. This gave a period of time, after response time was recorded, in which strategy use could be reported. The problems for the addition task in the second experiments were the same as those used in the first experiments (appendix A), except that the sub-set of two-carry problems was removed, giving 240, twelve sets of twenty, additions in total, with answers in the range from 157 to 199. The 240 subtraction problems were constructed to give a similar range of totals to the additions, 122 to 167, and again consisted of twelve sets of twenty problems, the four categories of zero-carry linear format, zero-carry columnar format, one-carry linear format and one-carry columnar format, at each of the three task levels; Single, AS and RLG. Examples of the subtraction problems, as bitmaps, are given in figure 23, (see appendix C) for the full set of problems.

$$198 - 34 = \underline{\qquad}$$

$$\begin{array}{r} 172 \\ -\ \underline{45} \\ = \underline{\quad\ \ } \end{array}$$

Figure 23. Examples of subtraction problems produced as bitmaps.

The multiplication problems were constructed to include another problem factor (difficulty), with two levels; *easy*, three-digit x one-digit problems, and *hard*, two-digit x two-digit problems, to investigate the possibility that WM involvement is influenced largely by the relative level of problem complexity, such that easy problems may only marginally involve WM resources, whereas difficult problems might require WM resources to the full. Figure 24 provides examples of the bitmap multiplications, but also illustrates the difficulty of identifying the decomposition of operations and the strategy used, when more difficult problems are compared to easy problems. Easy no-carry problems, in both linear and columnar format, can essentially be solved by *reading and doubling* without recourse to more complex calculation, while knowledge of table facts and appropriate use of strategy can in fact reduce carry problems to no-carry problems. This is further justification for the recording of strategy use with the presentation of each individual problem, for there is clearly potential for differences in response times with the use of different strategies.

Easy / No-Carry / Columnar

$$\begin{array}{r} 132 \\ \times\ \ 2 \\ \hline \\ \hline \end{array}$$

Read and Double?

Easy / No-carry / Linear

$134 \times 2 =$ ______

Read and Double?

Easy / Carry / Columnar

$$\begin{array}{r} 126 \\ \times\ \ 2 \\ \hline \\ \hline \end{array}$$

2 + carry + 40 + 200
12 + 240
240 + 12

Easy / Carry / Linear

$129 \times 2 =$ ______

8 + carry + 40 + 200
18 + 240
240 + 18

Hard / Carry / Columnar

$$\begin{array}{r} 26 \\ \times\ 12 \\ \hline \\ \hline \end{array}$$

2 + carry + 40 + 260
72 + 240
260 + 52
360 - 48

Hard / Carry / Linear

$29 \times 12 =$ ______

8 + carry + 40 + 290
18 + 40 + 290
360 - 12
240 + 108

Hard / No-Carry / Columnar

$$\begin{array}{r} 32 \\ \times\ 12 \\ \hline \\ \hline \end{array}$$

4 + 60 + 320
24 + 360
64 + 320

Hard / No-Carry / Linear

$41 \times 12 =$ ______

2 + 80 + 410
12 + 480
82 + 410

Figure 24. Examples of multiplications produced as bitmaps (given here with possible solution strategies)

A total of 216 multiplication problems were created (see appendix D), nine problems in each of the eight categories shown in figure 24, giving 72 problems for each of the three levels of task, Single, AS and RLG. The range of totals for the hard problems varied from 156 to 961, whereas the range of totals for the easy problems varied from 226 to 987.

Division problems were constructed to investigate the format variable in more detail, by including, in addition to linear and traditional formats, a format of presentation (columnar), not usually associated with division. The purpose of introducing a columnar format was to ascertain if this would lead to increased response times and errors, due to columnar format having stronger associations with the other arithmetic operations. Examples of these formats are given in figure 25.

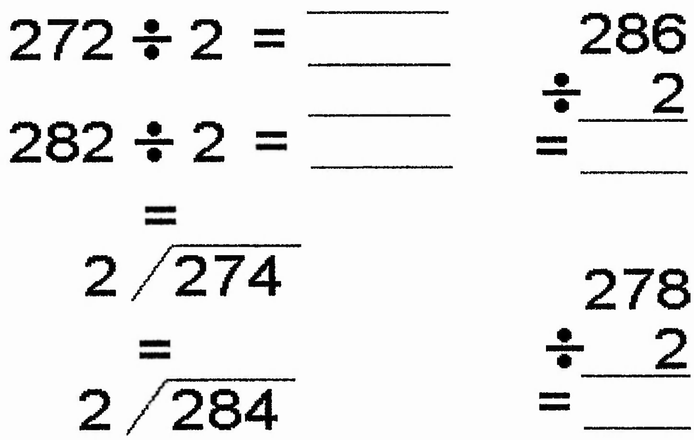

Figure 25. Examples of divisions produced as bitmaps

A total of 162 division problems were constructed (see appendix E), nine for each of the six categories represented in figure 25, at each of the three levels of the task variable, Single, AS and RLG. Answers to the divisions ranged in total from 112 to 448. The order of presentation for the blocks of trials in the addition, subtraction, multiplication and division experiments was counter-balanced in accordance with the table in appendix F.

5.3 The addition experiment

This dual-task addition experiment investigated the role of both phonological and central executive processes, through the use of articulatory suppression and random letter generation, when 20 participants solved a reduced sub-set of the addition problems from the first series of experiments. In effect, this experiment served as a partial replication of two of the earlier experiments, and as previously, evidence of WM involvement was sought in the pattern of interactions between task, carry and format variables.

5.3.1 Addition experiment – correct answer response times

As in the first series of experiments, the transformation y = log 10 x was used here, and for all further response time data, to give greater adherence to the assumptions underlying the use of ANOVA, and although tables will continue to report untransformed mean response times in milliseconds, graphs will be given for both transformed and untransformed scores where necessary. The response times in table 12, and in subsequent tables, reflect only correct answers, for incorrectly answered problems were removed from the data set, along with response times greater than two standard deviations from the mean, and means and standard deviations recalculated accordingly prior to analysis.

Level	Mean	Std. Deviation
Single Zero-Carry Linear	3525	1068
Single Zero Carry Columnar	3280	1004
Single One-Carry Linear	4421	1439
Single One-Carry Columnar	4126	1342
AS Zero-Carry Linear	4223	1339
AS Zero Carry Columnar	3942	1360
AS One-Carry Linear	5399	2146
AS One-Carry Columnar	5237	2254
RLG Zero-Carry Linear	5982	2521
RLG Zero Carry Columnar	5481	2230
RLG One-Carry Linear	7892	4408
RLG One-Carry Columnar	7306	4178

Table 12. Correct item mean response times and standard deviations in milliseconds (addition experiment)

A 3 (task) x 2 (carry) x 2 (format) repeated-measures ANOVA of log transformed response times demonstrated a significant difference between single-task (M = 3622 msec), AS (M = 4375 msec) and RLG (M = 6026 msec), $F_{(2, 38)} = 54.02$, $p < 0.001$. A significant difference between no-carry (M = 4055 msec) and one-carry (M = 5140 msec), $F_{(1, 19)} = 43.96$, $p <$ 0.001. A significant difference between linear format (M = 4732 msec) and columnar format (M = 4416 msec), $F_{(1, 19)} = 34.77$, $p < 0.001$. None of the interactions were significant.

Further analysis, using within-subjects contrasts and separate 2 x 2 x 2 ANOVA of paired task levels, allowed Single vs. AS, Single vs. RLG and AS vs. RLG comparisons. The Single level vs. AS comparison demonstrated a significant effect of task, $F_{(1, 19)}$ =

38.05, $p < 0.001$, a significant effect of carry, $F_{(1, 19)} = 50.26$, $p < 0.001$, a significant effect of format, $F_{(1, 19)} = 32.27$, $p < 0.001$, and a marginally significant task x carry x format interaction, $F_{(1, 19)} = 4.28$, $p < 0.06$, which reflected a significant carry x format interaction under AS, such that one-carry linear and columnar problems were relatively equally impaired in comparison to no-carry problems, where columnar format gave faster response times.

The Single vs. RLG comparison revealed a significant effect of task, $F_{(1, 19)} =$ 62.50, $p < 0.001$, a significant effect of carry, $F_{(1, 19)} = 41.82$, $p < 0.001$, and a significant effect of format, $F_{(1, 19)} = 33.07$, $p < 0.001$. The AS vs. RLG comparison demonstrated a significant effect of task, $F_{(1, 19)} = 45.14$, $p < 0.001$, a significant effect of carry, $F_{(1, 19)} =$ 38.17, $p < 0.001$, and a significant effect of format, $F_{(1, 19)} = 26.42$, $p < 0.001$. Analyses of response times and error scores, here and in subsequent experiments, routinely report Greenhouse-Geisser corrected probabilities where appropriate.

5.3.2 Addition experiment – error scores

One advantage of recording strategy use after the presentation of each problem, was that the slowing down of problem presentation, resulting from the use of a different onset key, eliminated problem omissions due to either the technical problem of *key-bounce*, or due to participants double-pressing the response key, thereby reducing the number of missed problems to zero, so that error scores could be given directly, e.g. 3 errors from a sub-category of 20 problems giving an error score of 3.

However, as in the first series of experiments, the transformation $y = \text{sqrt}(x + 0.5)$ gave better adherence to the underlying ANOVA assumptions, and all analyses and graphs are based on these transformed scores, although untransformed mean error rates will continue to be reported in tables. The untransformed mean error scores for the addition experiment are given in the table 13.

Level	Mean	Std. Deviation
Single Zero-Carry Linear	.70	.86
Single Zero Carry Columnar	.30	.73
Single One-Carry Linear	.85	.93
Single One-Carry Columnar	.90	.72
AS Zero-Carry Linear	.80	.69
AS Zero Carry Columnar	.70	.80
AS One-Carry Linear	1.50	.94
AS One-Carry Columnar	1.15	1.18
RLG Zero-Carry Linear	1.05	.94
RLG Zero Carry Columnar	.80	.69
RLG One-Carry Linear	2.90	1.29
RLG One-Carry Columnar	2.25	1.91

Table 13. Untransformed mean error scores and standard deviations (addition experiment)

A 3 x 2 x 2 repeated-measures ANOVA demonstrated a significant difference between single-task (M= 1.028), AS (M= 1.178) and RLG (M= 1.406), $F_{(2, 38)} = 12.51, p < 0.001$. A significant difference between no-carry (M= 1.050) and one-carry (M= 1.358), $F_{(1, 19)} = 16.45$, $p < 0.001$. A significant difference between linear format (M = 1.265) and columnar format (M = 1.143), $F_{(1, 19)} = 10.36$, $p < 0.01$. A significant task x carry interaction, $F_{(2, 38)} = 7.87$, $p < 0.001$, and a marginally significant task x carry x format interaction, $F_{(2, 38)} = 3.07, p < 0.07$.

Further 2 x 2 x 2 ANOVA and tests of within-subjects contrasts provided the following comparisons: Single vs. AS, a significant effect of task, $F_{(1, 19)} = 4.64, p < 0.05$, a significant effect of carry, $F_{(1, 19)} = 5.90, p < 0.05$, and a significant effect of format, $F_{(1, 19)} = 5.90, p < 0.05$. None of the interactions were significant.

Single vs. RLG, a significant effect of task, $F_{(1, 19)} = 18.03, p < 0.001$, a significant effect of carry, $F_{(1, 19)} = 22.95, p < 0.001$, a significant effect of format, $F_{(1, 19)} = 5.97, p < 0.05$, a significant task x carry interaction, $F_{(1, 19)}$ 12.90, $p < 0.01$. One-carry problems had greater errors under RLG, modified by a marginally significant task x carry x format interaction, $F_{(1, 19)} = 4.22$, $p < 0.06$, such that under single- task conditions, one and no-carry problems in linear format showed similar errors, whereas no-carry problems in columnar format showed significantly less errors than one-carry problems.

AS vs. RLG, a significant effect of task, $F_{(1,19)} = 11.34$, $p < 0.01$, a significant effect of carry, $F_{(1, 19)} = 18.95$, $p < 0.001$, a significant effect of format, $F_{(1, 19)} = 10.38$, $p < 0.004$, and a significant task x carry interaction, $F_{(1, 19)} = 12.82$, $p < 0.01$, showing that RLG affected the carry operation and not AS.

5.3.3 Addition experiment – interpretation and discussion of results

Figure 26 compares results from the first series of experiments with the addition experiment results.

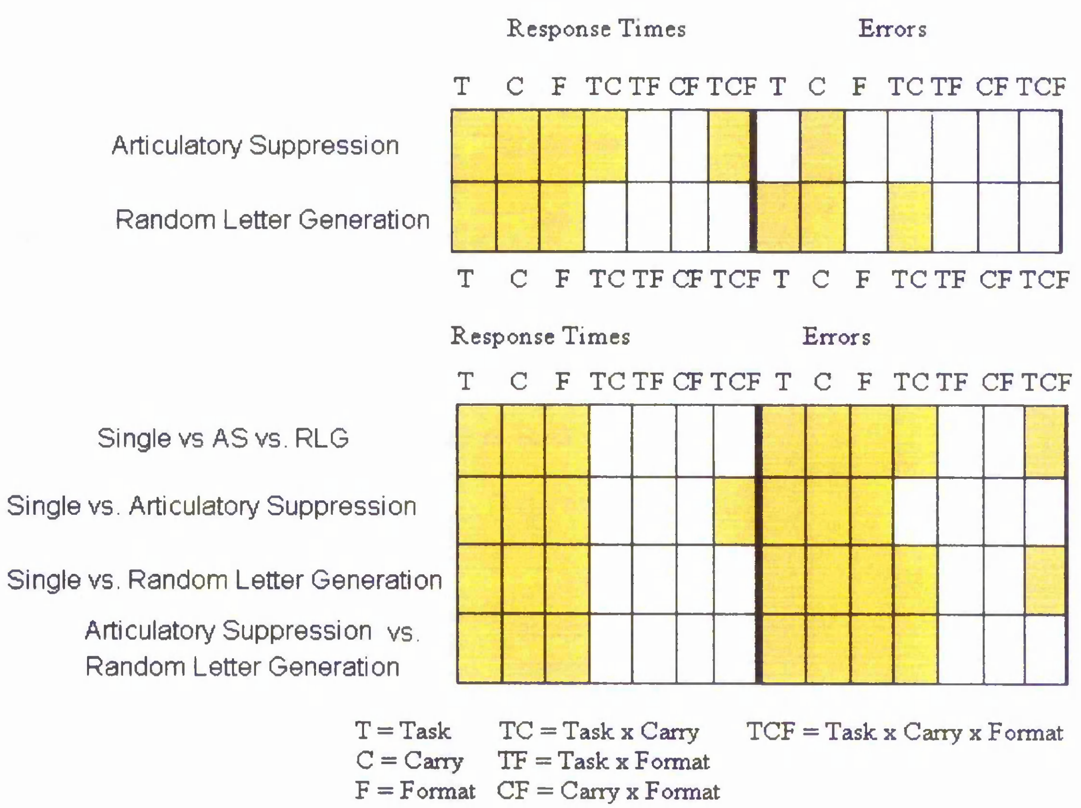

Figure 26. Comparison of first and second addition experiments

As can be seen from figure 26, the effect of articulatory suppression on response times is similar in both the first and second experiments for the main effects of task, carry and format. The task x carry x format interaction in the first experiment demonstrated that AS gave a differentially greater increase in response times for one-carry problems only when these were presented in linear format. However, as figures 27 and 28 demonstrate, the AS task x carry x format interaction in the second experiment gave a differentially greater increase in response times for one-carry problems only when these were presented in columnar format.

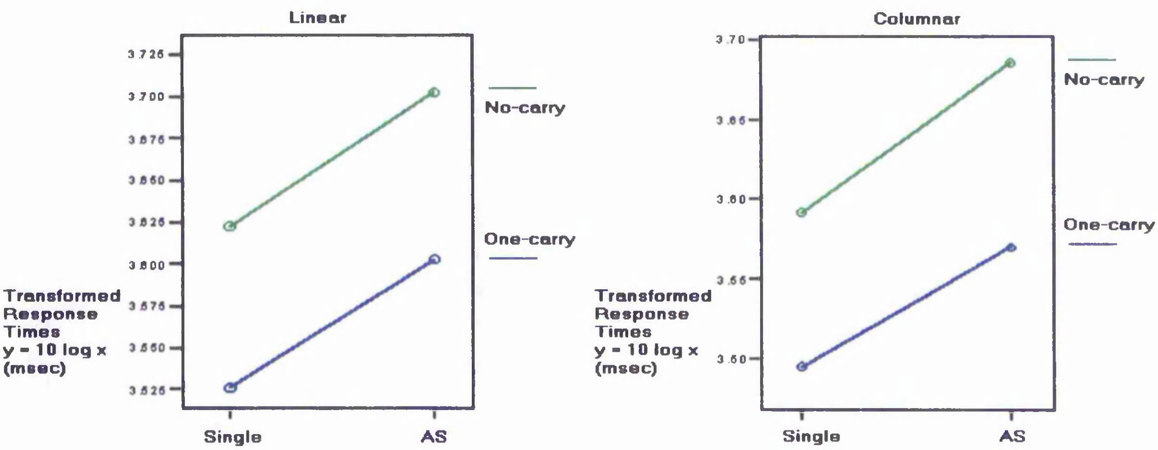

Figure 27. Addition experiment Single vs. AS task x carry x format interaction (transformed response times)

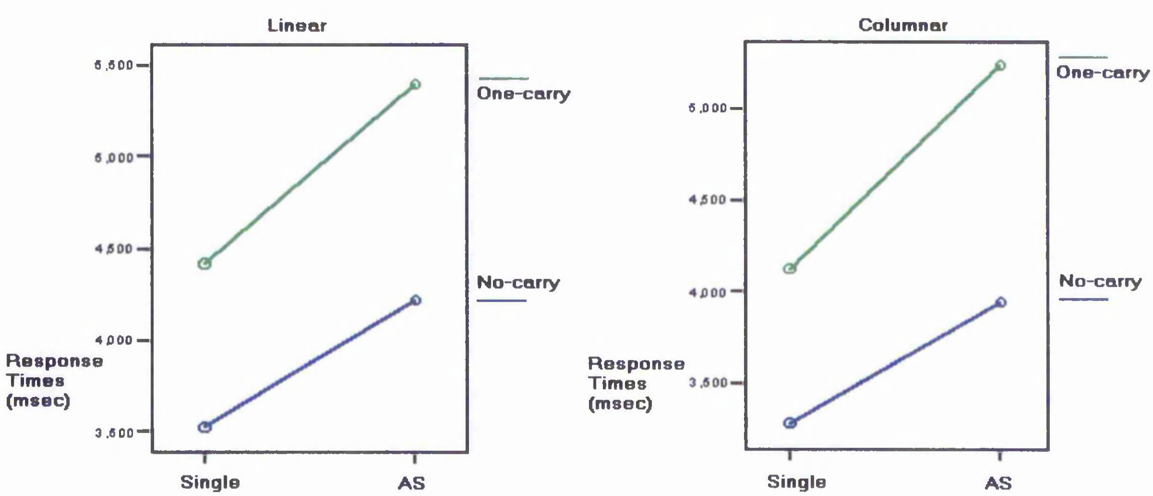

Figure 28. Addition experiment Single vs. AS task x carry x format interaction (untransformed response times)

The interaction is consistent across graphs of transformed and untransformed data, and is therefore unlikely to be a result of the log transformation alone. Obviously, this is not consistent with the tentative suggestion, from the first series of experiments, that phonological coding might be used in the transcoding of information from linear into columnar format.

The AS error scores for the first and second experiments show a different pattern of main effects in the second addition experiment, with the task and format variables achieving significance in addition to the carry variable, so that error scores increased under AS, with the carry operation, and with linear format. This is not consistent with the earlier suggestion that the visual format of presentation is important in the time taken for the extraction or transcoding of problem information, but not in terms of the processes underlying calculation accuracy, although the same lack of interactions in the response time data again provide no evidence for phonological coding in the carry operation.

The RLG response time results are the same in the first and second experiments, significant main effects of task, carry and format, but no interactions, providing no evidence that RLG disrupts either the carry operation or format of presentation, although the RLG error data does provide evidence that RLG disrupts both the carry operation and the format of presentation. The significant task x carry interaction shows that more errors occur in carry problems under RLG, and this is modified by the task x carry x format interaction, such that in single-task conditions no-carry linear problems have greater errors than no-carry columnar problems, while one-carry problems in linear and columnar format have similar errors. Figure 29 illustrates this three-way interaction, showing that RLG has a differential effect on the carry operation compared to single-task conditions, thereby implicating central executive involvement in the carry operation.

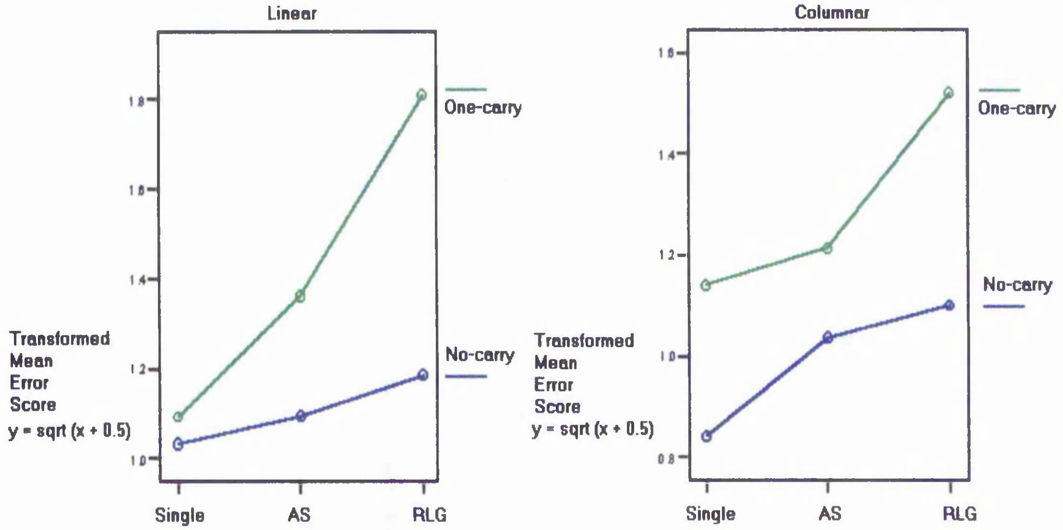

Figure 29. The task x carry x format interaction – addition experiment transformed error scores.

These graphs both demonstrate the required pattern of task x carry interaction, as identified earlier for the first RLG experiment (see figure 15), although clearly, linear format gives a greater effect. Once again, the suggestion is that response times and error rates are measuring different aspects of cognitive load, and that error scores give a better indication of WM involvement.

Baseline versus dual-task AS articulation rate was not significantly different, (t = -1.022, n.s.). As in the first experiments, this failed to meet the criterion for dual-task interference, suggesting perhaps that the articulatory suppression task did not provide a strong load on those components of WM used in the arithmetic task.

The comparison of randomness in the RLG baseline versus dual-task condition for the second experiments relied upon a simpler procedure than the randomness index used in the first experiments, Evans (1978), and used the Chi-square *goodness-of-fit* test to compare sequences of A, B, C, D, E articulations, the idea being that a random articulation of one-hundred of these letters should give an expectancy of twenty occurrences for each letter, and that the value of Chi-square would increase as observations departed from these expectations. This procedure gives a *pseudo-randomness* index rather than a true randomness index, because the repetitive articulation of twenty A, B, C, D, E sequences would not be significantly different to any other sequence of each of twenty A's, B's etc. However, a check of articulations was made to ensure that participants had not simply repeated each five-letter sequence, and here, all that is needed is an indication that the dual-task RLG articulations are less random than RLG articulations as a single-task, the actual levels of randomness are irrelevant. RLG baseline versus dual-task scores were significantly different, ($t = -10.729$, $p < 0.001$), thereby meeting the criterion for acceptance of resource competition between primary and secondary tasks. Correlations of response times, for each of the twelve addition categories, with associated error scores provided only one significant result, RLG zero-carry problem response times correlated negatively with error scores, ($r = -0.606$, $p < 0.01$), but given the lack of supporting correlations, and the multiple use of correlations, this cannot be taken as evidence of a speed / accuracy trade-off.

5.3.4 Item analysis of addition experiment problems

Item analysis for the addition problems involved comparison of the mean response time scores for each of the twenty addition item problems in each item category, averaged across the twenty participants, with incorrect items, and those with response time scores greater than two standard deviations from the mean, removed from the analysis. As can be seen from table 14, none of the twelve problem categories contained individual problems that gave response times significantly different to the other problems within the category, although naturally, there was some variation in response times, as might be expected with different addends and totals. This is evidence not only that the individual items within categories are consistent, but also that *problem-size effect* (see Ashcraft (1992, 1995) above) does not result in radically different response times for this limited range (totals 157 to 199) of multi-digit addition problems.

Level	F Ratio	Probability
Single Zero-Carry Linear	F (19, 357) = 0.86	p = 0.638 n.s.
Single Zero-Carry Columnar	F (19, 366) = 1.40	p = 0.121 n.s.
Single One-Carry Linear	F (19, 359) = 0.95	p = 0.527 n.s.
Single One-Carry Columnar	F (19, 349) = 1.44	p = 0.105 n.s.
AS Zero-Carry Linear	F (19, 350) = 0.70	p = 0.817 n.s.
AS Zero-Carry Columnar	F (19, 346) = 0.86	p = 0.633 n.s.
AS One-Carry Linear	F (19, 346) = 0.83	p = 0.665 n.s.
AS One-Carry Columnar	F (19, 350) = 0.76	p = 0.750 n.s.
RLG Zero-Carry Linear	F (19, 360) = 0.48	p = 0.970 n.s.
RLG Zero Carry Columnar	F (19, 366) = 0.63	p = 0.884 n.s.
RLG One-Carry Linear	F (19, 334) = 0.75	p = 0.762 n.s.
RLG One-Carry Columnar	F (19, 357) = 0.63	p = 0.885 n.s.

Table 14. Item analysis of within-category correct addition problem response times

5.3.5 Strategy use in the addition experiment

All participants were sent a detailed strategy sheet by e-mail in advance of the date of the experiments, and were also given ample time to familiarise themselves with this strategy sheet prior to full participation in the experiments. The strategy sheet contained seven different strategy categories based on the strategy types discussed in section 2.4 above, and these are briefly described below, but see appendix G for further details of addition and subtraction strategies.

The seven strategies were:

A) Counting on / Counting down – of hundreds, tens or units.

B) Overshooting / Undershooting and correcting.

C) Partitioning.

D) Make nearest hundred or nearest whole number.

E) Mental carrying – borrowing / paying back.

F) Retrieval without calculation.

G) Any other strategy, or a mixture of the strategies above.

The strategy sheets were directly available to participants throughout the experiments, and participants were required to state which of the strategy categories they had used immediately after giving each problem answer, and these strategy choices were recorded for later analysis. There were two separate analyses, the first to identify the percentage use of each strategy across the twelve levels of task, carry and format, and the second to identify any effect that strategy use might have on response times. The purpose of analysing percentage strategy use across task, carry and format levels was to determine if strategy use changed when participants faced more difficult problem conditions, e.g. did they use a different strategy for a single-task, zero-carry columnar problem than they did for an RLG one-carry linear problem?

If so, and if the second analysis indicated that these different strategies had different response times associated with them, then this had the potential to undermine the study, for the effect of the different task, carry and format levels would be confounded with the effect of strategy use. The strategies for both correct and incorrectly answered problems were analysed together to give percentage usage, as strategy choice was the important factor, but the strategy and response time analysis used only correct responses.

Table 15 gives the mean percentage strategy use for each of the twelve task, carry and format levels in the addition experiment, and the number of participants using each strategy within each task, carry and format level.

Strategy	A Counting	B Overshoot	C Partition	E Carry	F Retrieve	G Other
Single zero-carry linear	2.75	4	55	30	5.5	2.75
Number choosing strategy	1	3	17	10	3	1
Single zero-carry columnar	1.75	5	45	39.75	7	1.5
Number choosing strategy	1	4	15	13	3	1
Single one-carry linear	5.5	7.75	53	30.25	2.5	1
Number choosing strategy	3	5	15	11	3	2
Single one-carry columnar	1.5	7.75	45.5	41	3.5	0.75
Number choosing strategy	1	4	14	13	4	1
AS zero-carry linear	3.25	2.75	54.25	27.5	8.25	4
Number choosing strategy	1	3	16	9	4	2
AS zero-carry columnar	1.25	3.75	50.45	34.3	8.25	2
Number choosing strategy	1	5	14	12	5	1
AS one-carry linear	4.25	6.25	53.75	31	2.75	2
Number choosing strategy	2	4	16	10	2	2
AS one-carry columnar	0.5	5	47.75	39.75	3.25	3.75
Number choosing strategy	1	4	14	13	3	2
RLG zero-carry linear	2.75	1.75	53.25	30.5	7.5	4.25
Number choosing strategy	1	2	15	11	5	2
RLG zero-carry columnar	1.5	2	40.75	44.75	10	1
Number choosing strategy	1	3	12	14	4	2
RLG one-carry linear	4.5	4.5	56	30.75	1.25	3
Number choosing strategy	2	4	14	11	2	2
RLG one-carry columnar	2	4.75	40.75	48	3	1.5
Number choosing strategy	1	4	11	14	3	2

Table 15. Mean percentage strategy use and number of participants choosing strategy at each level of additions

Strategy D (make nearest hundred or nearest whole number) was not used by any of the participants, and was not included in table 15. Strategies A, B, F and G were used by either one single participant, or by very few participants. Strategy use proved to be strongly idiosyncratic; most participants made use of one predominant strategy and limited use of one other strategy, while some used one exclusive strategy throughout all levels of the experiment. Only one participant used more than three of the strategies.

The only strategies that could be meaningfully analysed were strategies C and E, and the percentage scores for those participants demonstrating some use of these two strategy groups (C, N = 17 and E, N = 15) were transformed using y = sqrt (x + 0.5), and subjected to separate 3 x 2 x 2 within-subjects ANOVA to reflect the task, carry and format levels.

The analysis of strategy C provided no significant main effects or interactions, with the same results for strategy E. This demonstrates that while certain individuals may indeed alter their strategy choice with different task, carry and format levels, most individuals, and the group as a whole, did not differ in their choice of strategy. This is important because it shows that here strategy choice did not tend to alter with WM load; participants largely used the same strategy for Single, AS and RLG problems, irrespective of format of presentation and number of carries.

This does not undermine the logic of the use of dual-tasks and the recording of response times. However, this might not be the case for every sample of participants, and it is important to know whether strategy choice relates to response times; are some strategies faster than others? Response times for correct items were compared using one-way ANOVA for all of the strategies used at each level of task, carry and format, with Games-Howell post-hoc multiple comparison tests used to identify the relationship of these strategies, although strategies C and E are the only ones reported due to the small number of participants choosing the other strategies, and the results are given in table 16.

Level	F Ratio	Probability	C vs. E
Single Zero-Carry Linear	$F\,(5, 371) = 5.10$	$p < 0.001$	C > E, $p < 0.01$
Single Zero-Carry Columnar	$F\,(5, 380) = 23.63$	$p < 0.001$	C > E, $p < 0.001$
Single One-Carry Linear	$F\,(5, 373) = 9.16$	$p < 0.001$	C > E, $p < 0.001$
Single One-Carry Columnar	$F\,(5, 363) = 19.80$	$p < 0.001$	C > E, $p < 0.001$
AS Zero-Carry Linear	$F\,(4, 366) = 9.04$	$p < 0.001$	C > E, $p = 0.05$
AS Zero-Carry Columnar	$F\,(5, 365) = 28.07$	$p < 0.001$	C > E, $p < 0.001$
AS One-Carry Linear	$F\,(5, 341) = 5.43$	$p < 0.001$	C > E, $p < 0.001$
AS One-Carry Columnar	$F\,(5, 345) = 16.20$	$p < 0.001$	C > E, $p < 0.001$
RLG Zero-Carry Linear	$F\,(5, 374) = 12.11$	$p < 0.001$	C > E, $p < 0.001$
RLG Zero Carry Columnar	$F\,(5, 380) = 18.43$	$p < 0.001$	C > E, $p < 0.001$
RLG One-Carry Linear	$F\,(5, 329) = 5.89$	$p < 0.001$	C > E, $p < 0.001$
RLG One-Carry Columnar	$F\,(4, 339) = 16.54$	$p < 0.001$	C > E, $p < 0.001$

Table 16. Analysis of response times and strategy use with post-hoc comparisons of strategies C and E

This appears to indicate that strategy E is consistently faster than strategy C, but the situation is complicated by the results from the previous analysis of percentage strategy use. For percentage strategy use it is clear that many participants demonstrated exclusive, or near exclusive, use of either strategy C or strategy E. There is simply insufficient data to compare the use of strategy C and strategy E for each participant, and consequently, it is not possible to disambiguate strategy effects from the potential effects of individual differences in speed of processing.

It could be that those participants using strategy E were simply faster at doing arithmetic than those using strategy C. Interestingly, of the two participants who used strategy E (mental carrying / borrowing and paying back) exclusively, one was a book-keeper, who demonstrated the fastest response times of all of the participants, and the other was an undergraduate student from a Chinese family, also demonstrating similarly fast response times. When asked, during debriefing, why they used only this strategy, both participants indicated that prior education or occupational experience had proved the success of this strategy, and consequently they no longer used other strategies, preferring to develop the performance of this single strategy.

5.4 The subtraction experiment

The triple-code model, Cohen and Dehaene (2000), claims that addition facts are normally retrieved as rote learned verbal word frames, whereas subtractions, because they have no store of rote learned facts, have to be solved by *'strategical quantity manipulations'*. In contrast, the abstract-modular model of McCloskey, Caramazza and Basili (1985) makes no distinction between the arithmetic operations in terms of different processing routes. The comparison of results for this subtraction experiment, with those from the addition experiment, should provide an indication of support for one or the other of these cognitive models of arithmetic processing. If this subtraction experiment demonstrates different patterns of WM activation to those activated during addition, then this would at least suggest some support for the triple-code model, whereas similar patterns of activation would suggest support for the abstract-modular model. However, this is complicated by the fact that different patterns of WM activation could possibly be attributed to differences in encoding problem information rather than in the calculation process itself, and further complicated by the possibility that response times might be more closely associated with encoding, while error rates might be more closely associated with the calculation process. Experimental procedure was identical to the addition experiment, with strategy recorded for every problem.

5.4.1 Subtraction experiment – correct answer response times

Table 17 gives the untransformed mean response times for each task x carry x format level.

Level	Mean	Std. Deviation
Single Zero-Carry Linear	4033	1268
Single Zero Carry Columnar	3791	1108
Single One-Carry Linear	5463	1685
Single One-Carry Columnar	5188	1658
AS Zero-Carry Linear	4804	1714
AS Zero Carry Columnar	4473	1539
AS One-Carry Linear	6512	2580
AS One-Carry Columnar	6229	2521
RLG Zero-Carry Linear	6342	2749
RLG Zero Carry Columnar	5800	2500
RLG One-Carry Linear	7760	3129
RLG One-Carry Columnar	7486	2956

Table 17. Correct item mean response times and standard deviations in milliseconds (subtraction experiment)

A 3 x 2 x 2 repeated-measures ANOVA demonstrated a significant difference between single-task ($M = 4355$ msec), AS ($M = 5093$ msec) and RLG ($M = 6324$ msec), $F_{(2, 38)} = 43.19$, $p < 0.001$. A significant difference between no-carry ($M = 4519$ msec) and one-carry ($M = 5984$ msec), $F_{(1, 19)} = 67.96$, $p < 0.001$. A significant difference between linear format ($M = 5358$ msec) and columnar format ($M = 5047$ msec), $F_{(1, 19)} = 26.80, p < 0.001$. A significant task x carry interaction, $F_{(2, 38)} = 5.47, p < 0.05$, and a marginally significant task x carry x format interaction, $F_{(2, 38)} = 3.33, p < 0.06$.

Further analysis, using within-subjects contrasts and separate 2 x 2 x 2 ANOVA of paired task levels, gave the following Single vs. AS results; a significant effect of task, $F_{(1,19)} = 37.58, p < 0.001$, a significant effect of carry, $F_{(1, 19)} = 70.95, p < 0.001$, and a significant effect of format, $F_{(1, 19)} = 26.16, p < 0.001$. None of the interactions achieved significance.

The Single vs. RLG results demonstrated a significant effect of task, $F_{(1, 19)} = 53.27, p < 0.001$, a significant effect of carry, $F_{(1, 19)} = 67.81, p < 0.001$, a significant effect of format, $F_{(1, 19)} = 23.77, p < 0.001$, a significant task x carry interaction, $F_{(1, 19)} = 5.18, p < 0.05$, such that the difference in response times for zero and one-carry problems was less under RLG, and a significant task x carry x format interaction, $F_{(1, 19)} = 5.12, p < 0.05$, such that under RLG columnar format gave a response time advantage for zero-carry problems, but not for one-carry problems.

The AS vs. RLG comparison provided a significant effect of task, $F_{(1, 19)} = 29.14, p < 0.001$, a significant effect of carry, F (1, 19) $= 58.34, p < 0.001$, a significant effect of format, $F_{(1, 19)} = 25.79, p < 0.001$, and a significant task x carry interaction, $F_{(1, 19)} = 8.32, p < 0.01$, again demonstrating that under RLG the difference in response times between zero-carry and one-carry problems was less than under single-task conditions.

5.4.2 Subtraction experiment – error scores

The untransformed mean error scores for each task x carry x format level are given in table 18.

Level	Mean	Std. Deviation
Single Zero-Carry Linear	.60	.68
Single Zero Carry Columnar	.15	.37
Single One-Carry Linear	1.45	1.32
Single One-Carry Columnar	1.15	1.14
AS Zero-Carry Linear	1.40	1.10
AS Zero Carry Columnar	.60	.99
AS One-Carry Linear	1.95	1.19
AS One-Carry Columnar	1.80	1.51
RLG Zero-Carry Linear	.95	1.19
RLG Zero Carry Columnar	.80	.77
RLG One-Carry Linear	2.35	1.39
RLG One-Carry Columnar	2.20	1.64

Table 18. Mean error scores and standard deviations (subtraction experiment)

A 3 x 2 x 2 repeated-measure ANOVA of transformed error scores demonstrated a significant difference between single-task (M = 1.080), AS (M = 1.313) and RLG (M = 1.359), $F_{(2, 38)} = 11.08, p < 0.001$. A significant difference between no-carry (M = 1.051) and one-carry (M = 1.451), $F_{(1, 19)} = 43.00, p < 0.001$. A significant difference between linear format (M = 1.320) and columnar format (M = 1.181), $F_{(1,19)} = 10.03, p < 0.01$.

Tests of within-subjects contrasts and further 2 x 2 x 2 ANOVA gave the following comparisons. Single vs. AS; a significant effect of task, $F_{(1, 19)} = 12.52, p < 0.01$, a significant effect of carry, $F_{(1, 19)} = 32.23, p < 0.001$, a significant effect of format, $F_{(1, 19)} = 10.64, p < 0.01$. Single vs. RLG; a significant effect of task, $F_{(1, 19)} = 23.49, p < 0.001$, a significant effect of carry, $F_{(1, 19)} = 41.56, p < 0.001$, and a significant effect of format, $F_{(1, 19)} = 4.45, p < 0.05$. AS vs. RLG; a significant effect of carry, $F_{(1, 19)} = 34.84, p < 0.001$, and a significant effect of format, $F_{(1, 19)} = 5.69, p = 0.028$.

5.4.3 Subtraction experiment – interpretation and discussion of results

Figure 30 provides a visual summary of the subtraction experiment results, and a visual comparison with the addition experiment results.

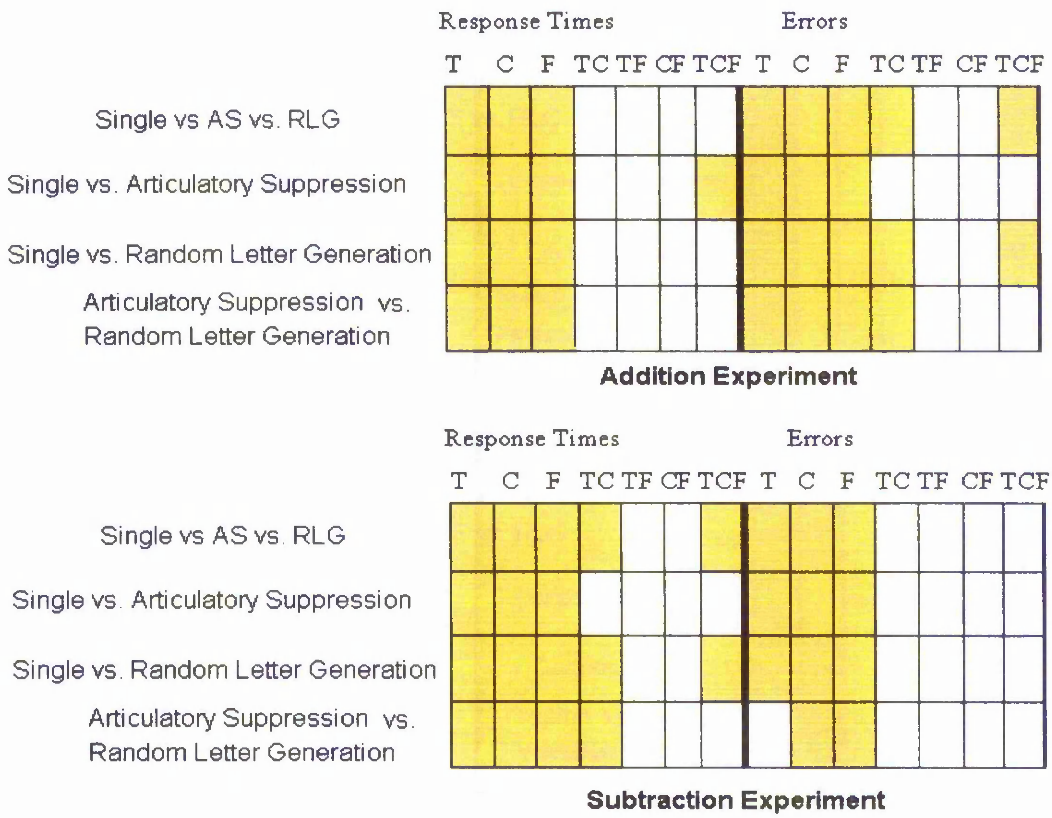

Figure 30. Comparison of results for addition and subtraction experiments

The results for subtraction seem to be almost reversed, in terms of response times and errors, when compared to the addition experiment results. Articulatory suppression provides no significant interactions for response times or errors, and its main effects are assumed to reflect cost of concurrence rather than dual-task interference.

The Single vs. RLG, response time, task x carry x format interaction can be understood as a carry x format interaction solely under RLG, where the advantage of columnar format is greater under zero-carry conditions than for one-carry conditions. Figures 31 and 32 provide a visual representation of this three-way interaction.

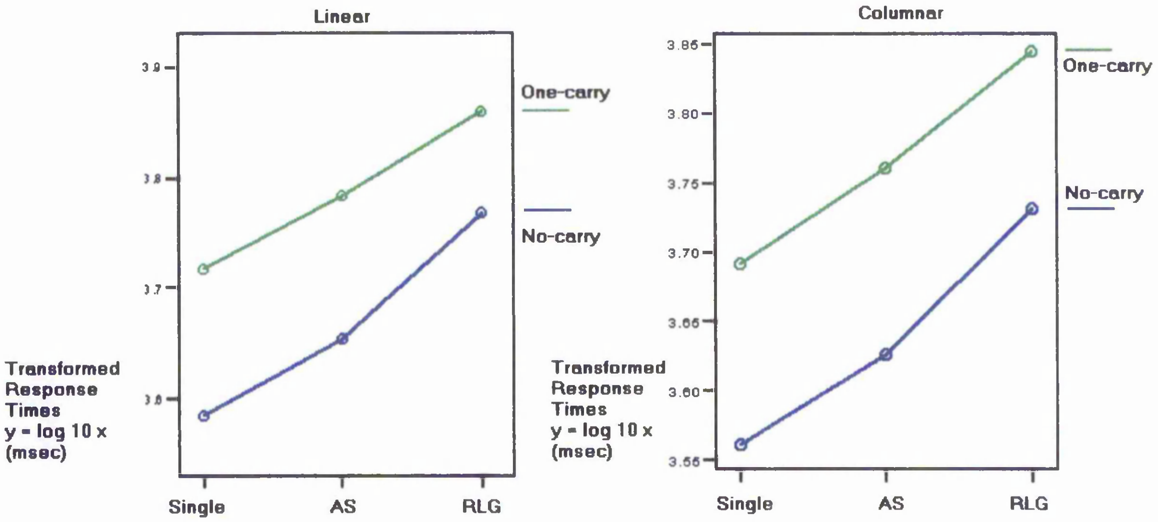

Figure 31. The task x carry x format interaction – subtraction experiment transformed response times

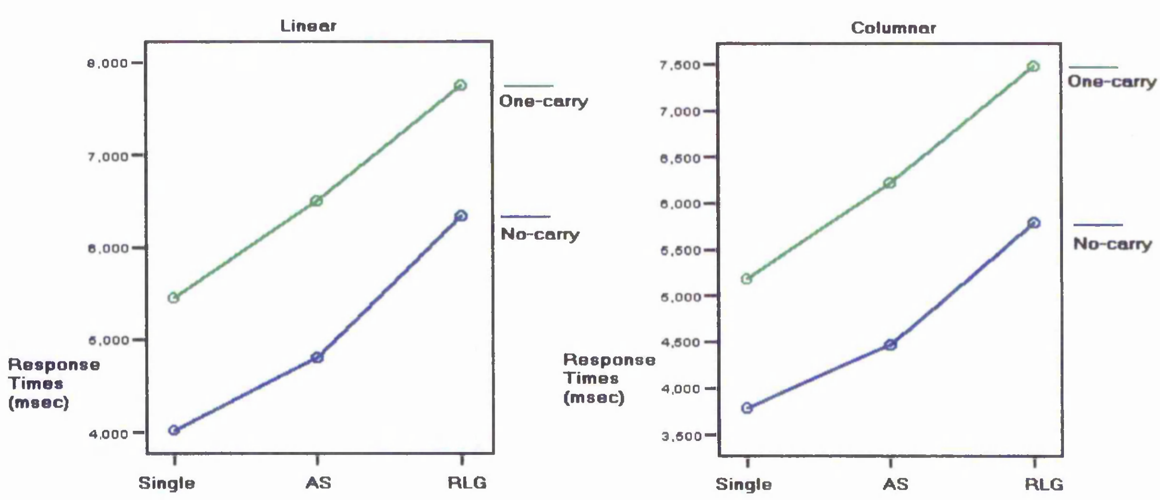

Figure 32. The task x carry x format interaction – subtraction experiment untransformed response times

Alternatively, the interaction in figures 31 and 32 can be thought of as a task x format interaction at each of the carry levels, columnar format providing an even greater advantage under RLG than under single-task conditions, but only for zero-carry problems. This is consistent with RLG operating during encoding, or as a cost of concurrence effect, rather than in relation to the carry operation. As the interaction is evident in the graphs of both transformed and untransformed data, there is no reason to suppose that this interaction stems from the log transformation of scores.

The lack of interactions in the error data also provides no evidence for RLG disruption to the carry operation, and no evidence of central executive involvement in the solution of these subtraction problems. The AS vs. RLG task x carry interaction shows that the difference between zero-carry and one-carry problem response times is less under RLG than under AS, and this again provides no support for RLG as a disruptor of the carry operation.

The error data for AS vs. RLG has no interactions, but is also notable for the lack of a main effect of task, which means that error rates under AS and RLG were not significantly different. Given that the same twenty participants completed the subtraction experiment and the addition experiment, in the same conditions using the same apparatus, it is reasonable to suppose that the difference between the addition experiment and subtraction experiment results are primarily due to the differences between the addition and subtraction problems, and by implication, between the addition and subtraction operations. The differences seem to indicate that RLG affects the carry operation in addition problems but not in subtraction problems, and insofar as RLG is a measure of one aspect, or aspects, of central executive involvement, that the central executive is involved in addition but not subtraction.

This would appear to be consistent with the prediction from the triple-code model that additions and subtractions are processed differently, but the triple-code model explains this as due to additions being normally retrieved as rote learned verbal word frames, whereas subtractions have no store of rote learned facts, relying on magnitude comparisons for their solution. It is not clear how the RLG results, and the implication of central executive involvement, can be understood in relation to the triple-code model unless access to the store of rote learned arithmetic facts requires central executive involvement, so that additions require the CE, but subtractions do not.

The argument postulated above is complicated by the response time results, for if response times are more strongly associated with encoding time than the calculation process itself, then subtractions seemed to demonstrate a different pattern of response time results to those found for additions, and given the absence of rote learned subtraction facts, this might reflect some sort of recoding between operations, subtractions perhaps being recoded as additions, e.g. 192-57 recoded as 57 +? = 192.

Articulatory suppression baseline articulation rate was not significantly different from dual-task articulation rate, again suggesting either that the articulatory suppression task did not load the phonological component of WM to any extent, or that phonological resources were not required by the subtraction task. However, as with the addition experiment, the baseline measurement of randomness in RLG as a single-task was significantly different than that for RLG as a dual-task, (t = - 8.397, $p < 0.001$). Multiple correlations of each of the twelve category mean response times with associated mean error scores produced only one significant positive correlation for one-carry linear problems under RLG, ($r = 0.679$, $p < 0.001$), which provides no evidence of speed / accuracy trade-off.

5.4.4 Item analysis of subtraction experiment problems

As for the addition experiment, item analysis for the subtraction problems also involved comparison of the mean response time scores for each of the twenty subtraction item problems in each item category, averaged across the twenty participants, with incorrect items, and those with response time scores greater than two standard deviations from the mean, removed from the analysis. Table 19 gives the results of one-way ANOVA on all twelve of the subtraction item categories.

Level	F Ratio	Probability
Single Zero-Carry Linear	$F\,(19, 347) = 0.52$	$p = 0.955$ n.s.
Single Zero-Carry Columnar	$F\,(19,354) = 0.84$	$p = 0.658$ n.s.
Single One-Carry Linear	$F\,(19, 329) = 1.36$	$p = 0.142$ n.s.
Single One-Carry Columnar	$F\,(19, 332) = 0.99$	$p = 0.476$ n.s.
AS Zero-Carry Linear	$F\,(19, 346) = 0.65$	$p = 0.868$ n.s.
AS Zero-Carry Columnar	$F\,(19, 364) = 0.48$	$p = 0.970$ n.s.
AS One-Carry Linear	$F\,(19, 332) = 0.91$	$p = 0.569$ n.s.
AS One-Carry Columnar	$F\,(19, 333) = 0.47$	$p = 0.973$ n.s.
RLG Zero-Carry Linear	$F\,(19, 358) = 0.45$	$p = 0.980$ n.s.
RLG Zero Carry Columnar	$F\,(19, 356) = 0.46$	$p = 0.977$ n.s.
RLG One-Carry Linear	$F\,(19, 332) = 0.34$	$p = 0.996$ n.s.
RLG One-Carry Columnar	$F\,(19, 332) = 0.59$	$p = 0.911$ n.s.

Table 19. Item analysis of within-category correct subtraction problem response times

As with the addition problem items, none of the twelve problem categories contained individual problems that gave response times significantly different to the other problems within the category. This demonstrates that the individual items within categories are consistent, and that *problem-size effect* does not result in different response times for this limited range of multi-digit subtraction problems.

5.4.5 Strategy use in the subtraction experiment

The same strategy sheet was used in the subtraction experiment as in the addition experiment, and as previously, participants were given ample time to familiarise themselves with this strategy sheet prior to participation in the experiments. The strategy sheets continued to be directly available to participants throughout the experiments. Participants were again required to state which of the strategy categories they had used immediately after giving each problem answer, and these were recorded for the two separate analyses described earlier, percentage strategy use, and response time variation with strategy. Table 20 gives the mean percentage strategy use for each of the twelve task, carry and format levels in the subtraction experiment, and the number of participants using each strategy within each task, carry and format level.

Again, strategy D (make nearest hundred or nearest whole number) was not used by any of the participants, and has been removed from table 20. Strategies A, F and G had small percentage use due to being chosen by either one single participant, or by very few participants.

Strategy	**A Counting**	**B Overshoot**	**C Partition**	**E Carry**	**F Retrieve**	**G Other**
Single zero-carry linear	5.5	3.25	47.5	36.5	3	4.25
Number choosing strategy	2	4	16	13	1	1
Single zero-carry columnar	1.75	4	32.75	53.5	3	5
Number choosing strategy	2	3	12	15	1	1
Single one-carry linear	7.5	18.3	39.5	32.5	1	1.3
Number choosing strategy	2	5	15	12	2	3
Single one-carry columnar	3.25	19	35.75	40.75	0.25	1
Number choosing strategy	3	6	10	16	5	1
AS zero-carry linear	4.5	3.75	46.25	37.4	3.85	4.25
Number choosing strategy	2	2	13	12	1	1
AS zero-carry columnar	1.75	7	31	51.25	4.25	4.8
Number choosing strategy	1	3	12	14	2	1
AS one-carry linear	5.75	21.1	39	32.6	0.8	0.8
Number choosing strategy	2	6	14	11	3	2
AS one-carry columnar	3.75	19.9	41.5	33.85	0.3	0.75
Number choosing strategy	2	7	13	10	1	2
RLG zero-carry linear	3.75	4.55	44.25	39.4	3.85	4.25
Number choosing strategy	2	4	12	13	2	1
RLG zero-carry columnar	2.75	4.75	35.25	49.25	4	4
Number choosing strategy	1	3	10	14	2	1
RLG one-carry linear	5	20.6	35.25	38.1	0.55	0.5
Number choosing strategy	2	6	13	13	2	2
RLG one-carry columnar	5.25	19	30	44.75	0.75	0.25
Number choosing strategy	2	6	11	15	2	1

Table 20. Mean percentage strategy use and number choosing each strategy at each level of subtractions

Strategies B, C and E were considered to offer sufficient numbers of participants for analysis, and the percentage scores for those participants demonstrating some use of these three strategy groups (B, N = 9, C, N = 16 and E, N = 17) were transformed using y = sqrt (x + 0.5), and subjected to separate 3 x 2 x 2 within-subjects ANOVA to reflect the task, carry and format levels.

Analysis of strategy B demonstrated a significant effect of carry, $F_{(1, 8)} = 8.20$, $p < 0.05$, due to greater percentage use of strategy B with one-carry problems. Further analysis using within-subjects contrasts revealed a significant task x carry interaction at the Single vs. AS level, such that greater percentage use of strategy B occurred under articulatory suppression with one-carry problems, $F_{(1, 8)} = 7.96$, $p < 0.05$. Strategy C demonstrated no significant difference in percentage usage across the different task x carry x format levels. Analysis of strategy E revealed a significant effect of format, $F_{(1,16)} = 4.50$, $p < 0.05$, such that the percentage use of strategy E was greater for problems in columnar format. Further analysis of within-subjects contrasts for the Single vs. AS comparison demonstrated a significant effect of task, $F_{(1, 16)} = 5.38$, $p \leq 0.05$, which was also evident for the AS vs. RLG comparison, $F_{(1, 16)} = 4.71$, $p < 0.05$, both of which are explained by the lower percentage use of strategy E under articulatory suppression.

These results are different to the addition results, in that here, strategy use does vary with task x carry x format levels, and this means that if the different strategies are associated with different response times, then this could constitute a confounding variable, potentially undermining the logic of the experiment to some extent. As in the addition experiment, response times for correct items were compared using one-way ANOVA for all of the strategies used at each level of task, carry and format, with Games-Howell post-hoc multiple comparison tests used to identify the relationship of strategies B, C and E. Results are given in table 21.

Level	F Ratio	Probability	B vs. C vs. E
Single Zero-Carry Linear	$F\,(5, 361) = 4.85$	$p < 0.001$	B = C = E, n.s.
Single Zero-Carry Columnar	$F\,(5, 368) = 13.68$	$p < 0.001$	B = C, C > E, $p < 0.001$
Single One-Carry Linear	$F\,(5, 343) = 10.34$	$p < 0.001$	B = C, C > E, $p < 0.001$
Single One-Carry Columnar	$F\,(4, 346) = 19.80$	$p < 0.001$	B = C, C > E, $p < 0.001$
AS Zero-Carry Linear	$F\,(5, 360) = 8.43$	$p < 0.001$	B > C, C = E, $p < 0.001$
AS Zero-Carry Columnar	$F\,(5, 378) = 18.33$	$p < 0.001$	B = C, C > E, $p < 0.001$
AS One-Carry Linear	$F\,(4, 347) = 7.65$	$p < 0.001$	B = C, C > E, $p < 0.001$
AS One-Carry Columnar	$F\,(4, 348) = 13.87$	$p < 0.001$	B = C, C > E, $p < 0.001$
RLG Zero-Carry Linear	$F\,(5, 372) = 4.34$	$p = 0.001$	B > E, C = E, $p < 0.05$
RLG Zero Carry Columnar	$F\,(5, 370) = 13.03$	$p < 0.001$	B = C, C > E, $p < 0.001$
RLG One-Carry Linear	$F\,(4, 347) = 4.53$	$p = 0.001$	B = C, C > E, $p < 0.05$
RLG One-Carry Columnar	$F\,(4, 347) = 6.03$	$p < 0.001$	B = C, C > E, $p < 0.001$

Table 21. Analysis of response times and strategy use with post-hoc comparisons of strategies B, C and E

Percentage use of strategy B was greater with one-carry problems under articulatory suppression, but as can be seen in table 21, those switching from strategy C to strategy B would not have gained any advantage or disadvantage in response time, for strategy B provides essentially similar response times to strategy C across eleven of the twelve task, carry and format levels. Those switching from strategy E to strategy B could have increased their response times, but what effect might this actually have? A simple, but relatively inaccurate, estimate of the effect can be obtained by considering two of the AS levels in more detail, the AS zero-carry columnar (ASZCC) and the AS one-carry columnar (ASOCC) levels, as in table 22.

Task Carry Format Level	Overall Mean response time	Strategy B response time	Strategy E response time	B – E difference	Number of problems strategy E
ASZCC	4473	6508	3875	2633	187
ASOCC	6229	7411	4882	2529	129
Difference	+ 1756				58

Table 22. Response time differences (milliseconds) strategy B vs. strategy E, at two AS levels (subtractions)

Fifty-eight fewer problems were solved using strategy E at the ASOCC level, and if it is assumed that all of these represent a change to strategy B, then the overall effect on mean response time can be calculated by multiplying this number by the difference in ASOCC strategy response times (B – E), giving 58 x 2529 = 146682 milliseconds. Dividing this number by the total number of ASOCC problems correctly answered by the twenty participants (352), gives 146682 / 352 = 417 milliseconds, which represents the maximum increase in ASOCC mean response time as a result of strategy change from E to B. This amounts to an approximate increase of 7%. The effect of percentage change in the use of strategy E can also be broadly, if inaccurately, estimated by the use of simplifying assumptions. The percentage change in strategy E use between linear and columnar format problems is 42% to 53%, and as there were nominally 120 problems in linear format and 120 problems in columnar format, these percentages represent 50 and 64 problems respectively, with the difference being 14 problems. Assuming that strategy E is on average some 2 seconds faster than strategy B or strategy C, then the overall contribution of strategy E to differences in format is 14 x 2 = 28 seconds, but this has to be divided across the 120 problems, giving 28/120 = 233 milliseconds per problem, which is potentially large enough to explain the advantage of columnar format over linear format, although, as with the addition experiment, this is complicated by individual differences in strategy use, for the effects of switching strategies applied only to some, and not all of the participants. Eight participants either made no use of strategy E, or made no use of strategy C, and again, the effect of strategy cannot be disambiguated from individual differences in speed of processing, as the faster response times for strategy E might not be due to the strategy itself, but due to these speed of processing differences, and any estimates of the effect of strategy necessarily include these faster average response times even if those participants who switched strategies did not demonstrate such faster times. Overall, there simply were not enough participants demonstrating switching responses to provide a sensible analysis.

5.5 The multiplication experiment

If the claims of the triple-code model are correct, then the pattern of interactions for multiplication should be different to those found in subtraction, and should perhaps be more similar to those found for addition, in that both addition and multiplication facts are considered to be stored as verbal word frames. Lee and Kang (2002) reported differences for single-digit subtractions and multiplications, but it is possible that single-digit arithmetic either fails to load WM, or loads WM in a different way to multi-digit arithmetic. The study by Seitz and Schumann-Hengsteler (2000) found that *easy* problems, e.g. 3 x 4, were not affected by any of their secondary tasks, but that *hard* problems, e.g. 8 x 17, were affected by both AS and RLG. To further investigate this issue, this multiplication experiment incorporated a difficulty factor into the design, utilising *easy*, three-digit x one-digit problems, and *hard*, two-digit x two-digit problems, as described in section 5.2 above, giving a Task (Single, AS, RLG) x Carry (zero, one) x Difficulty (easy, hard) x Format (linear, columnar) design.

5.5.1 Multiplication experiment – correct answer response times

The untransformed mean response times for each task x carry x format level are given in table 23. A 3 (task) x 2 (carry) x 2 (difficulty) x 2 (format) repeated-measures ANOVA of the log transformed response times demonstrated a significant difference between single-task (M = 6223 msec), AS (M = 7447 msec) and RLG (M = 8933 msec), $F_{(2, 38)} = 111.66$, $p < 0.001$. A significant difference between no-carry (M = 6138 msec) and one-carry (M = 9036 msec), $F_{(1, 19)} = 102.30$, $p < 0.001$. A significant difference between easy (M = 4864 msec) and hard ($M = 11402$ msec), $F_{(1, 19)} = 529.93$, $p < 0.001$. A significant difference between columnar format ($M = 7852$ msec) and linear format ($M = 7674$ msec), $F_{(1, 19)} = 38.28$, $p < 0.001$. A significant task x difficulty interaction, $F_{(2, 38)} = 8.50$, $p < 0.01$, a significant carry x difficulty interaction, $F_{(1, 19)} = 49.07$, $p < 0.001$, and a marginally significant carry x format interaction, $F_{(1, 19)} = 4.36$, $p < 0.06$.

LEVEL	Mean	Std. Deviation
Single no-carry easy columnar	3118	1027
Single no-carry easy linear	3329	1135
Single no-carry hard columnar	9037	3181
Single no-carry hard linear	9470	3902
Single one-carry easy columnar	5397	2108
Single one-carry easy linear	5496	2164
Single one-carry hard columnar	11340	3905
Single one-carry hard linear	12224	4413
AS no-carry easy columnar	3796	1317
AS no-carry easy linear	4184	1574
AS no-carry hard columnar	10070	3404
AS no-carry hard linear	11277	4174
AS one-carry easy columnar	6662	2712
AS one-carry easy linear	6724	2760
AS one-carry hard columnar	13411	4627
AS one-carry hard linear	14735	5533
RLG no-carry easy columnar	4642	1477
RLG no-carry easy linear	5113	1618
RLG no-carry hard columnar	11789	4076
RLG no-carry hard linear	12549	4582
RLG one-carry easy columnar	8467	3496
RLG one-carry easy linear	8737	3383
RLG one-carry hard columnar	15428	6046
RLG one-carry hard linear	16505	6235

Table 23. Correct item response times and standard deviations in milliseconds (multiplication experiment)

Further analysis using within-subjects contrasts and 2 x 2 x 2 x 2 ANOVA allowed the following comparisons: Single vs. AS, a significant effect of task, $F_{(1, 19)} = 133.28$, $p < 0.001$, a significant effect of carry, $F_{(1, 19)} = 134.94$, $p < 0.001$, a significant effect of difficulty, $F_{(1, 19)} = 525.63$, $p < 0.001$, and a significant effect of format, $F_{(1, 19)} = 22.42$, $p < 0.001$. A significant task x format interaction, $F_{(1, 19)} = 4.46$, $p < 0.05$, such that the advantage of columnar format was greater under AS than in single-task conditions, and a significant carry x difficulty interaction, $F_{(1, 19)} = 38.35, p < 0.001$, such that the carry operation appeared to give a relatively greater increase in response times for easy problems than for hard problems, but see the interpretation in section 5.5.3.

The Single vs. RLG comparison gave; a significant effect of task, $F_{(1, 19)} = 180.42$, $p < 0.001$, a significant effect of carry, $F_{(1, 19)} = 88.53, p < 0.001$, a significant effect of difficulty, $F_{(1, 19)} = 494.56, p < 0.001$, and a significant effect of format, $F_{(1, 19)} = 31.84, p < 0.001$. A significant task x difficulty interaction, $F_{(1, 19)} = 12.42, p < 0.002$, such that RLG appeared to give a relatively greater increase in response times for easy problems than it did for hard problems, and a significant carry x difficulty interaction, $F_{(1, 19)} = 49.42, p < 0.001$, also appearing to show, as in the Single vs. AS comparison, that the carry operation gave a relatively greater increase in response times for easy problems than it did for hard problems.

The AS vs. RLG comparison demonstrated a significant effect of task, $F_{(1, 19)} = 41.61, p < 0.001$, a significant effect of carry, $F_{(1, 19)} = 83.22, p < 0.001$, a significant effect of difficulty, $F_{(1, 19)} = 449.81, p < 0.001$, and a significant effect of format, $F_{(1, 19)} = 51.53, p < 0.001$. A significant task x difficulty interaction, $F_{(1, 19)} = 7.33, p < 0.05$, also appearing to show, as for the Single vs. RLG comparison, a relatively greater increase in response times for easy problems than for hard problems. A significant carry x difficulty interaction, $F_{(1, 19)} = 27.62, p < 0.001$, again appearing to show that the carry operation gave a relatively greater increase in response times for easy problems than for hard problems, but see the interpretation in section 5.5.3. A marginally significant carry x format interaction, $F_{(1, 19)} = 3.62, p < 0.07$, also appearing to demonstrate, as in the Single vs. AS comparison, a relatively greater advantage of columnar format for no-carry problems.

5.5.2 Multiplication experiment – error scores

As for previous error scores, the transformation y = sqrt (x + 0.5) was used to give greater adherence to the assumptions underlying ANOVA, but the untransformed scores are given in table 24.

LEVEL	Mean	Std. Deviation
Single no-carry easy columnar	.05	.22
Single no-carry easy linear	.30	.57
Single no-carry hard columnar	2.00	1.37
Single no-carry hard linear	1.20	.76
Single one-carry easy columnar	.55	.99
Single one-carry easy linear	.80	.76
Single one-carry hard columnar	2.35	1.49
Single one-carry hard linear	1.65	1.38
AS no-carry easy columnar	.20	.52
AS no-carry easy linear	.25	.63
AS no-carry hard columnar	2.15	1.59
AS no-carry hard linear	3.30	1.30
AS one-carry easy columnar	1.35	.93
AS one-carry easy linear	1.10	1.11
AS one-carry hard columnar	2.95	1.73
AS one-carry hard linear	3.20	1.93
RLG no-carry easy columnar	.75	.71
RLG no-carry easy linear	.35	.74
RLG no-carry hard columnar	2.15	1.46
RLG no-carry hard linear	2.55	1.43
RLG one-carry easy columnar	.95	.94
RLG one-carry easy linear	1.45	1.50
RLG one-carry hard columnar	2.90	1.61
RLG one-carry hard linear	3.30	1.86

Table 24. Mean error scores and standard deviations (multiplication experiment)

A 3 (task) x 2 (carry) x 2 (difficulty) x 2 (format) repeated-measures ANOVA demonstrated a significant difference between single-task (M = 1.178), AS (M = 1.405) and RLG (M = 1.414), $F_{(2, 38)} = 21.15$, $p < 0.001$. A significant difference between no-carry (M = 1.223) and one-carry (M = 1.442), $F_{(1, 19)} = 21.30$, $p < 0.001$. A significant difference between easy (M = 1.011) and hard (M = 1.654), $F_{(1, 19)} = 125.40$, $p < 0.001$. A marginally significant task x difficulty interaction, $F_{(2, 38)} = 3.26$, $p < 0.07$. A marginally significant task x carry x difficulty interaction, $F_{(2, 38)} = 3.34$, $p = 0.06$. A significant task x carry x format interaction, $F_{(2, 38)} = 3.80$, $p < 0.05$. A significant task x difficulty x format interaction, $F_{(2, 38)} = 9.80$, $p < 0.001$, and a significant carry x difficulty x format interaction, $F_{(1, 19)} = 8.75$, $p < 0.01$.

Further analysis using within-subjects contrasts and 2 x 2 x 2 x 2 ANOVA allowed comparisons of task levels as follows: Single vs. AS, a significant effect of task, $F_{(1,19)} = 23.45$, $p < 0.001$, a significant effect of carry, $F_{(1, 19)} = 21.88$, $p < 0.001$, and a significant effect of difficulty, $F_{(1, 19)} = 153.49$, $p < 0.001$.

Significant interactions included; task x difficulty, $F_{(1, 19)} = 11.07$, $p < 0.01$, such that under AS, hard problems had relatively greater errors than under single-task conditions, modified by a significant task x difficulty x format interaction, $F_{(1, 19)} = 24.68$, $p < 0.001$, such that under single-task conditions columnar format gave greater errors for hard problems, while under AS, linear format gave greater errors for hard problems.

A significant carry x difficulty interaction, $F_{(1, 19)} = 5.31$, $p < 0.05$, such that the carry operation led to relatively more errors for easy problems, modified by a significant task x carry x difficulty interaction, $F_{(1, 19)} = 6.30$, $p < 0.05$, such that AS had a greater effect on one-carry easy problems.

The Single vs. RLG comparison gave, a significant effect of task, $F_{(1, 19)} = 29.08$, $p < 0.001$, a significant effect of carry, $F_{(1, 19)} = 15.24$, $p < 0.001$, and a significant effect of difficulty, $F_{(1, 19)} = 95.19$, $p < 0.001$.

Significant interactions included; difficulty x format, $F_{(1, 19)} = 5.73$, $p < 0.05$, such that linear format gave greater errors for easy problems, whereas columnar format gave greater errors for hard problems, modified by a significant task x difficulty x format interaction, $F_{(1, 19)} = 8.56$, $p < 0.01$, such that under single-task conditions columnar format gave greater errors for hard problems, while under RLG, linear format gave greater errors for hard problems. A marginally significant carry x difficulty x format interaction, $F_{(1, 19)} = 3.71$, p

< 0.07, such that linear format gave a relatively greater increase in errors for easy, one-carry problems.

The AS vs. RLG comparison demonstrated a significant effect of carry, $F_{(1, 19)} = 18.00$, $p < 0.001$, and a significant effect of difficulty, $F_{(1, 19)} = 95.38$, $p < 0.001$. Results also included the marginally significant carry x difficulty interaction, $F_{(1, 19)} = 4.12$, $p < 0.06$, such that the carry operation led to relatively greater errors for easy problems, modified by the significant task x carry x difficulty interaction, $F_{(1, 19)} = 5.26$, $p < 0.05$, such that, as in the Single vs. AS comparison, AS had a greater effect on one-carry easy problems.

A significant difficulty x format interaction, $F_{(1, 19)} = 8.09$, $p < 0.01$, such that easy problems showed no difference in error rates for linear and columnar formats, whereas hard problems showed an advantage for columnar format, modified by a carry x difficulty x format interaction, such that linear format led to a relatively greater increase in error rates for easy, one-carry problems. A significant task x carry x format interaction, $F_{(1, 19)}$, $p < 0.01$, such that the advantage of columnar format for one-carry problems was lost under AS, but remained for RLG.

5.5.3 Multiplication experiment – interpretation and discussion of results

Figure 33 provides a visual summary of the multiplication experiment results, and a visual comparison with the addition and subtraction experiment results. The multiplication results are complicated by the inclusion of the difficulty factor, but in terms of both response times and error rates, are somewhat different to the addition and subtraction experiments. If the effect of difficulty is temporarily ignored, then the response time results for Single vs. AS show two interactions absent in the addition and subtraction experiments, task x format and carry x format.

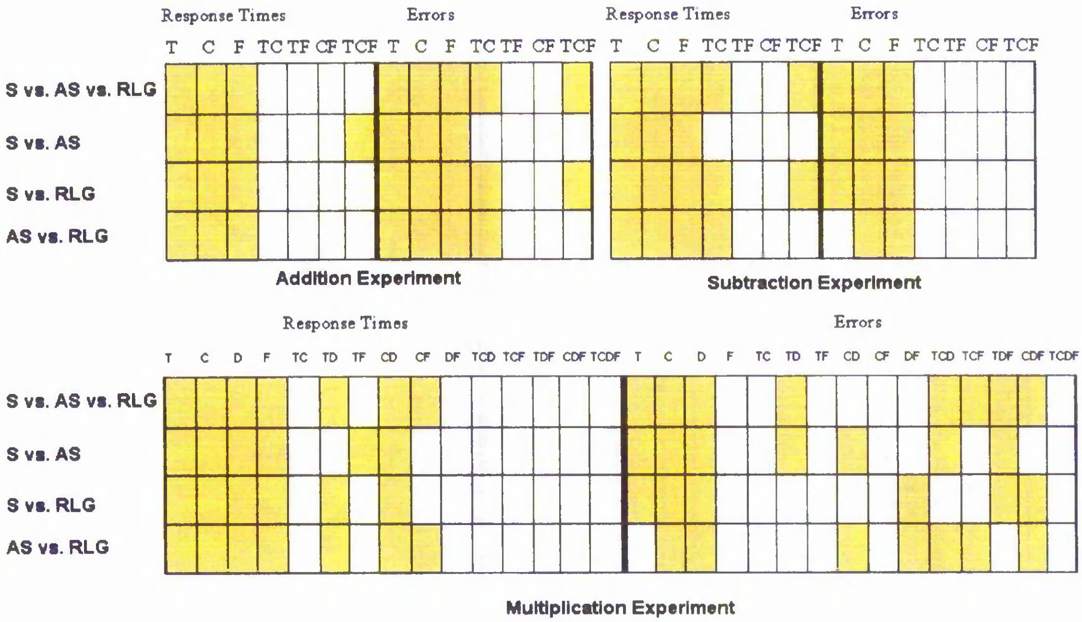

Figure 33. Comparison and summary of addition, subtraction and multiplication experiment results

The task x format interaction is shown in figure 34, and the carry x format interaction in figure 35.

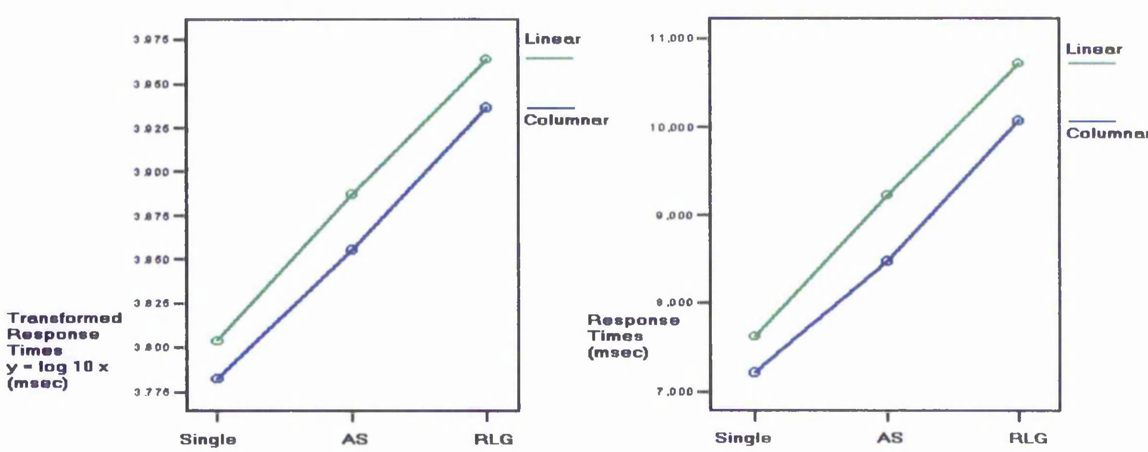

Figure 34. The task x format interaction for transformed and untransformed multiplication response times

Figure 34 indicates that the advantage of columnar format was greater under AS than single-task conditions, but that AS was not significantly different to RLG. This interaction is apparent in the graphs of both transformed and untransformed scores, and is therefore not considered to be due to the log transformation of response scores.

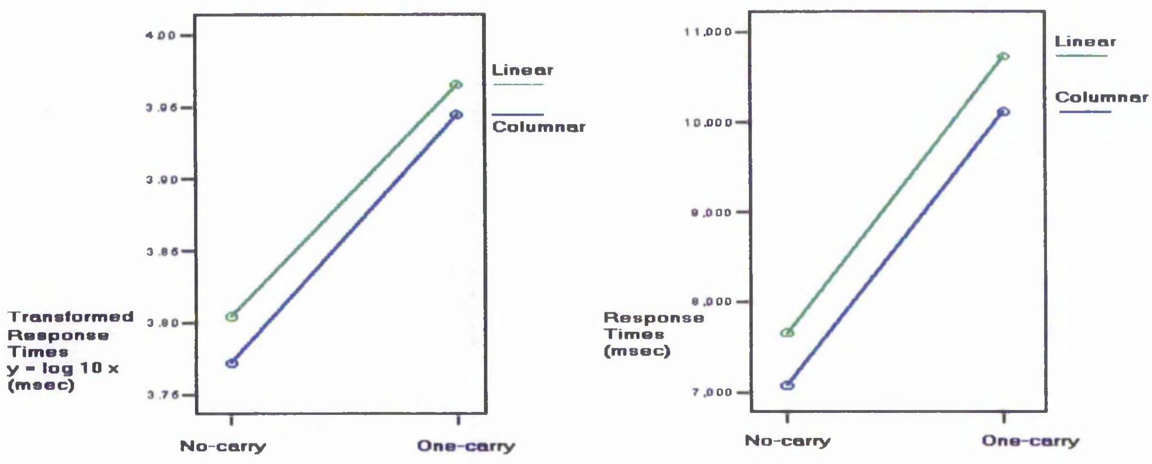

Figure 35. The carry x format interaction for transformed and untransformed multiplication response times

The carry x format interaction in figure 35 is only apparent in the transformed scores, and could therefore represent nothing more than a sub-additive interaction due to the log transformation of response times. The response time carry x difficulty interaction is shown in figure 36, and the task x difficulty interaction in figure 37.

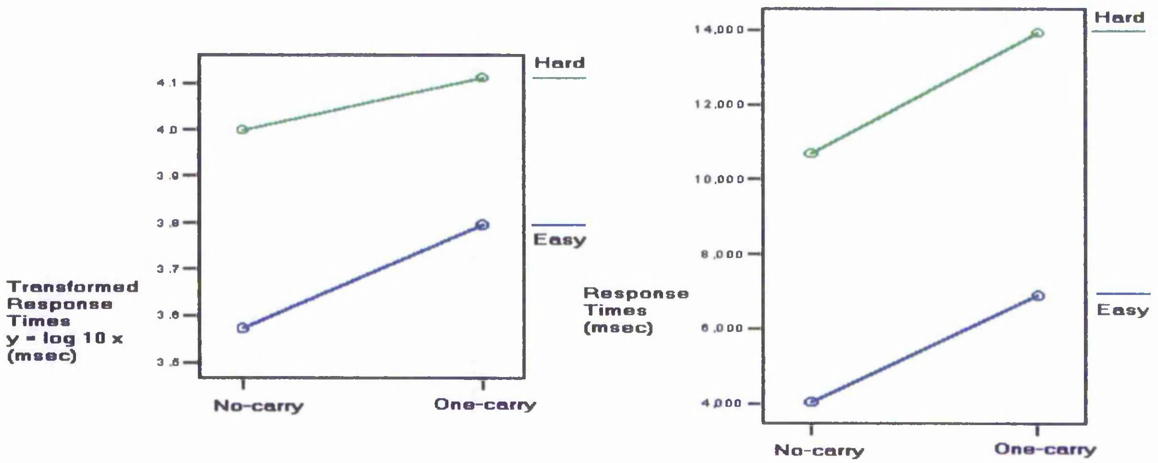

Figure 36. The carry x difficulty interaction for transformed and untransformed multiplication response times

The carry x difficulty interaction in the graph of transformed scores in figure 36 is not apparent in the graph of untransformed scores, and again, this could represent nothing more than a sub-additive effect due to the log transformation of response times.

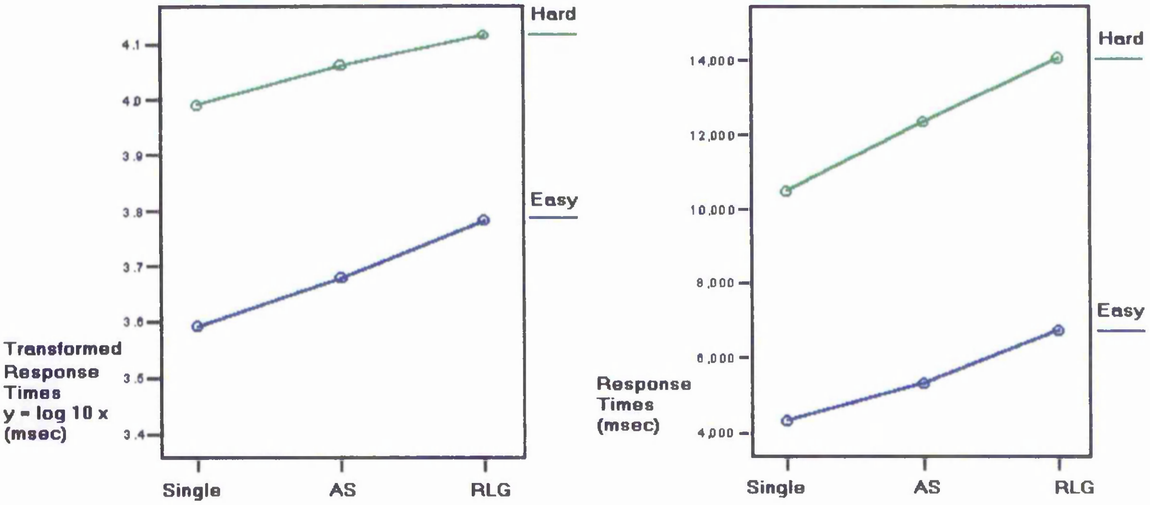

Figure 37. The task x difficulty interaction for transformed and untransformed multiplication response times

The graph of the task x difficulty interaction of transformed scores in figure 37 appears to show that RLG gives a relatively greater increase in response times for easy problems, but this interpretation is not supported by the graph of the same interaction in the untransformed scores. The graph of untransformed scores indicates a progressive increase in response times between easy and hard problems under AS and RLG, whereas the graph of transformed scores indicates a progressive decrease between easy and hard problems (Single vs. RLG, AS vs. RLG). This is evidence for the possibility of a sub-additive effect due to the log transformation of response times, and this appears to hide the increase in response times under AS and RLG as problem difficulty increases. The error score results for the multiplication experiment show no significant two-way task x carry interactions, but do show several more complex three-way interactions which allow the error scores to be interpreted, and which are shown in figures 38 - 41.

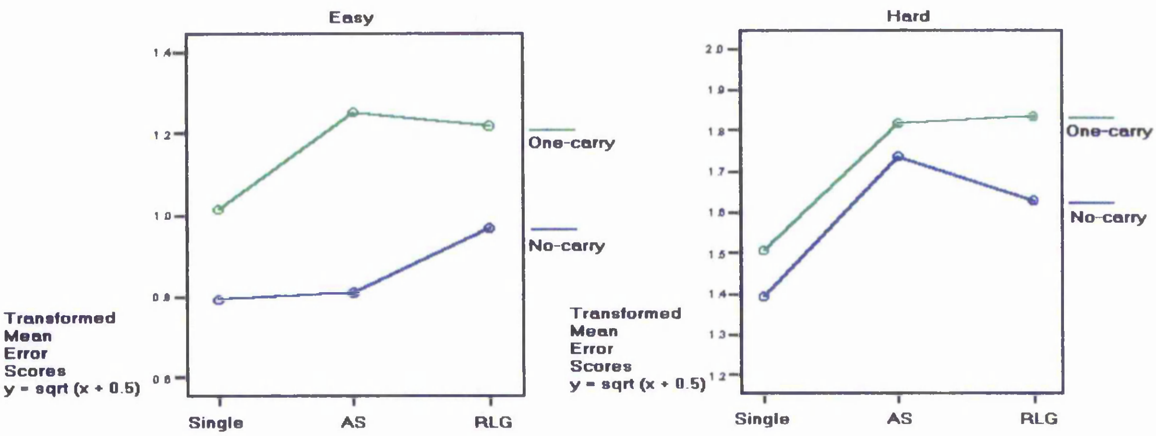

Figure 38. The task x carry x difficulty interaction– multiplication experiment transformed error scores

As can be seen in figure 38, articulatory suppression has virtually no effect on no-carry easy problems in comparison to its effect on no-carry hard problems, but overall, the effect of AS is not significantly different than the effect of RLG. This pattern of results appears to suggest that phonological coding is involved in the carry operation more for hard than for easy problems, and that while AS and RLG are about equally disruptive to one-carry problems at both difficulty levels, AS is more disruptive of hard no-carry problems, while RLG is more disruptive of easy no-carry problems. Again, a possible explanation is that easy, no-carry problems might be solved by a *reading / subitising* process rather than by calculation.

Trbovitch and LeFevre (2003) suggested that horizontal presentation appeared to activate phonological codes, whereas vertical presentation appeared to activate visual codes. If this is so, then it is possible that linear rather than columnar presentation would be more prone to disruption by the use of articulatory suppression, and perhaps to some extent by the articulatory component of RLG. Both linear and columnar no-carry problems can be *read* from left to right, but columnar format also requires up and down eye movements, whereas linear problems perhaps require repeated back and forwards scanning. Figure 39 gives the three-way task x carry x format interaction, as the task x carry interaction at each level of format, so that the effects of format can be further investigated.

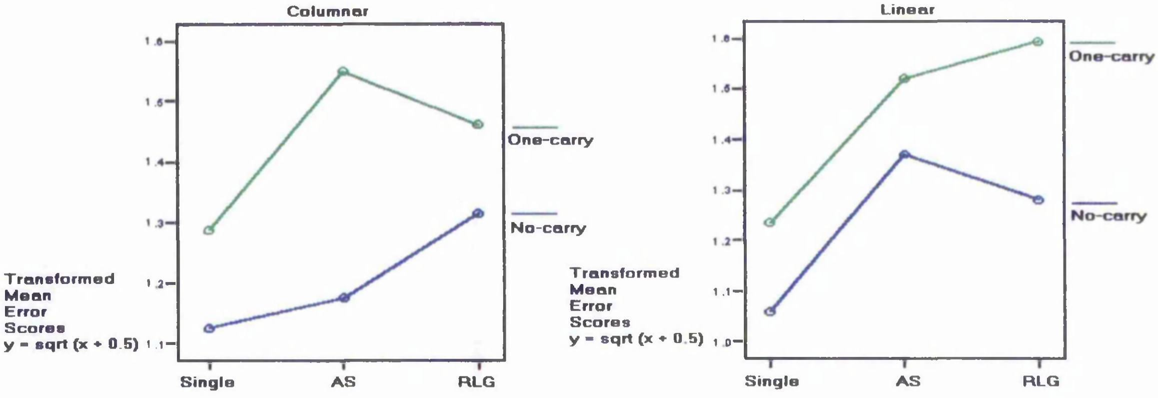

Figure 39. The task x carry x format interaction – multiplication experiment transformed error scores

The pattern of interaction in figure 39 is quite similar to that found in figure 38, but here AS appears to have little differential effect on problems in linear format, but a relatively large differential effect on problems in columnar format. This cannot be due to easy / hard differences, for they are collapsed over both columnar and linear formats, and it appears that linear format is impacted more by articulatory suppression, possibly because of the disruption of *reading / subitising* processes, but mainly because columnar no-carry problems have lower associated error scores. The interactions in figures 38 and 39 both arise as a result of the lack of effect of articulatory suppression on no-carry problems, and especially when these problems are easy and in columnar format.

Columnar format appears to support the carry operation under RLG, but RLG does not appear to impact no-carry columnar problems differently to no-carry linear problems.

Figure 40 illustrates the task x difficulty x format interaction, as the task x difficulty interaction at each level of format.

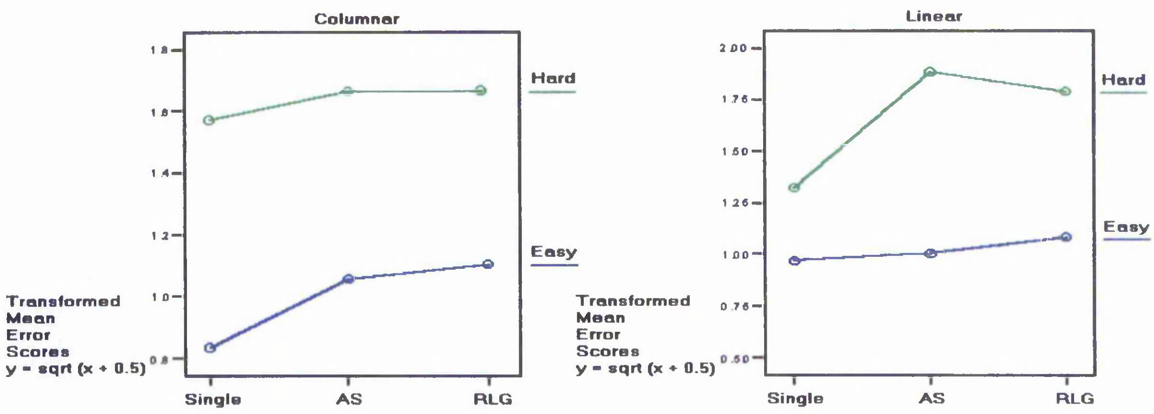

Figure 40. The task x difficulty x format interaction– multiplication experiment transformed error scores

Figure 40 appears to show that the AS and RLG tasks interact only with problems in linear format, but the interaction results partly from the difference between columnar hard single-task problems and linear hard single-task problems, suggesting that in single-task conditions, hard problems in linear format produce less errors. The final three-way interaction to consider is the carry x difficulty x format interaction, and figure 41 illustrates this as the carry x difficulty interaction at each level of format.

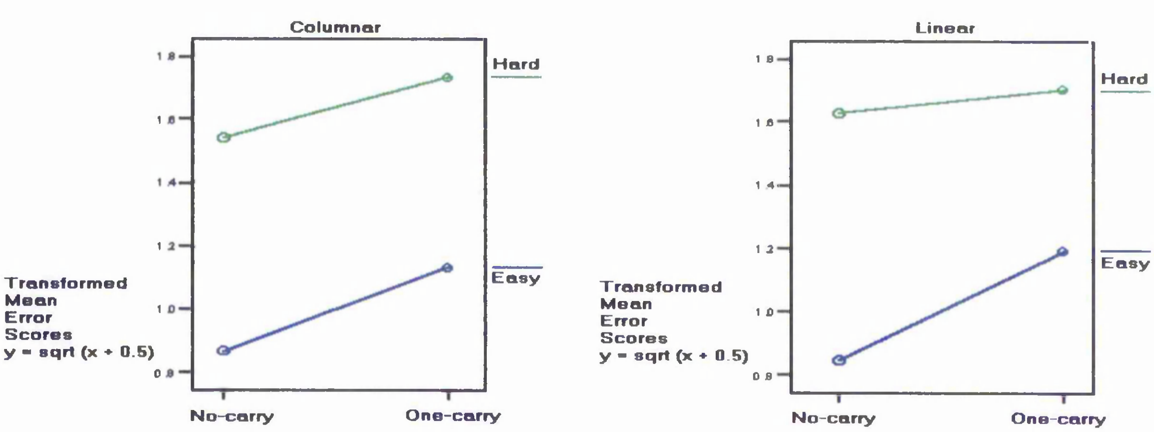

Figure 41. The carry x difficulty x format interaction– multiplication experiment transformed error scores

Figure 41 shows a relatively greater increase in errors for hard no-carry problems in linear format, but this is collapsed across the three task levels, and this three-way interaction was only significant for the Single vs. RLG and AS vs. RLG comparisons. To summarise the four interactions, AS appears not to disrupt easy no-carry problems, and especially easy no-carry columnar problems, but AS and RLG both increase errors for hard linear problems, and RLG more so for hard no-carry linear problems.

This suggests that easy no-carry problems, especially in columnar format, do not require phonological processes, but that problems in linear format do require phonological processing. It also suggests that the carry operation involves some element of phonological processing, but that this is mediated both by difficulty and format. The effect of RLG is also mediated by difficulty, carry and format. This is consistent both with the horizontal (phonological) vs. vertical (visuo-spatial) claims of Trbovitch and LeFevre (2003), and to some extent with the easy / hard differences found by Seitz and Schumann-Hengsteler (2000). The role of phonological processing could be due to *reading* processes being engaged, but it may also be due to the relative impairment of access to the store of multiplication facts. It is interesting that the results for multiplication implicated phonological coding whereas the results for addition and subtraction did not.

Clearly, addition, subtraction and multiplication seem to involve different processing. Baseline vs. dual-task measurement of AS articulation rate again demonstrated no significant difference, which, given the evidence for phonological processing found in this experiment, suggests that the articulation of 'A, B, C' is relatively automatic and not affected by other phonological processing. However, the pseudo-randomness index based on Chi-square goodness-of-fit differences was significantly different, (t = - 12.908, p < 0.001), such that articulation of 'A, B, C, D, E' was much less random under RLG. Correlations of all 24 pairs of response times and error scores provided five significant positive correlations; single carry hard columnar, (r = 0.469, p < 0.05), AS carry easy columnar, (r = 0.601, p < 0.01), AS carry hard columnar, (r = 0.541, p < 0.05), RLG carry easy columnar, (r = 0.450, p < 0.05), RLG carry hard linear, (r = 0.456, p < 0.05). Given the multiple use of correlation, the significance of these values is questionable, but the point is that they are all positive, offering no evidence for a speed / accuracy trade-off, and only limited evidence for error rates increasing with response times in 5 of 24 categories.

5.5.4 Item analysis of multiplication experiment problems

Table 25 gives the results of one-way ANOVA tests on problem response times from all twenty-four task x carry x difficulty x format levels.

Level	F Ratio	Probability
Single no-carry easy columnar	F (8, 162) = 0.83	p = 0.576 n.s.
Single no-carry easy linear	F (8, 163) = 1.13	p = 0.345 n.s.
Single no-carry hard columnar	F (8, 126) = 1.48	p = 0.170 n.s.
Single no-carry hard linear	F (8, 145) = 0.37	p = 0.935 n.s.
Single carry easy columnar	F (8, 157) = 0.55	p = 0.818 n.s.
Single carry easy linear	F (8, 154) = 1.44	p = 0.183 n.s.
Single carry hard columnar	F (8, 123) = 1.17	p = 0.324 n.s.
Single carry hard linear	F (8, 136) = 0.56	p = 0.811 n.s.
AS no-carry easy columnar	F (8, 166) = 1.16	p = 0.326 n.s.
AS no-carry easy linear	F (8, 152) = 0.86	p = 0.551 n.s.
AS no-carry hard columnar	F (8, 119) = 0.26	p = 0.977 n.s.
AS no-carry hard linear	F (8, 100) = 0.29	p = 0.967 n.s.
AS carry easy columnar	F (8, 136) = 0.77	p = 0.633 n.s.
AS carry easy linear	F (8, 145) = 1.19	p = 0.308 n.s.
AS carry hard columnar	F (8, 104) = 0.38	p = 0.929 n.s.
AS carry hard linear	F (8, 99) = 0.37	p = 0.936 n.s.
RLG no-carry easy columnar	F (8, 154) = 0.97	p = 0.466 n.s.
RLG no-carry easy linear	F (8, 164) = 0.76	p = 0.641 n.s.
RLG no-carry hard columnar	F (8, 131) = 0.74	p = 0.658 n.s.
RLG no-carry hard linear	F (8, 112) = 1.64	p = 0.122 n.s.
RLG carry easy columnar	F (8, 146) = 0.22	p = 0.987 n.s.
RLG carry easy linear	F (8, 138) = 0.26	p = 0.978 n.s.
RLG carry hard columnar	F (8, 110) = 0.29	p = 0.968 n.s.
RLG carry hard linear	F (8, 94) = 1.04	p = 0.413 n.s.

Table 25. Item analysis of within-category correct multiplication problem response times

The results of this item analysis demonstrate that although there is some variation within categories, there are no significant differences in response times within any of the 24 categories, and individual problems within each category can be taken to give essentially similar response times.

5.5.5 Strategy use in the multiplication experiment

The strategy choice sheet used for the multiplication and division experiments differed from that used in the addition and subtraction experiments, in that there were now only six categories, (see appendix F), but in brief, the six categories were as follows:

A) Overshooting / Undershooting

B) Partitioning

C) Mental Carrying

D) Change numbers in proportion

E) Retrieval

F) Any other strategy

Participants were provided with strategy choice sheets in advance, but were also allowed a period of familiarisation prior to the experimental trials, and the strategy choice sheets were directly available to participants during the period of the experiments.

As previously, participants stated strategy use after the answer to each problem, and table 26 gives the mean percentage strategy use at each experimental level. As can be seen from an inspection of table 26, low percentage strategy use, and the small numbers of participants using strategies A, D, E and F provided no basis for the analysis of these strategies, and analysis was therefore restricted to strategies B and C.

Strategy	A Overshoot	B Partition	C Carry	D Proportion	E Retrieval	F Other
Single no-carry easy columnar	0	33	36	9	13	9
Number using strategy	0	14	11	2	4	2
Single no-carry easy linear	0	29	38	8	16	9
Number using strategy	0	13	12	2	5	3
Single no-carry hard columnar	3	53	37	1	5	1
Number using strategy	2	13	11	1	2	1
Single no-carry hard linear	3	50	37	1	8	1
Number using strategy	2	13	10	1	3	1
Single carry easy columnar	0	36	41	3	13	7
Number using strategy	0	15	13	2	5	2
Single carry easy linear	2	33	37	6	14	8
Number using strategy	3	15	13	2	5	2
Single carry hard columnar	5	42	42	1	8	2
Number using strategy	3	11	13	1	4	3
Single carry hard linear	5	56	33	1	5	1
Number using strategy	3	13	9	1	2	1
AS no-carry easy columnar	0	26	36	10	18	10
Number using strategy	0	9	11	2	5	2
AS no-carry easy linear	0	28	35	10	18	9
Number using strategy	0	11	9	2	6	2
AS no-carry hard columnar	2	54	36	3	5	0
Number using strategy	1	12	10	2	2	0
AS no-carry hard linear	4	51	37	3	4	1
Number using strategy	2	12	9	3	1	1
AS carry easy columnar	0	41	38	5	10	6
Number using strategy	0	14	11	2	4	2
AS carry easy linear	1	35	36	8	13	7
Number using strategy	1	12	10	2	6	2
AS carry hard columnar	4	52	36	3	4	1
Number using strategy	3	13	10	2	2	1
AS carry hard linear	5	54	33	2	5	1
Number using strategy	2	13	9	1	2	1
RLG no-carry easy columnar	0	26	37	9	18	10
Number using strategy	0	7	8	2	5	2
RLG no-carry easy linear	0	28	36	9	17	10
Number using strategy	0	9	11	2	4	2
RLG no-carry hard columnar	3	54	36	1	6	0
Number using strategy	1	13	10	1	3	0
RLG no-carry hard linear	3	46	38	4	9	0
Number using strategy	2	10	10	2	5	0
RLG carry easy columnar	1	40	39	4	9	7
Number using strategy	1	14	11	2	5	2
RLG carry easy linear	1	35	39	6	13	6
Number using strategy	2	11	9	2	5	2
RLG carry hard columnar	4	53	32	3	7	1
Number using strategy	1	12	10	2	4	2
RLG carry hard linear	5	52	36	1	5	1
Number using strategy	3	12	10	1	2	1

Table 26. Mean percentage strategy use and number choosing strategy at each level of multiplication

A 3 (task) x 2 (carry) x 2 (diff) x 2 (format) repeated-measures ANOVA investigated the use of strategy B at each of the experimental levels, and gave the following results. A significant effect of carry, $F_{(1, 19)} = 8.17$, $p < 0.01$, such that strategy B was used less for no-carry problems, modified by a significant task x carry interaction, $F_{(2, 38)} = 4.40$, $p < 0.05$, such that this was true only under AS and RLG.

A carry x difficulty interaction, $F_{(1, 19)} = 5.97$, $p < 0.05$, such that strategy B was used less with easy no-carry problems, further modified by a significant carry x difficulty x format interaction, $F_{(1, 19)} = 5.44$, $p < 0.05$, such that strategy B was used less in columnar format easy no-carry problems.

A task x format interaction, $F_{(1, 19)} = 11.85$, $p < 0.01$, such that Strategy B was used less with problems in linear format only under RLG, but further modified by a task x carry x format crossover interaction, $F_{(1, 19)} = 4.68$, $p < 0.05$, such that strategy B was used less in single no-carry problems in columnar format, but more in RLG no-carry problems in columnar format. A similar analysis of strategy C revealed only a carry x format interaction, $F_{(1, 19)}$, $p < 0.05$, such that linear one-carry problems used strategy C less than columnar one-carry problems.

Once again, it is possible that differential use of strategies across experimental levels might have an impact, but this can only be checked by analysing response time differences for each strategy. Table 27 reports a series of one-way ANOVA tests on strategy response time differences within each experimental level, with Games-Howell post-hoc multiple comparison tests for strategies B vs. C.

Level	F Ratio	Probability	B vs. C
Single no-carry easy columnar	$F(4, 166) = 6.01$	$p < 0.001$	B = C n.s.
Single no-carry easy linear	$F(4, 167) = 4.80$	$p < 0.001$	B = C n.s.
Single no-carry hard columnar	$F(4, 130) = 9.60$	$p < 0.001$	B = C n.s.
Single no-carry hard linear	$F(4, 149) = 17.26$	$p < 0.001$	B = C n.s.
Single carry easy columnar	$F(4, 161) = 3.88$	$p < 0.01$	B = C n.s.
Single carry easy linear	$F(5, 157) = 14.40$	$p < 0.001$	B = C n.s.
Single carry hard columnar	$F(3, 128) = 25.60$	$p < 0.001$	B > C, $p < 0.01$
Single carry hard linear	$F(4, 140) = 7.33$	$p < 0.001$	B = C n.s.
AS no-carry easy columnar	$F(4, 170) = 12.39$	$p < 0.001$	B = C n.s.
AS no-carry easy linear	$F(4, 156) = 5.37$	$p < 0.001$	B = C n.s.
AS no-carry hard columnar	$F(3, 124) = 2.49$	p = n.s.	B = C n.s.
AS no-carry hard linear	$F(4, 104) = 17.09$	$p < 0.001$	B > C, $p < 0.05$
AS carry easy columnar	$F(4, 140) = 7.20$	$p < 0.001$	B = C n.s.
AS carry easy linear	$F(4, 149) = 13.65$	$p < 0.001$	B > C, $p < 0.05$
AS carry hard columnar	$F(3, 109) = 10.17$	$p < 0.001$	B > C, $p < 0.001$
AS carry hard linear	$F(3, 104) = 15.09$	$p < 0.001$	B > C, $p < 0.001$
RLG no-carry easy columnar	$F(4, 158) = 5.72$	$p < 0.001$	B = C n.s.
RLG no-carry easy linear	$F(4, 168) = 2.47$	$p < 0.05$	B = C n.s.
RLG no-carry hard columnar	$F(3, 136) = 8.26$	$p < 0.001$	B > C, $p < 0.05$
RLG no-carry hard linear	$F(4, 116) = 4.17$	$p < 0.01$	B = C n.s.
RLG carry easy columnar	$F(4, 150) = 6.75$	$p < 0.001$	B = C n.s.
RLG carry easy linear	$F(5, 141) = 9.04$	$p < 0.001$	B = C n.s.
RLG carry hard columnar	$F(4, 114) = 10.13$	$p < 0.001$	B > C, $p < 0.001$
RLG carry hard linear	$F(3, 99) = 15.11$	$p < 0.001$	B > C, $p < 0.001$

Table 27. Analysis of response times and strategy use with post-hoc comparisons of strategies B and C

There is no clear pattern of results in table 27, although the 8 out of 24 instances where strategy B provides significantly greater response times than strategy C, seem to involve hard problems, 7 from 8, and to involve mostly AS and RLG, 7 from 8. As with the previous addition and subtraction experiments, interpretation is complicated by individual differences, in that not only was there evidence of exclusive, or near exclusive, use of strategies B and C by a relatively large number of participants, but those using strategy C may simply have been faster overall, and there is insufficient data from those switching strategies to disambiguate the effect of speed of processing from effect of strategy.

5.6 The division experiment

The division operation is different to the other three arithmetic operations in that unlike addition and multiplication, there is no suggestion of storage of division facts as verbal word frames, and unlike subtraction, no suggestion that divisions can be solved by comparison of magnitude. Division problems might require transcoding between operations, with multiplication facts being used, e.g. 286 / 2 = (1) x 2, (4) x 2 and (3) x 2, answer 143. Details of the division problems are given in section 5.2, but this experiment differs from the others in that there are three levels of the format variable (linear, traditional and columnar), with the expectation being that as columnar format is not usually encountered in division it might have a negative impact upon response times and errors due to the association of this format with the other arithmetic operations.

5.6.1 Division experiment – correct answer response times

Untransformed mean response times for each task x carry x format level are given in table 28.

LEVEL	Mean	Std. Deviation
Single no-carry linear	3143	828
Single no-carry traditional	3105	978
Single no-carry columnar	3265	925
Single one-carry linear	5988	2219
Single one-carry traditional	5703	2121
Single one-carry columnar	5715	2058
AS no-carry linear	3797	998
AS no-carry traditional	3716	1101
AS no-carry columnar	3721	1001
AS one-carry linear	6825	2324
AS one-carry traditional	6682	2326
AS one-carry columnar	6864	2602
RLG no-carry linear	4742	1356
RLG no-carry traditional	4645	1225
RLG no-carry columnar	4669	1292
RLG one-carry linear	7723	2627
RLG one-carry traditional	7769	2697
RLG one-carry columnar	7744	2486

Table 28. Correct item mean response times and standard deviations – division experiment

A 3 (task) x 2 (carry) x 3 (format) repeated-measures ANOVA demonstrated a significant difference between single-task (M = 4055 msec), AS (M = 4786 msec) and RLG (M = 5768 msec), $F_{(2, 38)} = 76.75$, $p < 0.001$. A significant difference between no-carry (M = 3673 msec) and one-carry (M = 6324 msec), $F_{(1, 19)} = 107.49$, $p < 0.001$. A significant difference between linear format (M = 4864 msec), traditional format (M = 4724 msec) and columnar format (M = 4842 msec), $F_{(2, 38)} = 4.30$, $p < 0.05$. A marginal task x carry x format interaction, $F_{(4, 76)} = 2.78$, $p < 0.06$.

Further analysis using within-subjects contrasts and 2 x 2 x 3 ANOVA provided the following comparisons: Single vs. AS, a significant effect of task, $F_{(1, 19)} = 51.20$, $p < 0.001$, a significant effect of carry, $F_{(1, 19)} = 112.10$, $p < 0.001$, a significant effect of format, $F_{(2, 38)} = 4.69$, $p < 0.05$, and a significant task x carry x format interaction, $F_{(2, 38)} = 3.72$, $p < 0.05$.

Single vs. RLG, a significant effect of task, $F_{(1, 19)} = 92.54$, $p < 0.001$, a significant effect of carry, $F_{(1, 19)} = 102.68$, $p < 0.001$, a significant effect of format, $F_{(2, 38)} = 4.66$, $p < 0.05$, and a marginal task x carry x format interaction, $F_{(2, 38)} = 3.16$, $p < 0.07$. AS vs. RLG, a significant effect of task, $F_{(1, 19)} = 63.20$, $p < 0.001$, and a significant effect of carry, $F_{(1, 19)} = 94.34$, $p < 0.001$.

5.6.2 Division experiment – error scores

Table 29 gives the untransformed mean error scores for each task x carry x format level, but as in all previous experiments, error scores were transformed using y = sqrt (x + 0.5).

A 3 (task) x 2 (carry) x 3 (format) ANOVA, gave the following results, a significant difference between single-task (M = 0.989), AS (M = 1.060) and RLG (M = 1.153), $F_{(2, 38)} = 6.25$, $p < 0.01$. A significant difference between no-carry (M = 0.865) and one-carry (M =

1.270), $F_{(1, 19)} = 36.00$, $p < 0.001$. A significant difference between linear format (M = 1.134), traditional format (M = 1.052) and columnar format (M = 1.017), $F_{(2, 38)} = 4.44$, $p < 0.05$. A significant task x carry x format interaction, $F_{(4, 76)} = 3.12$, $p < 0.05$.

LEVEL	Mean	Std. Deviation
Single no-carry linear	.25	.44
Single no-carry traditional	.10	.31
Single no-carry columnar	.15	.37
Single one-carry linear	1.45	1.50
Single one-carry traditional	.95	1.23
Single one-carry columnar	.95	1.10
AS no-carry linear	.40	.60
AS no-carry traditional	.30	.47
AS no-carry columnar	.05	.22
AS one-carry linear	1.60	1.57
AS one-carry traditional	1.05	.69
AS one-carry columnar	1.35	1.27
RLG no-carry linear	.80	.62
RLG no-carry traditional	.40	.50
RLG no-carry columnar	.35	.49
RLG one-carry linear	1.40	1.60
RLG one-carry traditional	1.80	1.32
RLG one-carry columnar	1.35	1.42

Table 29. Mean error scores and standard deviations (division experiment)

Further analysis using within-subjects contrasts and 2 x 2 x 3 ANOVA provided the following comparisons: Single vs. AS, a significant effect of carry, $F_{(1, 19)} = 49.58$, $p < 0.001$, and a significant effect of format, $F_{(2, 38)} = 4.34$, $p < 0.05$.

Single vs. RLG, a significant effect of task, $F_{(1, 19)} = 7.77$, $p < 0.01$, a significant effect of carry, $F_{(1, 19)}$ 27.87, $p < 0.001$, a significant effect of format, $F_{(2, 38)} = 3.55$, $p < 0.05$, and a significant task x carry x format interaction, $F_{(2, 38)} = 4.52$, $p < 0.05$.

AS vs. RLG, a significant effect of task, $F_{(1, 19)} = 9.36$, $p < 0.01$, a significant effect of carry, $F_{(1, 19)} = 25.95$, $p < 0.001$, and a significant task x carry x format interaction, $F_{(2, 38)} = 4.89$, $p < 0.01$.

5.6.3 Division experiment – interpretation and discussion of results

Figure 42 gives a visual summary of division experiment response times and error results.

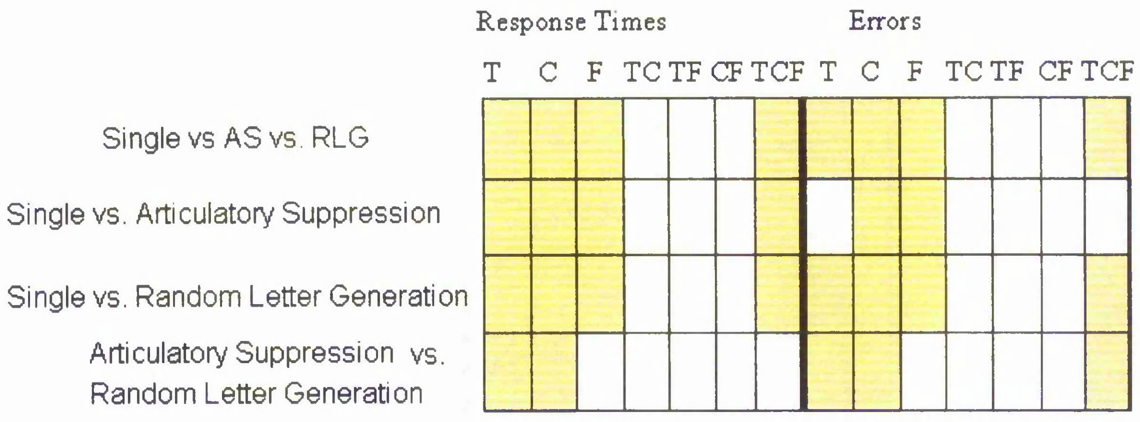

Figure 42. Summary of division experiment results

The response time results are best explored through a visual inspection of the three-way task x carry x format interaction, and figures 43 and 44 present this, for both transformed and untransformed scores, as the task x format interaction at each of the carry levels.

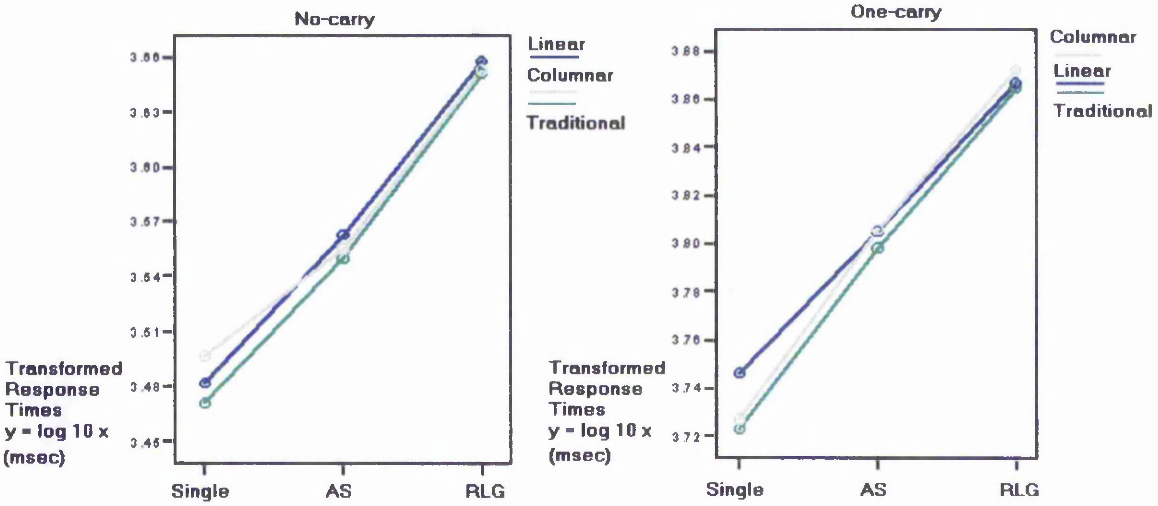

Figure 43. The task x format x carry interaction –division experiment transformed response times

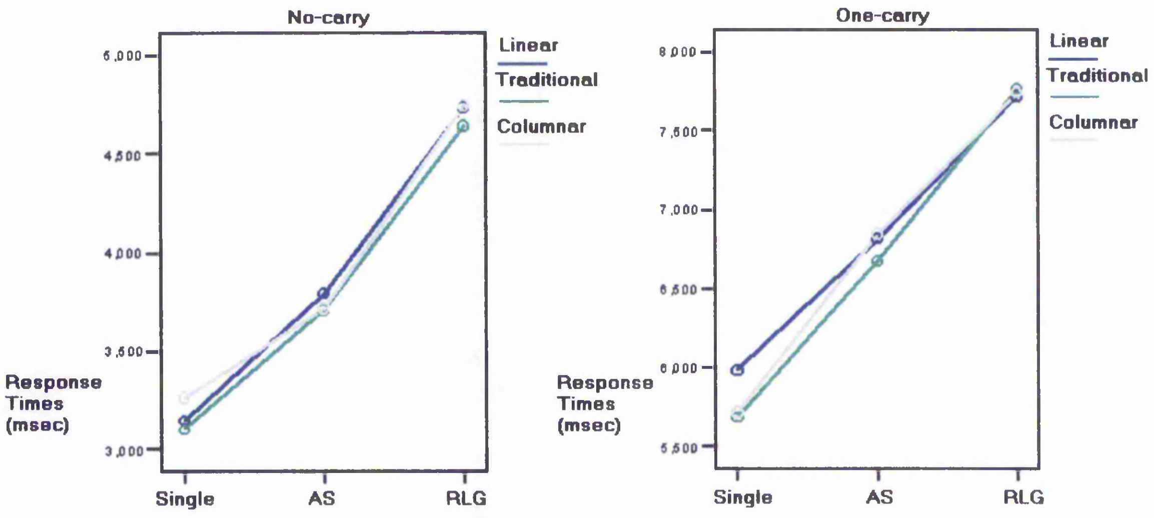

Figure 44. The task x format x carry interaction –division experiment untransformed response times

As seen in both figure 43 and figure 44, overall there is no significant effect of format between the AS vs. RLG levels, but for Single vs. AS, traditional format provides lower response times than either columnar format, $F_{(1, 19)} = 8.82$, $p < 0.01$, or linear format, $F_{(1, 19)} = 8.00$, $p < 0.01$. This interaction is therefore unlikely to be due solely to the log transformation of response times.

The Single vs. AS significant task x carry x format interaction is due to differences between linear and columnar format at the two carry levels, $F_{(1, 19)} = 7.22$, $p < 0.05$, such that for single-task no-carry problems, columnar format gives longer response times than both linear format, whereas for single-task carry problems, linear format gives longer response times than columnar format.

For Single vs. RLG, traditional format also gives lower response times than either columnar format, $F_{(1, 19)} = 5.93$, $p < 0.05$, or linear format, $F_{(1, 19)} = 7.28$, $p < 0.01$, and the task x carry x format interaction is due to the same differences between linear and columnar format at the two carry levels, $F_{(1, 19)} = 4.60$, $p < 0.05$, as described for the Single vs. AS task x carry x format interaction.

Overall, there is no evidence in the response time data for AS and RLG having a differential impact on the carry process, or that AS and RLG have a differential impact upon format. The single-task differences are again suggestive of *reading* processes, linear format allowing answers to be *read* for single-task no-carry problems, but AS and the carry operation removing this possibility. However, the advantage of traditional format is not lost under AS and the carry operation, and it is not clear why traditional format should provide faster encoding.

As with response times, the error scores are also best investigated through examining the task x carry x format interaction. This is illustrated in figure 45, where overall, it can be seen that there is no significant difference in the number of errors between Single and AS levels, although differences do occur for Single vs. RLG and AS vs. RLG comparisons. Overall, the error rates for the three levels of format are not significantly different for the AS vs. RLG comparison, but as can be seen, error scores do vary considerably with format and carry.

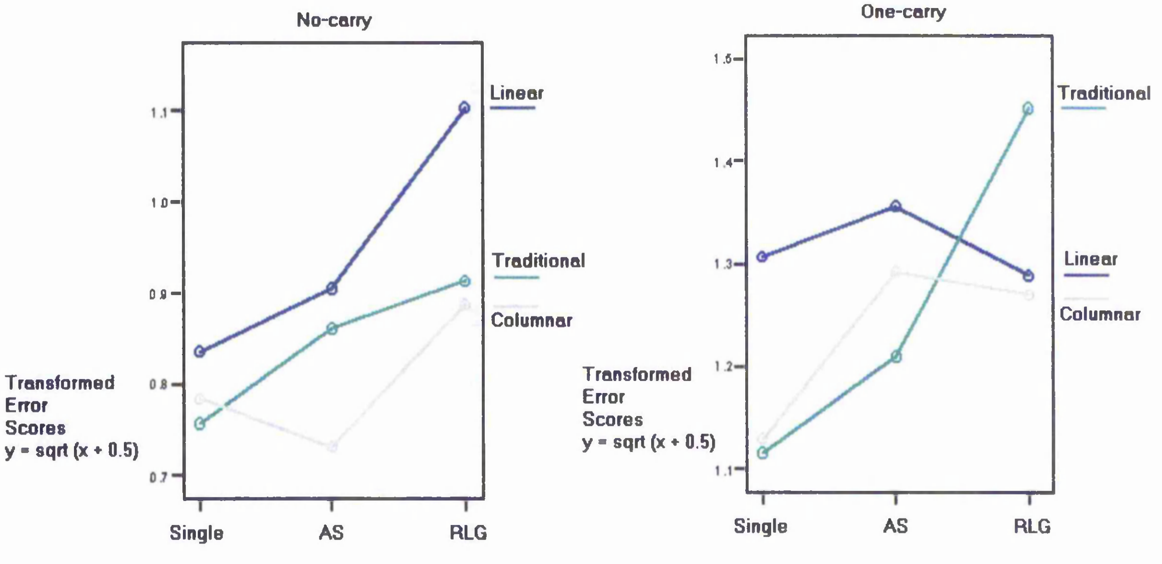

Figure 45. The task x carry x format interaction – division experiment transformed error scores

Format levels are not significantly different for one-carry problems, but for no-carry problems, linear format gave significantly greater errors than either columnar format, $F_{(1, 19)} = 4.51$, $p < 0.05$, or traditional format, $F_{(1, 19)} = 9.04$, $p < 0.01$.

There is no evidence in the division experiment error data for the interaction of RLG with the carry operation, for RLG has a greater impact on no-carry problems, and appears to interact with format rather than carry. Similarly, there is no evidence that AS errors are greater than single-task errors, and therefore no evidence for phonological coding in the carry operation.

The results of the division experiment seem to be closer to those for the addition and subtraction experiments than those for the multiplication experiment, but the lack of appropriate task x carry interactions under RLG, perhaps suggests a closer link to the subtraction results. It is equally possible that the four arithmetic operations are all different in terms of WM involvement, and for comparison, results for all four experiments are given in figure 46.

Baseline versus dual-task comparison of AS articulation rate demonstrated no significant difference, but the baseline versus dual-task comparison of RLG pseudo-randomness index did demonstrate a significant difference, ($t = -12.908$, $p < 0.001$), such that under dual-task conditions RLG was more random. Correlation of category mean response times with category mean error rates provided only one significant result, for RLG one-carry linear problems, ($r = 0.496$, $p < 0.05$), and this is consistent with the lack of correlation, or slight positive correlation, found in the previous experiments.

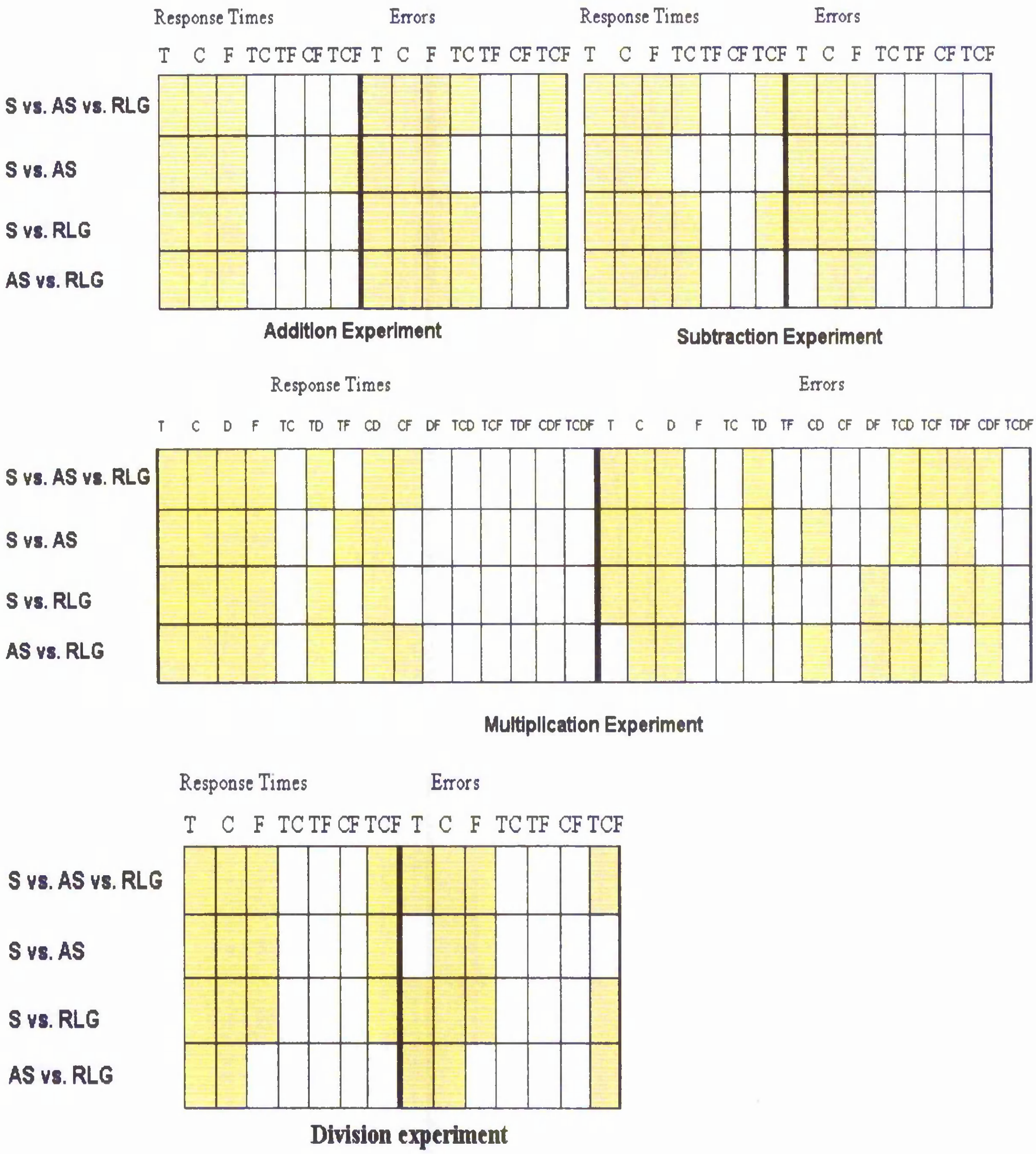

Figure 46. Summary of results for all four second series experiments

5.6.4 Division experiment – item analysis

Table 30 gives the results of one-way ANOVA tests on problem response times from all eighteen task x carry x format levels. There are no significant differences between these levels, and each problem within the categories can be assumed to provide essentially similar response times.

Level	F Ratio	Probability
Single no-carry linear	$F_{(8, 163)} = 1.54$	$p = 0.148$ n.s.
Single no-carry traditional	$F_{(8, 165)} = 0.80$	$p = 0.606$ n.s.
Single no-carry columnar	$F_{(8, 153)} = 1.48$	$p = 0.170$ n.s.
Single one-carry linear	$F_{(8, 141)} = 0.67$	$p = 0.717$ n.s.
Single one-carry traditional	$F_{(8, 152)} = 0.27$	$p = 0.974$ n.s.
Single one-carry columnar	$F_{(8, 152)} = 0.61$	$p = 0.767$ n.s.
AS no-carry linear	$F_{(8, 161)} = 0.48$	$p = 0.869$ n.s.
AS no-carry traditional	$F_{(8, 164)} = 0.62$	$p = 0.761$ n.s.
AS no-carry columnar	$F_{(8, 169)} = 1.48$	$p = 0.169$ n.s.
AS one-carry linear	$F_{(8, 137)} = 0.47$	$p = 0.875$ n.s.
AS one-carry traditional	$F_{(8, 148)} = 0.17$	$p = 0.994$ n.s.
AS one-carry columnar	$F_{(8, 141)} = 0.21$	$p = 0.988$ n.s.
RLG no-carry linear	$F_{(8, 152)} = 0.52$	$p = 0.841$ n.s.
RLG no-carry traditional	$F_{(8, 158)} = 0.92$	$p = 0.503$ n.s.
RLG no-carry columnar	$F_{(8, 160)} = 1.07$	$p = 0.385$ n.s.
RLG one-carry linear	$F_{(8, 136)} = 0.93$	$p = 0.497$ n.s.
RLG one-carry traditional	$F_{(8, 134)} = 0.46$	$p = 0.886$ n.s.
RLG one-carry columnar	$F_{(8, 143)} = 0.57$	$p = 0.799$ n.s.

Table 30. Item analysis of within-category correct division problem response times.

5.6.5 Strategy use in the division experiment

The division experiment used the same strategy choice sheet as in the multiplication experiment, and as in previous experiments, the sheet was directly available throughout the experiment and participants stated which strategy they used after each problem answer. Table 31 shows that the number of participants and mean percentage usage of strategies A, D, E and F are too small for analysis, so analysis was restricted to strategies B and C.

Table 31 gives the mean percentage strategy use at each experimental level, and the number of participants choosing each strategy.

Strategy	A Overshoot	B Partition	C Carry	D Proportion	E Retrieval	F Other
Single no-carry linear	0	24	57	7	3	9
Number using strategy	0	9	15	2	2	2
Single no-carry traditional	0	24	60	4	3	9
Number using strategy	0	9	14	2	2	2
Single no-carry columnar	0	24	59	4	3	10
Number using strategy	0	10	15	2	1	2
Single carry linear	2	34	56	2	1	5
Number using strategy	1	8	15	1	1	1
Single carry traditional	1	37	55	3	2	2
Number using strategy	1	9	14	2	2	1
Single carry columnar	2	31	56	4	3	4
Number using strategy	2	8	14	2	2	2
AS no-carry linear	0	24	56	8	2	10
Number using strategy	0	8	13	2	2	2
AS no-carry traditional	0	24	56	8	2	10
Number using strategy	0	9	14	2	1	2
AS no-carry columnar	0	21	57	8	4	10
Number using strategy	0	5	12	2	2	2
AS carry linear	3	31	62	2	0	2
Number using strategy	1	9	16	1	0	1
AS carry traditional	3	39	53	3	1	1
Number using strategy	1	11	14	1	1	1
AS carry columnar	3	39	53	3	1	1
Number using strategy	1	10	15	1	1	1
RLG no-carry linear	0	28	52	8	2	10
Number using strategy	0	8	13	2	1	2
RLG no-carry traditional	0	30	49	8	3	10
Number using strategy	0	9	12	2	1	2
RLG no-carry columnar	1	21	57	8	4	9
Number using strategy	1	8	14	2	1	2
RLG carry linear	2	39	55	1	1	2
Number using strategy	1	10	15	1	2	1
RLG carry traditional	2	40	49	6	2	1
Number using strategy	1	10	13	2	2	1
RLG carry columnar	2	35	58	4	0	1
Number using strategy	1	8	15	2	0	2

Table 31. Mean percentage strategy use and number choosing strategy at each level of division

A 3 (task) x 2 (carry) x 3 (format) repeated-measures ANOVA of the percentage use of strategy B demonstrated no significant effects or interactions. The same analysis of the percentage use of strategy C demonstrated a task x format interaction only under RLG, $F_{(1, 19)} = 4.68$, $p < 0.05$, such that the percentage use of C was less in columnar than traditional format, and although this could only affect two of the eighteen experimental levels, it is informative to know if the two strategies are associated with different response times. Table

32 reports a series of one-way ANOVA tests on response times and overall strategy use, with Games-Howell post-hoc multiple comparison tests for strategies B vs. C.

Level	F Ratio	Probability	B vs. C
Single no-carry linear	$F_{(4, 170)} = 3.97$	$p = 0.004$	B = C n.s.
Single no-carry traditional	$F_{(4, 169)} = 3.08$	$p = 0.018$	B = C n.s.
Single no-carry columnar	$F_{(4, 157)} = 1.55$	$p = 0.190$ n.s.	B = C n.s.
Single one-carry linear	$F_{(4, 145)} = 9.85$	$p < 0.001$	B > C, $p < 0.001$
Single one-carry traditional	$F_{(5, 155)} = 12.41$	$p < 0.001$	B > C, $p < 0.001$
Single one-carry columnar	$F_{(5, 155)} = 7.15$	$p < 0.001$	B > C, $p < 0.001$
AS no-carry linear	$F_{(4, 165)} = 2.47$	$p = 0.047$	B = C n.s.
AS no-carry traditional	$F_{(4, 168)} = 3.71$	$p = 0.006$	B = C n.s.
AS no-carry columnar	$F_{(4, 173)} = 2.78$	$p = 0.028$	B = C n.s.
AS one-carry linear	$F_{(3, 142)} = 15.50$	$p < 0.001$	B > C, $p < 0.001$
AS one-carry traditional	$F_{(4, 152)} = 14.85$	$p < 0.001$	B > C, $p < 0.001$
AS one-carry columnar	$F_{(4, 145)} = 17.99$	$p < 0.001$	B > C, $p < 0.001$
RLG no-carry linear	$F_{(4, 156)} = 3.73$	$p = 0.006$	B = C n.s.
RLG no-carry traditional	$F_{(4, 162)} = 1.92$	$p = 0.109$ n.s.	B = C n.s.
RLG no-carry columnar	$F_{(4, 164)} = 1.31$	$p = 0.269$ n.s.	B = C n.s.
RLG one-carry linear	$F_{(2, 142)} = 13.25$	$p < 0.001$	B > C, $p < 0.001$
RLG one-carry traditional	$F_{(4, 138)} = 15.74$	$p < 0.001$	B > C, $p < 0.001$
RLG one-carry columnar	$F_{(3, 148)} = 10.71$	$p < 0.001$	B > C, $p < 0.001$

Table 32. Analysis of response times and strategy use with post-hoc comparisons of strategies B and C.

The pattern of results in table 32 is very clear. For no-carry problems strategy B does not produce significantly greater response times than strategy C, but for carry problems the use of strategy B always leads to significantly greater response times. This implies that those participants who demonstrated exclusive, or near exclusive use, of strategy B throughout the experiment were at a disadvantage when faced with carry problems. Another way of stating this is to say that strategy C better supports carry problems, and those that switched to strategy C when faced with carry problems would be at an advantage. However, the analysis of percentage use of strategy C did not demonstrate that this switching is widespread across the task x carry x format levels, so the logic of the experiment is not undermined.

Chapter five has presented evidence for WM involvement in addition and multiplication, and evidence for the potential impact of strategy use on response times. A general discussion of the evidence from all these experiments and the earlier experiments will form the basis of the next chapter.

Chapter Six – general discussion of both series of experimental studies

This chapter will draw together elements from the analyses of all of the experimental studies to identify the following areas of interest: comparison and effectiveness of dual-tasks, evidence of WM involvement in task, carry, format and difficulty interactions, response time and error rates as differential indicators of cognitive load, links to cognitive models, and effects of strategy.

6.1 Comparison of dual-tasks and the effectiveness of dual-task methodology

As previously argued, the studies of Hegarty, Shah and Miyake (2000) and Oberauer et al (2001), make it difficult, if not impossible, to accept significant main effects, of response times or errors, as unambiguous evidence of dual-task interference. Such effects can be assumed to reflect nothing more than simple cost of concurrence – it not only takes longer to do two tasks, but the increase in time might also make the process more error prone. It could be that the two tasks are processed sequentially, but in order to demonstrate that they are in fact competing for the same processing resources, it is necessary to show a performance decrement to both tasks compared to levels of performance when either task is undertaken as a single task. In addition, the secondary interference task must be shown to interact with other experimental factors; it must have a differential impact rather than adding a constant amount, if the task is to be properly accepted as an interference task. WM involvement in the four arithmetic operations, as evidenced by interactions involving the various dual-tasks, will be considered in the next section, but here, the comparison and effectiveness of the secondary interference tasks will be considered. Articulatory suppression, as canonical articulation of the sequence 'A, B, C', failed to demonstrate any decrement in single vs. dual-task articulation rate in any of the experiments. This could indicate that the AS task did not load the phonological loop component of WM to any great extent but, as the pattern of interactions in some experiments provides a different source of evidence for the involvement of phonological processing, this lack of dual-task decrement is assumed to be due to the relative automaticity of articulating 'A, B, C', rather than a complete lack of effect of the AS task. A more complex articulation task might provide greater interference and evidence of secondary task decrement, but there are reasons to be wary of using overly long or complex articulations. Some experimental participants reported an inability to break into the 'A, B, C' sequence, having to complete the sequence before providing the answer to the problem, and although this has relatively little effect with a short articulation sequence, the effect with a longer more complex sequence would be larger. Further experimentation would be required to find the most appropriate articulation task to disrupt the sort of multi-digit arithmetic tasks used in these experiments, but clearly articulatory suppression tasks can differ, and any given articulation task cannot be assumed to load the phonological loop merely because it involves articulation.

Spatial tapping is regarded as a secondary task that loads the visuo-spatial sketchpad in WM, or at least the spatial component of it, but the figure-of-eight spatial sequences in the first experiments were not tapped less accurately under dual-task conditions than under single-task conditions, and there was no evidence that spatial tapping had any impact on the addition task, although there is plenty of experimental evidence that the visual format of arithmetic problems does have an impact. The conclusion is that spatial tapping cannot be used to identify the role of visuo-spatial processing in arithmetic problems, and that manipulations of the visual form of problems are more appropriate to ascertain visuo-spatial involvement.

The three random generation tasks used in the first series of experiments are clearly not equivalent. Random interval generation demonstrated no significant single vs. dual-task performance decrement, and no impact on the addition problems beyond cost of concurrence, while random key-pressing did demonstrate a dual-task performance decrement, but again, no impact on the addition problems beyond cost of concurrence. Random letter generation provided significant single vs. dual-task performance decrement for all experiments, and evidence of impact upon arithmetic problem performance. It is clear that different random generation tasks might not indicate the same aspects of central executive involvement, or might not reflect central executive involvement at all.

The potential for strategy choice to undermine the logic of dual-task methodology is considered in a separate section, but while strategy use was assessed for each arithmetic problem in the second series of experiments, no assessment was made of potential strategy differences in secondary task performance, and it is reasonable that future research should investigate this possibility.

6.2 Evidence of WM involvement as identified by task interactions with carry, format and difficulty

The involvement of the phonological loop component of WM was investigated through the use of articulatory suppression, and the first addition experiment did not find any impact of articulatory suppression on error rates in terms of significant task interactions, but it did produce a three-way task x carry x format interaction in response times such that there was a significant task x carry interaction for problems presented in linear format. This increase in response times for one-carry problems in linear format under AS was tentatively interpreted as a possible role for phonological processing in the re-representation of linear problems into columnar format. However, the second addition experiment also produced a significant task x carry x format interaction in response times, but this demonstrated that AS increased response times for one-carry problems presented in columnar format. This is obviously inconsistent with the tentative interpretation offered for the first addition experiment, and as there were no significant interactions in the error data for the second experiment, there is overall no evidence for phonological processing in the solution of these multi-digit additions.

The subtraction experiment identified no significant interactions under AS for response times and errors, and therefore no evidence that phonological coding is involved in the solution of these particular multi-digit subtraction problems.

The multiplication experiment demonstrated an impact of AS on response times in the form of a significant task x format interaction, and an impact on error rates through a task x difficulty interaction, a task x carry x difficulty interaction, and a task x difficulty x format interaction. The task x format interaction in response times indicated that AS had a differentially greater impact on linear problems, which suggested support for the Trbovitch and LeFevre (2003) claim that horizontal format engages phonological codes, but in the absence of task x carry interactions, this was not interpreted as providing evidence of phonological processing in the carry operation. The task x difficulty interaction in multiplication error scores was modified by the task x carry x difficulty interaction, such that the impact of AS was significantly less only for easy no-carry problems, and this is interpreted as the possibility of these problems being solved directly by *reading* rather than calculation processes. The task x difficulty x format interaction, demonstrated a differentially greater effect of AS on hard problems in linear format, but the interaction is essentially due to differences at the single-task level for linear and columnar formats. Overall, the Single vs. AS task x carry x format interaction was not significant, but if only columnar format results are analysed, then there is a significant task x carry interaction, suggesting phonological processing in the carry operation, and this is the result of the relatively smaller impact of AS on no-carry columnar problems, which again supports the

Trbovitch and LeFevre (2003) claim that vertical / columnar format engages visuo-spatial codes, and would therefore be unlikely to be disrupted by articulatory suppression.

The division experiment provided no task interactions for error scores, but a significant task x carry x format interaction for response times, which was due to differences between linear and columnar format at the two carry levels in single-task conditions. Linear format gave significantly lower response times for no-carry problems, whereas columnar format gave significantly lower response times for carry problems, which is consistent with some sort of *reading* process being used in the solution of the easier no-carry linear problems, and this possibility being removed by articulatory suppression. The involvement of central executive processes was investigated through the use of random letter generation, and the first addition experiment provided a significant task x carry interaction in the RLG error scores possibly indicating that RLG had a differentially larger impact on carry problems, although also possibly interpreted as a 'ceiling effect', thereby providing only limited support for central executive processing in the carry operation.

The second addition experiment also provided a significant task x carry interaction for RLG error scores, but this was modified by a significant task x carry x format interaction, which showed a relatively greater impact of RLG on carry problems in linear format, and this does implicate central executive processing in the carry operation.

The subtraction experiment provided no RLG task interactions for error scores, but a significant task x carry x format interaction for response times, such that RLG appears to reduce the advantage of columnar format for carry problems, and this is not consistent with central executive involvement in the carry operation for these subtraction problems.

The multiplication experiment RLG response times provided a significant task x difficulty interaction, showing that RLG had a differentially greater impact on easy problems, again suggesting the possibility that easy problems in single-task conditions are solved by *reading* processes, and that the articulatory or central executive components of RLG prevent the use of such processes. The multiplication experiment RLG error scores provided a significant task x difficulty x format interaction, but this reflected differences at the single-task level rather than under RLG, such that there were significantly less errors for hard problems in linear format. This interaction, and the lack of task x carry interactions provide no evidence for the involvement of central executive processes in the solution of these multiplication problems.

The division experiment response times provided a significant task x carry x format interaction, and this is the same as the task x carry x format interaction under AS, in that it reflects differences at the single-task level rather than under RLG. There is also a significant task x carry x format interaction in the RLG error scores, but this indicates a significant effect of RLG on no-carry problems in different formats rather than on carry problems, and is therefore not consistent with central executive involvement in the carry operation, although it does indicate a significant carry x format interaction under RLG, such that for no-carry problems linear format gives more errors than either traditional or columnar format, whereas for carry problems linear format gives less errors than either traditional or columnar format.

To summarise, the task interactions demonstrate that the central executive is implicated in addition, and that the phonological loop is implicated in multiplication. However, no evidence was found for the involvement of the central executive or the phonological loop in either subtraction or division, which suggests that subtraction and division problems are solved without recourse to WM resources.

6.3 Effect of carries, format and difficulty independent of task

The main effects of carry, difficulty and format are all assumed to indicate additional processing time due to either increasing the number of operations or transcoding processes, but task independent interactions between these factors are discussed here. The carry operation increased both response times and error scores for every experiment reported here, but in the second experiments carry interactions not involving task were only apparent in the multiplication experiment, and these may have been sub additive effects due to the log transform of response times.

Format provided significant differences in all but two instances of experimental response times, the Single vs. AS subtraction response times were not different for linear vs. columnar format, and the AS vs. RLG division response times were also not significantly different.

In additions, subtractions and multiplications, columnar format generally resulted in faster response times, but this was not always the case, and the potential for linear format supporting fast solution *reading* processes has already been explained.

Divisions were presented in three formats, linear, traditional and columnar, and traditional format provided faster response times throughout, while linear and columnar format differences at the single-task level have already been discussed in relation to the possible involvement of *reading* processes.

Format did not always lead to significant differences in the error scores, and the first addition experiment demonstrated no significant effects of format in any of the five dual-task conditions, while the second experiments demonstrated significant effects of format in AS and RLG error scores for additions, subtractions and divisions, but not for multiplications. The difficulty factor only appeared in the multiplication experiment, and the difficulty x format interaction in error scores has already been discussed in relation to the significant task x difficulty x format interaction, but easy and hard problems also interact with other experimental factors, so that the Seitz and Schumann-Hengsteler (2000) finding that easy problems were not affected by AS and RLG is only partially supported.

6.4 Response times and error scores as differential indicators of cognitive load

On the basis of the first series of addition experiments, the suggestion was made that response times might be a measure of encoding processes – the time taken to extract information from the problem representation, while error scores might provide a more direct measure of calculation processes, or at least central executive involvement in the carry operation. This suggestion was due to three strands of evidence; the lack of appropriate task x carry interactions in the response time data, main effects of task and format for all five dual-task response times but no effect of format for error scores, and RLG error scores providing the only instance of main effect of task and appropriate, albeit ambiguous, task x carry interaction.

Is this suggestion supported in the second series of experiments for all arithmetic operations? The second addition experiment response times indicated that differences in format affected error rates, and this is inconsistent with format being involved with encoding but not with calculation accuracy, for the significant task x carry x format interaction demonstrated not only that RLG interacts with the carry operation, but that this effect is greater for problems in linear format. However, in other respects the second addition experiment is similar to the first addition experiment, with the same lack of appropriate interactions in the response time data, and the suggestion that response times and errors are measuring different aspects of cognitive load is still plausible.

The subtraction experiment showed no interactions under AS, and a task x carry x format interaction only for response times under RLG, which did not indicate disruption to the carry process, and there is nothing in the subtraction experiment to indicate that response times and error rates are measuring different aspects of cognitive load.

The multiplication response time results and error scores are very different, in that there is a main effect of format, but no three-way interactions in AS and RLG response times, whereas in the error scores, there is no main effect of format, but four different three-way interactions, which implicates AS in the carry operation, but only for easy problems, and RLG in format and difficulty interactions. Response times do not implicate phonological processes or central executive processes, whereas error rates do, so the multiplication experiment does support the suggestion that response times and error rates measure different aspects of cognitive load.

The division results are similar for response times and error rates, but there is no evidence for AS or RLG having an effect on the carry process, although RLG does interact with format, and in this respect, the division experiment, like the subtraction experiment, does not support response times and error scores as differential indicators of cognitive load. However, for additions and multiplications, it is the error score interactions that provide the evidence to implicate RLG and AS in the carry operation, and this suggests that error scores are a more direct indicator of calculation processes, whereas response times are not a good measure of calculation processes, but are likely to indicate differences in both encoding and calculation time. The conclusion is that future studies should seek to identify WM involvement through the use of error scores rather than through differences in response times.

6.5 Links to cognitive models of arithmetic

The cognitive models of arithmetic, reviewed in chapter two, suggested particular differences in the processing of arithmetic. The abstract – modular model of McCloskey, Caramazza and Basili (1985) has separate encoding, calculation and output stages, so that external features of the problem, such as differences in presentation format, should not have any effect on the calculation process.

The encoding complex hypothesis – interactive model, Campbell and Clarke (1988), predicts exactly the opposite, in that the surface form of any problem is assumed to have a direct impact on problem solution. In short, this model predicts format effects on the calculation process, such as task x format and carry x format interactions.

The triple-code model of Dehaene and Cohen (1995), proposes different processing routes for subtraction as opposed to multiplication and addition, so that it should be possible to find patients who have impaired subtraction but preserved multiplication and addition, but never patients who have preserved addition but impaired subtraction and multiplication, although as stated in chapter two, this claim has been challenged.

The independent number module proposed by Butterworth (2000), suggests that the solution of arithmetic problems is not dependent on the use of WM resources, and consequently no evidence of WM involvement should be found in experimental studies of WM and arithmetic.

Do the experimental results presented here contradict, or provide support, for any of these models? The answer is that they do. The task x carry x format interactions in the error data demonstrate that linear and columnar format of presentation can both increase and decrease the number of errors depending on the carry level and the type of secondary task, and while a theoretical defence of the abstract-modular model might be possible by claiming increased encoding and output times due to different formats, this would only seem to be capable of providing a constant additive increase in response times and errors, whereas interactions provide evidence of a differential increase in response time and errors. For this reason, it would appear that the interactions are better explained by the encoding complex hypothesis – interactive model, although there are reasons to suggest that this cannot provide the full explanation of the experimental results, for the pattern of interactions are not the same for all arithmetic operations, and do appear to be largely consistent with the predictions of the triple-code model.

Subtractions and divisions show different patterns of interactions to additions and multiplications, and in fact, multiplications are different to additions. It is RLG that disrupts the carry operation in additions, whereas articulatory suppression disrupts the carry operation in multiplications. RLG is a central executive task whereas articulatory suppression disrupts the phonological loop, and it seems possible that this difference arises because access to, or calculation of, addition facts requires central executive processing, while access to rote-learned multiplication facts requires phonological processing. The lack of significant task x carry interactions in the subtraction results is consistent with a different processing route, although there is no direct evidence that subtractions are solved by *strategical quantity manipulations* as the triple-code model suggests.

The situation for division is puzzling, for the lack of appropriate task x carry interactions does not implicate central executive or phonological processing in their solution, and to this extent, they seem similar to subtractions, but while comparisons of magnitude offer a plausible account of the subtraction process, the division process does not seem to be explicable in these terms, and appears to require some sort of recoding operation, e.g. to change divisions to multiplications, but if so, why are the same interactions not evident? Perhaps the multi-digit division problems used here were too easy in comparison to the multiplication problems, with further research possibly benefiting from the introduction of a difficulty factor into the division problems. However, this overlooks the possibility that there is an independent number-processing module as claimed by Butterworth (2000). The solution of subtractions and divisions has to involve some sort of cognitive processing even if this is independent of WM resources, and the experimental results in this study do demonstrate only limited WM involvement in arithmetic, and are to this extent at least, consistent with a partially independent number-processing module.

6.6 Effects of strategy choice and causes of strategy change

In the first series of experiments, the assessment of strategy use was too imprecise, and the classification of participants into five strategy groups, on the basis of verbal protocols and negotiation, did not allow an analysis of changes in strategy with varying experimental levels, or the analysis of strategy response times. The second series of experiments adopted the procedure suggested by Hecht (2002) for the analysis of strategy use following the presentation of each arithmetic problem, and this did allow the analysis of percentage strategy use at each experimental level, and the analysis of strategy response times. The use of different strategies was strongly idiosyncratic; many participants maintained the same strategy throughout each experiment, and this limited the analysis, due to some strategies having insufficient values to provide a meaningful analysis.

The addition experiment provided two strategy choices that did have sufficient values for analysis, strategy C (partitioning), and strategy E (mental carrying) – see appendix G for details. Separate ANOVA tests demonstrated that the percentage use of both strategies was not significantly different over task, carry and format levels, which is important, in that it suggests that strategy choice is not dependent upon WM load. However, strategy C gave significantly longer response times at every experimental level, and although this is not a problem for a repeated-measures design, it demonstrates that the analysis of response time data for different participants on the same experiment cannot be assumed to represent the same underlying processes.

A further problem for the analysis of strategy, is that what is needed is the response times for different strategies undertaken by the same participants, but here, the lack of participants demonstrating a switch in strategy choice at different task, carry and format levels made such analysis impossible, and the possible effects of speed of processing could not be disambiguated from the effect of strategy. Those using the 'faster' strategy may simply have been those who were faster at processing addition problems, and this could potentially provide an explanation of the response time advantage of particular strategies. The subtraction experiment provided three strategies with sufficient values for analysis, strategy B (estimating/correcting), and strategies C and E, as above. Separate ANOVA tests revealed no significant differences in percentage strategy use for strategy C, but strategies B and E did demonstrate significant differences in percentage use at different experimental levels, and if the response times for each strategy are different, then this could potentially undermine the logic of the experiment.

Further analysis demonstrated that on ten of the twelve experimental levels, strategy B did not provide response times significantly different to strategy C, but on ten of the twelve experimental levels, strategy C gave significantly longer response times to strategy E. Given that the effect of speed of processing could again not be disambiguated from effect of strategy, an estimate of the impact of strategy differences was calculated, and this showed that between task and carry levels the impact of strategy is likely to be relatively minor, but where response time differences are smaller, as in the case of format, the strategy differences are potentially large enough to influence results. Further experimental studies are needed to compare different strategy response times for the same participants, and perhaps to obtain sufficiently large samples of participants demonstrating the same strategy use and speed of processing.

The strategy choices for the multiplication and division experiments were slightly different to the ones provided for the addition and subtraction experiments, see appendix H, and although, in the multiplication experiment, the strategies with sufficient values for analysis were still identified as partitioning and mental carrying, they were now labelled strategy B and strategy C. Separate ANOVA tests of strategy B and strategy C identified significant difference in percentage strategy use at different experimental levels, but further analysis of response time differences for these two strategies revealed that they were not significantly different at sixteen of the twenty-four experimental levels. Instances where strategy B provided significantly greater response times than strategy C were for *hard* problems, seven out of eight, under dual-task conditions, seven out of eight, and for the carry operation, six out of eight. Again, interpretation is complicated by the possibility of individual difference in speed of processing, but the suggestion is that strategy B (partitioning) does not provide the same support for the carry operation as provided by strategy C (mental carrying), but as strategy C is essentially the mental representation of borrowing/paying-back in HTU (hundreds, tens, units) columnar format, this is not altogether surprising.

The division experiment also provided strategies B and C as the only ones with sufficient values for analysis, and separate ANOVA tests indicated no significant differences in percentage use of strategy B at different experiment levels, and a significant difference in percentage use of strategy C at only two of the eighteen experimental levels. However, further analysis of strategy response times revealed a very clear pattern of difference, such that strategy B and C response times were not significantly different at any of the nine no-carry experimental levels, but strategy B gave significantly longer response times than strategy C at all nine experimental levels involving the carry operation. This is further evidence for the superiority of strategy C when solving carry problems, and it would seem to indicate that those with a preference for the use of strategy B could significantly improve their response times by switching to strategy C, but as the earlier analysis demonstrates, they generally do not switch strategies.

Summary and conclusions

The assumption that particular secondary tasks disrupt particular components of working memory is questionable. This research has shown that three random generation tasks designed to disrupt the central executive component of WM are not equivalent in effect, and that spatial tapping cannot be assumed to be an overall disruptor of visuo-spatial processing, although it may disrupt spatial processing.

The ability of articulatory suppression to load the phonological loop would appear to depend not simply on the articulation process, but on the complexity and sequencing of the articulation task. Single versus dual-task measurement of secondary task performance decrement can indicate when secondary tasks are competing for resources, but it would appear that a failure to find significant differences in single and dual-task performance does not necessarily indicate that the secondary task is not having an effect.

The articulatory suppression task used in these experiments consistently demonstrated no significant difference in articulation rate in single vs. dual-task conditions, but was shown to have an effect on response times and errors. The explanation for this is that the articulation of 'A, B, C' appears to be relatively automatic for most, if not all, participants, so that although this had some effect on phonological processing, articulation rate did not vary. The possibility for the *automatization* of secondary task procedures and the possibility of using other strategies, e.g. rhythm to aid a tapping sequence, suggests that secondary tasks do not necessarily impact all individuals to the same extent, and this is potentially another source of variation affecting response times and error scores. Comparisons of dual-task studies in cognitive arithmetic are effectively meaningless if the studies do not use the same primary and secondary tasks, and dual-task studies need to demonstrate that the secondary tasks used do in fact impact those components of WM that are intended to be impacted. The involvement of WM cannot be considered to be the same for all four arithmetic operations. Table 33 gives a summary of the components of WM identified with each of the four arithmetic operations in this research.

	Phonological Loop	Central Executive
Addition	No	Yes
Subtraction	No	No
Multiplication	Yes	No
Division	No	No

Table 33. Summary of the involvement of WM components in the four arithmetic operations

The visuo-spatial sketch pad is not included in this summary table because the involvement of this component was not assessed through the use of a separate visuo-spatial dual-task experiment, but through the use of the two levels of the format variable, linear vs. columnar presentation, and the conclusion is that the results of this study support the suggestion of Trbovitch and LeFevre (2003) that linear (horizontal) format engages phonological processes while columnar (vertical format) engages visuo-spatial processes. Although the arithmetic problems used in these experiments were continuously visible for the duration of each problem presentation, the experimental results, and the strategy results, suggest that many participants regularly use a mental representation of arithmetic problems in columnar format, and that this does provide better support for the carry operation, which is consistent with the view expressed in Heathcote (1994) that the VSSP is a *mental blackboard* upon which calculations are represented. The suggestion from the results of the current study is that linear format engages phonological processing because of *reading* processes, partly because the reading process occurs left-to-right, and partly because some problems can be solved directly as they are read without further calculation, especially relatively easy problems without any carry component.

The finding by Seitz and Schumann-Hengsteler (2000) that articulatory suppression and random letter generation had no effect on *easy* multiplication problems, is only partially supported in the multiplication experiment reported in this study, in that AS had no effect on easy no-carry problems, but did have an effect on easy carry problems, whereas RLG had an effect on both easy and hard problems. The results of the addition and multiplication experiments in the current research also support the suggestion that response times and error rates measure different aspects of cognitive load. It is the error score analysis that provided evidence of the task x carry interactions implicating phonological and central executive involvement, and it is suggested that error scores are a more direct measure of calculation processes, whereas response times measure a mixture of encoding time and calculation time, but do not necessarily indicate the impact of secondary tasks on the calculation process.

The evidence of differences in WM involvement for the different arithmetic operations provides some support for the triple-code model of Dehaene and Cohen (1995), although the addition experiment results supports central executive processing while the multiplication results support phonological processing, which is consistent with multiplication facts being stored as rote-learned verbal word frames, but not with addition facts being stored in this way, the suggestion being that addition facts are not accessed, but are calculated. The interactions involving the format variable support Campbell and Clarke's (1988) encoding complex model, and raise doubts about the adequacy of the abstract-modular model of McCloskey et al (1985). The independent number-processing module proposed by Butterworth (2000) is also partially supported by the differences in WM involvement across the four arithmetic operations, for if WM is not involved in the solution of problems requiring a particular arithmetic operation, then what do we use to solve such problems?

The results of the analysis of strategy choice indicate that this is major concern for dual-task studies, or other studies, of arithmetic operations involving the measurement of response times and error rates. There is evidence that relatively few participants change strategy use when faced with different task, carry and format levels, but for those that do, this could have considerable impact upon both response times and errors. The analysis in the current research could not disambiguate the effect of strategy from the possible effect of individual differences in speed of processing, and could only provide broad estimates of the likely effect of strategy on overall response times, but clearly, any studies not assessing and attempting to control for strategy choice response time differences, are likely to provide inadequate and inaccurate results.

References

Adams, J.W. & Hitch, G.J. (1997) Working memory and children's mental addition. *Journal of Experimental Child Psychology*, 67, 21-38.

Anderson, J.R., Reder, L.M. & Lebiere, C. (1996) Working memory: Activation limits on retrieval. *Cognitive Psychology*, 30, 221-256.

Andrade, J. (2001a) An introduction to working memory. In Andrade, J. (Ed). *Working Memory in Perspective.* Hove: Psychology Press. pp.3-30.

Andrade, J. (2001b) The working memory model: Consensus, controversy, and future directions. In Andrade, J. (Ed). *Working Memory in Perspective.* Hove: Psychology Press. pp. 281-310.

Ashcraft, M.H. (1992) Cognitive arithmetic: A review of data and theory. *Cognition*, 44, 75-106.

Ashcraft, M.H. (1995) Cognitive psychology and simple arithmetic: A review and summary of new directions. *Mathematical Cognition*, 1, 3-34.

Ashcraft, M.H & Kirk, E.P. (2001) The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General*, 130, 224-237.

Atkinson, R.C. & Shiffrin, R.M. (1968) Human memory: A proposed system and its control processes. In Spence, K.W. (Ed). *The psychology of learning and motivation: Advances in research and theory*, pp. 89-195, Academic Press.

Baddeley, A.D., (1990) The phonological loop. Extract from *Human Memory.* Erlbaum.

Baddeley, A.D., (1996) Exploring the Central Executive. *The Quarterly Journal of Experimental Psychology*, 49, 5-28.

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

Baddeley, A.D., Emslie, H., Kolodny, J. & Duncan, J. (1998) Random generation and the executive control of working memory. *The Quarterly Journal of Experimental Psychology*, 51, 819-852.

Baddeley, A.D., Gathercole, S. & Papagno, C. (1998) The phonological loop as a language learning device. *Psychological Review*, 105, 158-173.

Baddeley, A.D. & Logie, R.H. (1999) Working memory: The multiple-component model. In Miyake, A. & Shah, P. (1999) *Models of working memory: Mechanisms of active maintenance and executive control.* Cambridge: Cambridge University Press. pp.28-61.

Baddeley, A.D. & Hitch, G.J. (2000) Development of working memory: Should the Pascual-Leone and Baddeley and Hitch models be merged? *Journal of Experimental Child Psychology*, 77, 128-137.

Baddeley, A.D. & Hitch, G.J. (2001) Foreword. In Andrade, J. (Ed). (2001) *Working Memory in Perspective.* Hove: Psychology Press.

Baroody, A.J. (1999) The roles of estimation and the commutativity principle in the development of third grader's mental multiplication. *Journal of Experimental Child Psychology,* 74, 157-193.

Berch, D.B., Foley, E.J., Hill, R. & Ryan, P.M. (1999) Extracting parity and magnitude from Arabic numerals: Developmental changes in number processing and mental representation. *Journal of Experimental Child Psychology*, 74, 286-308.

Bisanz, J. (1999) The development of mathematical cognition: Arithmetic. *Journal of Experimental Child Psychology*, 74, 153-156.

Bisanz, J. (1999) The development of mathematical cognition: Numerical processes and concepts. *Journal of Experimental Child Psychology*, 74, 283-285.

Blanken, G., Dorn, M. & Sinn, H. (1997) Inversion errors in Arabic number reading: Is there a non-semantic route? *Brain and Cognition*, 34, 404-423.

Brandimonte, M.A., Hitch, G.J. & Bishop, D.V.M. (1992) Verbal recoding of visual stimuli impairs mental image transformations. *Memory and Cognition*, 20, 449-455.

Brenner, M.E., Herman, S., Ho, H-Z. & Zimmer, J.M. (1999) Cross-national comparison of representational competence. *Journal for Research in Mathematics Education*, 30, 541-558.

Bruyer, R. & Scailquin, J-C. (1998) The visuo-spatial sketchpad for mental images: Testing the multi-component model of working memory. *Acta Psychologica*, 98, 17-36.

Bryant, P., Christie, C. & Rendu, A. (1999) Children's understanding of the relation between addition and subtraction: Inversion, identity and decomposition. *Journal of Experimental Child Psychology, 74, 194-212.*

Bull, R. & Johnston, R.S. (1997) Children's arithmetical difficulties: Contributions from processing speed, item identification, and short term memory. *Journal of Experimental Child Psychology*, 65, 1-24.

Bull, R., Johnston, R S. & Roy, J.A. (1999) Exploring the roles of the visual-spatial sketch pad and the central executive in children's arithmetic skills: Views from cognition and developmental neuropsychology. *Developmental Neuropsychology*, 15, 421-442.

Butterworth, B. (2000) *The Mathematical Brain.* London: Papermac.

Butterworth, B., Cipolotti, L. & Warrington, E.K. (1996) Short-term memory impairment and arithmetical ability. *The Quarterly Journal of Experimental Psychology*, 49, 251-261.

Butterworth, B., Zorzi, M., Girelli, L. & Jonckheere, A.R. (2001) Storage and retrieval of addition facts: The role of number comparison. *The Quarterly Journal of Experimental Psychology*, 54, 1005-1029.

Campbell, J.I.D. (1994) Architectures for numerical cognition. *Cognition*, 53, 1-44.

Campbell, J.I.D. (1998) Linguistic influences in cognitive arithmetic: Comment on Noel, Fias and Brysbaert (1997). *Cognition*, 67, 353-364.

Campbell, J.I.D. (1999) The surface form x problem size interaction in cognitive arithmetic: Evidence against an encoding locus. *Cognition*, 70, B25-B33.

Campbell, J.I.D. & Fugelsang, J. (2001) Strategy choice for arithmetic verification: Effects of numerical surface form. *Cognition*, 80, B21-B30.

Campbell, J.I.D. & Xue, Q. (2001) Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General,* 130, 299-315.

Chincotta, D., Underwood, G., Abd Ghani, K., Papadopoulou, E. & Wresinski, M. (1999) Memory span for Arabic numerals and digit words: Evidence for a limited capacity visuo-spatial storage system. *The Quarterly Journal of Experimental Psychology*, 52, 325-351.

Chocon, F., Cohen, L., van de Moortele, P.F. & Dehaene, S. (1999) Differential contributions of the left and right inferior parietal lobules to number processing. *Journal of Cognitive Neuroscience*, 11, 617-630.

Cohen, L. & Dehaene, S. (2000) Calculating without reading: Unexpected residual abilities in pure alexia. *Cognitive Neuropsychology*, 17, 563-583.

Cohen, L., Dehaene, S., Chochon, F., Lehericy, S. & Naccache, L. (2000) Language and calculation within the parietal lobe: A combined cognitive, anatomical and fMRI study. *Neuropsychologia*, 38, 1426-1440.

Conway, A.R.A. & Engle, R. W. (1994) Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General*, 123, 354-373.

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

Dehaene, S. (1992) Varieties of numerical abilities. *Cognition*, 44, 1-42.

Dehaene, S. (1997) *The Number Sense.* London: Penguin.

Dehaene, S. & Cohen, L. (1995) Towards an anatomical and functional model of number processing. *Current Psychology of Cognition*, 18, 83-120.

Dehaene, S., Tzourio, N., Frak, F., Raynaud, L., Cohen, L., Mehler, J. & Mazoyer, B. (1996) Cerebral activations during number multiplication and comparison. *Neuropsychologia*, 34, 1097-1106.

Dehaene, S., Molko, N., Cohen, L. & Wilson, A.J. (2004) Arithmetic and the brain. *Current Opinion in Neurobiology,* 14, 218-224.

De Rammelaere, S., Stuyven, E. & Vandierendonck, A. (2001) Verifying simple arithmetic sums and products: Are the phonological loop and the central executive involved? *Memory and Cognition*, 29, 267-273.

DeStefano, D. & Lefevre, J-A. (2004) The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology*, 16, 353-386.

Dixon, J.A., Deets, J.K. & Bangert, A. (2001) The representations of the arithmetic functions include functional relationships. *Memory and Cognition*, 29, 462-477.

Ellis, A.W. (1993) *Reading, Writing and Dyslexia: A cognitive analysis* 2nd Edition, Hove: Lawrence Erlbaum Associates.

Engle, R.W., Kane, M.J. & Tuholski, S.W. (1999) Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In Miyake, A & Shah, P. (Eds), *Models of working memory: Mechanisms of active maintenance and executive control.* Cambridge: Cambridge University Press.

Evans, F.J. (1978) Monitoring attention deployment by random number generation: An index to measure subjective randomness. *Bulletin of the Psychonomic Society*, 12, 35-38.

Ehrenstein, A., Schweikert, R., Sangsup, C. & Proctor, R.W. (1997) Scheduling processes in working memory: Instructions control the order of memory search and mental arithmetic. *The Quarterly Journal of Experimental Psychology*, 50, 766-802.

Ericsson, K.A. & Kintsch, W. (1995) Long-term working memory. *Psychological Review*, 102, 211-245.

Faul, F. & Erdfelder, E. (1992) GPOWER: A priori, post-hoc and compromise power analyses for MS-DOS [*Computer program*] Bonn, FRG: Bonn University, Department of Psychology.

Furst, A.J. & Hitch G.J. (2000) Executive and phonological components of mental arithmetic. *Memory and Cognition*, 28, 774-782.

Fuson, K.C., Wearne, D., Hiebert, J.C., Murray, H.G., Kuman, P.G., Oliver, A.I., Carpenter, T.P. & Fenema, E. (1997) Children's conceptual structures for multidigit numbers and multidigit addition and subtraction. *Journal for Research in Mathematics Education*, 28, 130-163.

Gathercole, S.E. (1999) Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*. 3, 410-418.

Geary, D.C., Hoard, M.K. & Hamson, C.O. (1999) Numerical and arithmetical cognition: Patterns of function and deficits in children at high risk for a mathematical disability. *Journal of Experimental Child Psychology*, 74, 213-239.

Geary, D.C., Hamson, C.O. & Hoard, M.K. (2000) Numerical and arithmetical cognition: A longitudinal study of process and concept deficits in children with learning disability. *Journal of Experimental Child Psychology*, 77, 236-263.

Geary, D.C., Saults, S.J., Liu, F. & Hoard, M.K. (2000) Sex differences in spatial cognition, computational fluency, and arithmetical reasoning. *Journal of Experimental Child Psychology*, 77, 337-353.

Gobet, F. (2000) Some shortcomings of long-term working memory. *The British Journal of Psychology*, 91, 551-570.

Goolkasian, P. & Foos, P.W. (2002) Presentation format and its effect on working memory. *Memory and Cognition*, 30, 1096-1105.

Graham, T.A. (1999) The role of gestures in children's learning to count. *Journal of Experimental Child Psychology*, 74, 333-355.

Gregory, R.L. (1961) The brain as an engineering problem. In O.L Zangwill & W.H. Thorpe (Eds.), *Current Problems in Animal Behaviour*. Cambridge, UK: Cambridge University Press. pp. 307-330, cited in Andrade, J. (Ed). *Working Memory in Perspective*. Hove: Psychology Press. p.17.

Heathcote, D. (1994) The role of visuo-spatial working memory in the mental addition of multi-digit addends. *Current Psychology of Cognition*, 13, 207-245.

Hecht, S.A. (2002) Counting on working memory in simple arithmetic when counting is used for problem solving. *Memory and Cognition*, 30, 447-455.

Hegarty, M., Shah, P. & Miyake, A. (2000) Constraints on using the dual-task methodology to specify the degree of central executive involvement in cognitive tasks. *Memory and Cognition*, 28, 376-385.

Hitch, G.J., Towse, J.N. & Hutton, U. (2001) What limits children's working memory span? Theoretical accounts and applications for scholastic development. *Journal of Experimental Psychology: General*, 130, 184-198.

Howell, D.C. (1997) *Statistical Methods for Psychology* (4th Ed.). Belmont. Duxbury Press.

Jones, D.M. (1999) The cognitive psychology of auditory distraction: The 1997 BPS Broadbent Lecture. *British Journal of Psychology*, 90, 167-187.

Kail, R. & Hall, L.K. (2001) Distinguishing short-term memory from working memory. *Memory and Cognition*, 29, 1-9.

Kane, M.J., Bleckley, M.K., Conway, A.R.A. & Engle, R.W. (2001) A controlled attention view of working memory capacity. *Journal of Experimental Psychology: General*, 130, 169-183.

Kellogg, J.S., Hopko, D.R. & Ashcraft, M.H. (1999) The effect of time pressure on arithmetic performance. *Journal of Anxiety Disorders*, 13, 591-600.

Kemps, E. (1999) Effects of complexity on visuo-spatial working memory. *European Journal of Cognitive Psychology*, 11, 335-356.

Kemps, E., De Rammelaere, S. & Desmet, T. (2000) The development of working memory: Exploring the complementarity of two models. *Journal of Experimental Child Psychology*, 77, 89-109.

Kotovsky, K., Hayes, J.R. and Simon, H.A. (1985) Why are some problems hard? Evidence from the Tower of Hanoi. *Cognitive Psychology*, 17, 248-294.

Larsen, J.D. & Baddeley, A. (2003) Disruption of verbal STM by irrelevant speech, articulatory suppression, and manual tapping: Do they have a common source? *The Quarterly Journal of Experimental Psychology*, 56, 1249-1268.

LeBlanc, M.D. & Weber-Russell, S. (1996) Text integration and mathematical connections: A computer model of arithmetic word problem solving. *Cognitive Science*, 20, 357-408.

Lee, K-M. & Kang, S-Y. (2002) Arithmetic operation and working memory: Differential suppression in dual tasks. *Cognition*, 83, B63-B68.

Lehto, J. (1996) Are executive function tests dependent on working memory capacity. *The Quarterly Journal of Experimental Psychology*, 49, 29-50.

Lemaire, P., Abdi, H. & Fayol, M. (1996) The role of working memory resources in simple cognitive arithmetic. *European Journal of Cognitive Psychology*, 8, 73-103.

Liebeck, P. (1988) *How Children Learn Mathematics*. London: Penguin.

Logie, R.H. (1995) *Visuo-Spatial Working Memory*. Hove: Lawrence Erlbaum Associates.

Logie, R.H. & Helstrup, T. (1999) Introduction to special edition on mental imagery. *European Journal of Cognitive Psychology*, 11, 289-293.

Logie, R.H., Gilhooly, K.J. & Wynn, V. (1994) Counting on working memory in arithmetic problem solving. *Memory and Cognition*, 22, 395-410.

Lucangeli, D., Tressoldi, P.E., Bendotti, M. Bonanomi, M. & Siegel, L.S. (2003) Effective strategies for mental and written arithmetic calculation from the third to the fifth grade. *Educational Psychology*, 23, 507-520.

Macken, W., Tremblay, S., Alford, D. & Jones, D.M. (1999) Attentional selectivity in short-term memory: Similarity of process, not similarity of content, determines disruption. *International Journal of Psychology*, 34, 322-327.

Mathewson, J.H. (1998) Visual-spatial thinking: An aspect of science overlooked by educators. *Science Education*, 83, 33-54.

May, J. (2001) Specifying the central executive may require complexity. In Andrade, J. (Ed). *Working Memory in Perspective*. Hove: Psychology Press. pp. 261-278.

McCloskey, M. (1992) Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*, 44, 107-157.

McConnell, J. & Quinn, J.G. (2000) Interference in visual working memory. *The Quarterly Journal of Experimental Psychology*, 53, 53-67.

McLean, J.F. & Hitch, G.J. (1999) Working memory impairments in children with specific arithmetic learning difficulties. *Journal of Experimental Child Psychology*, 74, 240-260.

McNamara, D.S. & Scott, J.L. (2001) Working memory capacity and strategy use. *Memory and Cognition*, 29, 10-17.

Miura, I.T., Okamoto, Y., Vlahovic-Stetic, V., Chungsoon, C.K. & Jong Hye, H. (1999) Language supports for children's understanding of numerical fractions: Cross-national comparisons. *Journal of Experimental Child Psychology*, 74, 356-365.

Miyake, A. (2001) Individual differences in working memory: Introduction to the special section. *Journal of Experimental Psychology: General*, 130, 163-168.

Miyake, A. & Shah, P. (1999) *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge: Cambridge University Press.

Miyake, A., Friedman, N.P., Emerson, M.J. & Witzki, A.H. (2000) The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49-100.

Noel, M-P. & Seron, X. (1992) Arabic number reading deficit: A single case study. *Cognitive Neuropsychology*, 10, 317-339.

Noel, M-P., Fias, W. & Brysbaert, M. (1997) About the influence of the presentation format on arithmetic fact retrieval processes. *Cognition*, 63, 335-374.

Noel, M-P., Robert, A. & Brysbaert, M. (1998) Does language really matter when doing arithmetic? Reply to Campbell (1998). *Cognition*, 67, 365-373.

Noel, M-P., Desert, M., Aubrun, A. & Seron, X. (2001) Involvement of short-term memory in complex mental calculation. *Memory and Cognition*, 29, 34-42.

Norman, D.A. & Shallice, T. (1980) Attention and action: Willed and automatic control of behaviour. *Technical Report 8006, Personnel and Training Research Programs,* Office of Naval Research (Code 458). Arlington. Virginia.

Nunes, T. & Bryant, P. (1996) *Children doing mathematics*. Blackwell.

Oberauer, K., Demmrich, A., Mayr, U. & Kliegl, R. (2001) Dissociation retention and access in working memory: An age-comparative study of mental arithmetic. *Memory and Cognition*, 29, 18-33.

Pascual-Leone, J. (2000) Reflections on working memory: Are the two models complementary? *Journal of Experimental Child Psychology*, 77, 138-154.

Pelizzon, L., Brandimonte, M.A. & Favretto, A. (1999) Imagery and recognition: Dissociable measures of memory? *European Journal of Cognitive Psychology*, 11, 429-423.

Pesenti, M., Thioux, M., Seron, X. & De Volder, A. (2000) Neuroanatomical substrates of Arabic number processing, numerical comparison, and simple addition: A PET study. *Journal of Cognitive Neuroscience*, 12, 461-479.

Quinn, J.G. & McConnell, J. (1999) Manipulation of interference in the passive visual store. *European Journal of Cognitive Psychology*, 11, 373-389.

Reichle, E.D., Carpenter, P.A. & Just, M.A. (2000) The neural basis of strategy and skill in sentence-picture verification. *Cognitive Psychology*, 40, 261-295.

Rickard, T.C., Romero, S.G., Basso, G., Wharton, C., Flitman, S. & Grafman, J. (2000) The calculating brain: An fMRI Study. *Neuropsychologia*, 38, 325-335.

Roder, B. & Rosler, F. (1998) Visual input does not facilitate the scanning of mental images. *Journal of Mental Imagery*, 22, 165-168.

Roussel, J-L., Fayol, M. & Barroulet, P. (2002) Procedural vs. direct retrieval strategies in arithmetic: A comparison between additive and multiplicative problem solving. *European Journal of Cognitive Psychology*, 14, 61-104.

Schunn, C.D., Lovett, M.C. & Reder, L.M. (2001) Awareness and working memory in strategy adaptivity. *Memory and Cognition*, 29, 254-266.

Seitz, K. & Schumann-Hengsteler, R. (2000) Mental multiplication and working memory. *European Journal of Cognitive Psychology*, 12, 552-570.

Shallice, T. & Burgess, P. (1996) The domain of supervisory processes and temporal organization of behaviour. *Philosophical Transactions,* Royal Society London, B, 1405-1412.

Shanahan, T. & Lefevre, J-A. (2003) Addition strategies: An algorithm for solving multi-digit problems. Poster presented at *BASICS,* Banff.

Shrager, J. & Siegler, R.S. (1998) SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science,* 9, 405-410.

Siegler, R.S. & Jenkins, E.A. (1989) *How children discover new strategies*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Siegler, R.S. & Shipley, F. (1995) Variation, selection and cognitive change. In G. Halford and T. Simon (Eds.) *Developing cognitive competence: New approaches to process modelling* (pp.31-76) Hillsdale NJ: Erlbaum.

Siegler, R.S. & Lemaire, P. (1997) Older and younger adults' strategy choice in multiplication: Testing predictions of ASCM using the choice/no-choice method. *Journal of Experimental Psychology: General,* 126, 71-92.

Smith, E.E. & Jonides, J. (1997) Working memory: A view from neuroimaging. *Cognitive Psychology*, 33, 5-42.

Smyth, M.M., Collins, A.F., Morris, P.E. & Levy, P. (1994) *Cognition in Action.* Ch.6. Doing mental arithmetic: Holding information and operations for a short time. Hove: Erlbaum.

Tabachnick, B.G. & Fidell, L.S. (2001) *Multivariate Statistics* (4^{th} Ed.). Boston. Allyn and Bacon.

Temple, C.M. & Sherwood, S. (2002) Representation and retrieval of arithmetic facts: Developmental difficulties. *The Quarterly Journal of Experimental Psychology*, 55, 733-752.

Thompson, I. (2000) Understanding and assessing children's mathematical thinking through mental calculation. *[ICME (9)] International Conference on Mathematics Education: Tokyo*. http://www.nku.edu/~sheffield/ithompson.html.

Towse, J.N. (1998) On random generation and the central executive of working memory. *British Journal of Psychology*, 89, 77-102.

Towse, J.N. & Houston-Price, C.M.T. (2001) Combining representations in working memory: A brief report. *British Journal of Developmental Psychology*, 19, 319-324.

Trbovich, P.L. & Lefevre, J-A. (2003) Phonological and visual working memory in mental addition. *Memory and Cognition*, 33, 738-745.

Tuholski, S.W., Engle, R.W. & Bayliss, G.C. (2001) Individual differences in working memory capacity and enumeration. *Memory and Cognition*, 29, 484-492.

Vandierendonck, A. (2000) Analysing human random time generation behaviour: A methodology and a computer program. *Behaviour Research Methods, Instruments and Computers*, 32, 555-565.

Vandierendonck, A., De Vooght, G. & Van der Goten, K. (1998) Does random time interval generation interfere with working memory executive functions? *European Journal of Cognitive Psychology*, 10, 413-442.

Van Harskamp, N.J. & Cipolotti, L. (2001) Selective impairments for addition, subtraction and multiplication – implications for organisation of arithmetic Facts. *Cortex*, 37, 363-388.

Vecchi, T. & Cornoldi, C. (1999) Passive storage and active manipulation in visuo-spatial working memory: Further evidence from the study of age differences. *European Journal of Cognitive Psychology*, 11, 391-406.

Whalen, J., McCloskey, M, Lindemann, M. & Bouton, G. (2002) Representing arithmetic table facts in memory: Evidence from acquired impairments. *Cognitive Neuropsychology*, 19, 505-522.

Whetstone, T. (1998) The representation of arithmetic facts in memory: Results from retraining a brain-damaged patient. *Brain and Cognition*, 36, 290-309.

Zago, L., Pesenti, M., Mellet, E., Crivello, F., Mazoyer, B. & Tzuro-Mazoyer, N. (2001) Neural correlates of simple and complex mental calculation. *Neuroimage*, 13, 314-327.

Appendix A. Addition problems

N	1zcl	2zcl	1zcc	2zcc	1ocl	2ocl	1occ	2occ	1tcl	2tcl	1tcc	2tcc
1	123+34 (157)	125+32 (157)	126+42 (168)	125+43 (168)	129+62 (191)	139+52 (191)	134+57 (191)	135+56 (191)	123+78 (201)	136+65 (201)	127+74 (201)	132+69 (201)
2	134+63 (197)	132+65 (197)	134+43 (177)	135+42 (177)	128+69 (197)	129+68 (197)	139+58 (197)	138+59 (197)	124+89 (213)	137+76 (213)	129+84 (213)	128+85 (213)
3	127+32 (159)	124+35 (159)	123+56 (179)	136+43 (179)	128+34 (162)	125+37 (162)	127+35 (162)	123+39 (162)	137+79 (216)	129+87 (216)	127+89 (216)	147+69 (216)
4	125+42 (167)	124+43 (167)	135+53 (188)	123+65 (188)	123+48 (171)	125+46 (171)	124+47 (171)	129+42 (171)	138+62 (200)	136+64 (200)	126+74 (200)	124+76 (200)
5	125+43 (168)	123+45 (168)	127+62 (189)	125+64 (189)	138+53 (191)	132+59 (191)	127+64 (191)	128+63 (191)	123+89 (212)	127+85 (212)	129+83 (212)	137+75 (212)
6	147+52 (199)	124+75 (199)	142+54 (196)	134+62 (196)	125+59 (184)	128+56 (184)	138+46 184)	139+45 (184)	137+89 (226)	139+87 (226)	169+57 (226)	167+59 (226)
7	126+72 (198)	145+53 (198)	132+65 (197)	125+72 (197)	126+49 (175)	128+47 (175)	127+48 (175)	129+46 (175)	136+75 (211)	127+84 (211)	124+87 (211)	125+86 (211)
8	134+62 (196)	132+64 (196)	123+75 (198)	136+62 (198)	135+48 (183)	124+59 (183)	126+57 (183)	136+47 (183)	123+87 (210)	136+74 (210)	132+78 (210)	128+82 (210)
9	145+52 (197)	124+73 (197)	125+74 (199)	126+73 (199)	128+64 (192)	127+65 (192)	135+57 (192)	134+58 (192)	128+97 (225)	136+89 (225)	127+98 (225)	129+96 (225)
10	126+43 (169)	124+45 (169)	125+53 (178)	132+46 (178)	127+69 (196)	139+57 (196)	129+67 (196)	137+59 (196)	137+84 (221)	136+85 (221)	127+94 (221)	124+97 (221)
11	132+43 (175)	123+52 (175)	124+62 (186)	134+52 (186)	123+58 (181)	125+56 (181)	127+54 (181)	124+57 (181)	137+96 (233)	135+98 (233)	139+94 (233)	146+87 (233)
12	134+42 (176)	124+52 (176)	125+63 (188)	132+56 (188)	125+47 (172)	129+43 (172)	123+49 (172)	127+45 (172)	146+85 (231)	148+83 (231)	147+84 (231)	137+94 (231)
13	123+72 (195)	132+63 (195)	123+35 (158)	126+32 (158)	125+39 (164)	126+38 (164)	128+36 (164)	129+35 (164)	137+85 (222)	128+94 (222)	125+97 (222)	124+98 (222)
14	137+52 (189)	136+53 (189)	135+63 (198)	146+52 (198)	125+49 (174)	126+48 (174)	128+46 (174)	129+45 (174)	137+95 (232)	135+97 (232)	145+87 (232)	147+85 (232)
15	134+45 (179)	137+42 (179)	142+57 (199)	132+67 (199)	126+59 (185)	127+58 (185)	139+46 (185)	136+49 (185)	147+68 (215)	148+67 (215)	167+48 (215)	168+47 (215)
16	123+54 (177)	124+53 (177)	123+34 (157)	125+32 (157)	136+57 (193)	134+59 (193)	139+54 (193)	135+58 (193)	149+68 (217)	169+48 (217)	128+89 (217)	168+49 (217)
17	132+53 (185)	123+62 (185)	125+34 (159)	127+32 (159)	125+38 (163)	127+36 (163)	128+35 (163)	126+37 (163)	123+79 (202)	127+75 (202)	124+78 (202)	128+74 (202)
18	123+76 (199)	146+53 (199)	123+45 (168)	126+42 (168)	126+69 (195)	127+68 (195)	136+59 (195)	139+56 (195)	137+83 (220)	136+84 (220)	134+86 (220)	146+74 (220)
19	135+43 (178)	132+46 (178)	127+42 (169)	124+45 (169)	125+48 (173)	128+45 (173)	126+47 (173)	127+46 (173)	127+76 (203)	136+67 (203)	128+75 (203)	125+78 (203)
20	134+52 (186)	132+54 (186)	124+65 (189)	132+57 (189)	136+58 (194)	138+56 (194)	129+65 (194)	125+69 (194)	125+98 (223)	128+95 (223)	137+86 (223)	136+87 (223)
21	125+62 (187)	123+64 (187)	135+43 (178)	136+42 (178)	139+43 (182)	134+48 (182)	125+57 (182)	124+58 (182)	136+69 (205)	127+78 (205)	137+68 (205)	138+67 (205)
22	123+36 (159)	125+34 (159)	134+42 (176)	124+52 (176)	126+39 (165)	127+38 (165)	128+37 (165)	129+36 (165)	126+98 (224)	135+89 (224)	128+96 (224)	146+78 (224)
23	143+52 (195)	142+53 (195)	135+52 (187)	134+53 (187)	127+59 (186)	129+57 (186)	139+47 (186)	137+49 (186)	136+68 (204)	128+76 (204)	125+79 (204)	126+78 (204)
24	123+74 (197)	143+54 (197)	126+62 (188)	136+52 (188)	128+59 (187)	129+58 (187)	139+48 (187)	138+49 (187)	159+68 (227)	158+69 (227)	169+58 (227)	168+59 (227)
25	123+46 (169)	127+42 (169)	132+53 (185)	123+62 (185)	123+57 (180)	127+53 (180)	134+46 (180)	137+43 (180)	139+67 (206)	149+57 (206)	147+59 (206)	157+49 (206)
26	126+53 (179)	132+47 (179)	127+52 (179)	125+54 (179)	123+67 (190)	127+63 (190)	128+62 (190)	132+58 (190)	148+82 (230)	143+87 (230)	183+47 (230)	147+83 (230)
27	132+45 (177)	125+52 (177)	135+42 (177)	124+53 (177)	123+47 (170)	124+46 (170)	127+43 (170)	128+42 (170)	149+58 (207)	139+68 (207)	148+59 (207)	138+69 (207)
28	142+56 (198)	125+73 (198)	124+72 (196)	132+64 (196)	137+56 (193)	125+68 (193)	128+65 (193)	129+64 (193)	149+85 (234)	148+86 (234)	159+75 (234)	158+76 (234)
29	134+53 (187)	124+63 (187)	132+43 (175)	123+52 (175)	127+39 (166)	129+37 (166)	128+39 (167)	129+38 (167)	139+75 (214)	146+68 (214)	149+65 (214)	168+46 (214)
30	126+52 (178)	136+42 (178)	145+54 (199)	143+56 (199)	127+49 (176)	129+47 (176)	129+48 (177)	128+49 (177)	146+89 (235)	148+87 (235)	157+78 (235)	178+57 (235)

Appendix B.

Counter-Balanced Task Order - First Series of Experiments

	M	D	D	M	M	D	D	M	M	D	D	M N
S1	A1M	B1D	A2D	B2M	A3M	B3D	A4D	B4M	A5M	B5D	A6D	B6M
S2	B2M	A2D	B3D	A3M	B4M	A4D	B5D	A5M	B6M	A6D	B1D	A1M
S3	A3M	B3D	A4D	B4M	A5M	B5D	A6D	B6M	A1M	B1D	A2D	B2M
S4	B4M	A4D	B5D	A5M	B6M	A6D	B1D	A1M	B2M	A2D	B3D	A3M
S5	A5M	B5D	A6D	B6M	A1M	B1D	A2D	B2M	A3M	B3D	A4D	B4M
S6	B6M	A6D	B1D	A1M	B2M	A2D	B3D	A3M	B4M	A4D	B5D	A5M
S7	B1M	A1D	B2D	A2M	B3M	A3D	B4D	A4M	B5M	A5D	B6D	A6M
S8	A2M	B2D	A3D	B3M	A4M	B4D	A5D	B5M	A6M	B6D	A1D	B1M
S9	B3M	A3D	B4D	A4M	B5M	A5D	B6D	A6M	B1M	A1D	B2D	A2M
S10	A4M	B4D	A5D	B5M	A6M	B6D	A1D	B1M	A2M	B2D	A3D	B3M

Twelve Blocks of Equivalent Trials (A1 – A6, B1 – B6)
M = Presented as 'Maths Alone'
D = Presented as 'Dual – Task'

Appendix C. Subtraction problems

	szcl	szcc	socl	socc	aszcl	aszcc	asocl	asocc	rzcl	rzcc	rocl	rocc
1	159-36	156-33	172-45	172-45	158-35	155-32	181-54	185-58	157-34	157-34	195-68	194-67
2	158-34	159-35	191-67	192-26	176-52	178-54	172-48	173-49	156-32	159-35	162-38	163-29
3	157-35	156-34	184-59	183-58	167-45	168-46	164-39	173-48	154-32	158-36	174-49	162-37
4	156-24	157-25	163-37	162-36	175-43	174-42	193-57	194-58	196-64	197-65	183-47	182-56
5	169-36	165-32	161-29	161-29	178-45	179-45	161-39	171-39	167-34	168-35	181-49	181-49
6	179-53	198-72	162-25	164-27	158-32	168-42	172-35	174-37	199-73	189-63	185-48	182-45
7	197-34	198-35	192-58	193-59	187-26	187-24	182-48	183-49	186-23	188-25	172-38	173-39
8	187-53	186-52	184-49	183-48	176-42	177-43	163-28	162-27	196-62	197-63	164-29	161-26
9	177-42	178-43	173-37	172-48	187-52	187-53	193-57	192-56	198-63	199-64	192-56	194-58
10	178-42	179-43	191-49	191-49	189-53	188-52	181-39	181-39	198-62	199-63	171-29	171-29
11	197-54	196-53	172-29	171-28	187-54	186-53	192-49	191-48	167-24	168-25	181-38	182-39
12	195-53	196-54	182-38	183-39	187-45	188-46	172-28	171-27	184-42	185-43	183-39	181-37
13	176-32	178-34	192-47	193-48	186-42	187-43	183-38	182-37	198-56	197-53	173-28	172-47
14	199-54	198-53	161-29	161-29	179-34	177-32	171-39	171-39	168-23	169-24	181-39	181-47
15	177-52	178-53	192-39	191-38	187-62	188-63	181-28	182-29	197-72	198-73	182-29	181-28
16	198-46	197-45	182-28	183-29	178-26	179-27	193-39	192-38	188-36	189-37	181-37	181-36
17	199-46	198-45	183-28	182-27	189-36	187-34	192-37	193-38	179-26	178-25	191-36	191-36
18	186-32	189-45	151-29	151-29	178-34	179-25	161-39	161-39	197-43	196-42	171-49	171-49
19	197-42	198-43	192-35	191-24	179-24	178-23	184-27	185-28	187-32	189-34	182-25	183-26
20	198-34	199-37	182-35	185-38	189-27	188-36	174-27	173-26	187-25	186-24	194-47	195-48

Appendix D. Multiplication problems

Single Mults	CEL	CEC	CHL	CHC	NCEL	NCEC	NCHL	NCHC
1	126 x 2	125 x 2	26 x 12	25 x 12	123 x 2	124 x 2	23 x 12	24 x 12
2	119 x 2	118 x 2	19 x 12	18 x 12	131 x 2	113 x 2	31 x 12	13 x 12
3	128 x 3	118 x 3	28 x 13	18 x 13	121 x 3	123 x 3	21 x 13	23 x 13
4	239 x 2	245 x 2	39 x 22	45 x 22	231 x 2	221 x 2	31 x 22	21 x 22
5	225 x 3	227 x 3	25 x 23	27 x 23	211 x 3	221 x 3	11 x 23	21 x 23
6	316 x 2	317 x 2	16 x 32	17 x 32	312 x 2	321 x 2	12 x 32	21 x 32
7	325 x 3	329 x 3	25 x 33	29 x 33	312 x 3	311 x 3	12 x 33	31 x 31
8	345 x 2	329 x 2	45 x 12	29 x 12	332 x 2	331 x 2	32 x 21	31 x 22
9	226 x 2	237 x 2	26 x 22	22 x 37	212 x 2	214 x 2	12 x 22	21 x 42
AS Mults	CEL	CEC	CHL	CHC	NCEL	NCEC	NCHL	NCHC
1	127 x 2	128 x 2	27 x 12	28 x 12	134 x 2	132 x 2	32 x 12	32 x 12
2	118 x 2	119 x 2	18 x 12	19 x 12	113 x 2	121 x 2	13 x 12	21 x 12
3	116 x 3	126 x 3	16 x 13	26 x 13	112 x 3	121 x 3	23 x 13	21 x 13
4	228 x 2	236 x 2	28 x 22	36 x 22	221 x 2	231 x 2	21 x 22	31 x 22
5	217 x 3	228 x 3	17 x 23	28 x 23	213 x 3	212 x 3	13 x 23	12 x 23
6	317 x 2	316 x 2	17 x 32	16 x 32	331 x 2	311 x 2	31 x 32	11 x 32
7	317 x 3	328 x 3	17 x 33	28 x 33	321 x 3	312 x 3	21 x 33	33 x 12
8	336 x 2	337 x 2	36 x 12	37 x 12	321 x 2	323 x 2	21 x 32	23 x 12
9	227 x 2	225 x 2	27 x 22	25 x 22	213 x 2	212 x 2	13 x 22	22 x 12
RLG Mults	CEL	CEC	CHL	CHC	NCEL	NCEC	NCHL	NCHC
1	129 x 2	126 x 2	29 x 12	26 x 12	141 x 2	141 x 2	41 x 12	41 x 12
2	117 x 2	136 x 2	17 x 12	36 x 12	114 x 2	142 x 2	41 x 11	41 x 11
3	127 x 3	129 x 3	27 x 13	29 x 13	123 x 3	113 x 3	12 x 13	13 x 13
4	237 x 2	216 x 2	37 x 22	16 x 22	231 x 2	213 x 2	31 x 22	13 x 22
5	226 x 3	215 x 3	26 x 23	15 x 23	221 x 3	213 x 3	21 x 23	21 x 23
6	327 x 2	327 x 2	27 x 32	27 x 32	321 x 2	331 x 2	21 x 32	31 x 32
7	327 x 3	316 x 3	27 x 33	16 x 33	311 x 3	321 x 3	11 x 33	33 x 21
8	338 x 2	346 x 2	38 x 12	46 x 12	324 x 2	332 x 2	33 x 21	13 x 22
9	217 x 2	235 x 2	17 x 22	35 x 22	211 x 2	211 x 2	11 x 22	22 x 11

Appendix E. Division problems

Single Divisions	ZCL	ZCT	ZCS	OCL	OCT	OCS
1	242 / 2	246 / 2	248 / 2	238 / 2	252 / 2	254 / 2
2	369 / 3	363 / 3	339 / 3	372 / 3	348 / 3	375 / 3
3	424 / 2	442 / 2	428 / 2	432 / 2	478 / 2	436 / 2
4	484 / 4	448 / 4	488 / 4	492 / 4	456 / 4	476 / 4
5	636 / 3	639 / 3	633 / 3	654 / 3	681 / 3	672 / 3
6	624 / 2	626 / 2	628 / 2	634 / 2	674 / 2	678 / 2
7	842 / 2	846 / 2	848 / 2	852 / 2	854 / 2	856 / 2
8	936 / 3	939 / 3	963 / 3	951 / 3	957 / 3	972 / 3
9	884 / 4	844 / 4	848 / 4	852 / 4	856 / 4	864 / 4
AS Divisions	ZCL	ZCT	ZCS	OCL	OCT	OCS
1	262 / 2	264 / 2	268 / 2	272 / 2	274 / 2	256 / 2
2	339 / 3	369 / 3	369 / 3	384 / 3	357 / 3	351 / 3
3	484 / 2	486 / 2	468 / 2	456 / 2	472 / 2	474 / 2
4	448 / 4	484 / 4	488 / 4	452 / 4	476 / 4	496 / 4
5	663 / 3	696 / 3	669 / 3	678 / 3	657 / 3	651 / 3
6	642 / 2	648 / 2	646 / 2	652 / 2	658 / 2	654 / 2
7	868 / 2	864 / 2	862 / 2	892 / 2	896 / 2	894 / 2
8	939 / 3	963 / 3	993 / 3	978 / 3	987 / 3	954 / 3
9	848 / 4	884 / 4	844 / 4	892 / 4	896 / 4	868 / 4
RLG Divisions	ZCL	ZCT	ZCS	OCL	OCT	OCS
1	282 / 2	284 / 2	286 / 2	292 / 2	276 / 2	278 / 2
2	396 / 3	396 / 3	336 / 3	381 / 3	378 / 3	387 / 3
3	462 / 2	464 / 2	482 / 2	452 / 2	434 / 2	476 / 2
4	488 / 4	484 / 4	448 / 4	472 / 4	464 / 4	468 / 4
5	639 / 3	699 / 3	693 / 3	675 / 3	687 / 3	684 / 3
6	686 / 2	682 / 2	684 / 2	676 / 2	638 / 2	692 / 2
7	884 / 2	884 / 2	882 / 2	876 / 2	878 / 2	874 / 2
8	993 / 3	996 / 3	969 / 3	984 / 3	981 / 3	975 / 3
9	844 / 4	848 / 4	884 / 4	872 / 4	836 / 4	876 / 4

Appendix F.

Counter-Balanced Task Order – Second Series of Experiments

Additions and Subtractions

Sequence of Twelve Blocks	Participant Starting Order				
Single 1	P1	P5	P9	P13	P17
AS 1					
RLG 1					
AS 2					
RLG 2					
Single 2	P2	P6	P10	P14	P18
RLG 1					
Single 3	P3	P7	P11	P15	P19
AS 3					
Single 4	P4	P8	P12	P16	P20
RLG 4					
AS 4					

Twelve Blocks of Equivalent Trials (Single 1 – AS 4)
All twenty participants started the sequence of twelve trials from one of the four different (S 1 – S 4) positions.

Multiplications and Divisions

Sequence of Nine Blocks	Participant Starting Order						
Single 1	P1	P4	P7	P10	P13	P16	P19
AS 1							
RLG 1							
AS 2							
RLG 2							
Single 2	P2	P5	P8	P11	P14	P17	P20
RLG 1							
Single 3	P3	P6	P9	P12	P15	P18	
AS 3							

Nine Blocks of Equivalent Trials (Single 1 – AS 3)
All twenty participants started the sequence of nine trials from one of the three different (S 1 – S 3) positions.

Appendix G. Strategies for addition and subtraction

A) COUNTING ON / COUNTING DOWN of hundreds, tens or units.

178 + 26 = 178, 188, 198, 199, 200, 201, 202, 203, 204.

178 – 26 = 178, 168, 158, 157, 156, 155, 154, 153, 152.

B) OVERSHOOTING / UNDERSHOOTING

178 + 26 = 178 + 30 = 208, 208 – 4 = 204

178 – 26 = 178 – 30 = 148, 148 + 4 = 152

C) PARTITIONING

178 + 26 = (170 + 20) + (8 + 6) = 190 + 14 = 204.

178 –26 = (170 – 20) + (8 – 6) = 150 + 2 = 152.

D) MAKE NEAREST HUNDRED OR NEAREST WHOLE NUMBER

178 + 26 = (178 + 22) + (26 – 22) = 200 + 4 = 204.

178 – 26 = (178 + 2) – (26 + 2) = 180 – 28 = 152.

E) MENTAL CARRYING – BORROWING / PAYING BACK

178 + 26 =

H	T	U
1	9	14
1	10	4
2	0	4

176 – 28 =

H	T	U
1	5	(-2)
1	4	8

F) RETRIEVAL – JUST KNOW THE ANSWER WITHOUT CALCULATION.

G) ANOTHER STRATEGY NOT GIVEN ABOVE, OR A MIXTURE OF STRATEGIES FROM THOSE ABOVE.

Appendix H. Strategies for division and multiplication

A) OVERSHOOTING / UNDERSHOOTING

29 x 12 = (30 x 12) – 12 = 360 – 12 = 348.

951 / 3 = (960 / 3) – (9 / 3) = 320 – 3 = 317.

B) PARTITIONING

29 x 12 = (20 x 12) + (9 x 12) = 240 + 108 = 348.

Or (29 x 10) + (29 x 2) = 290 + 58 = 348.

C) MENTAL CARRYING – BORROWING / PAYING BACK

	H	T	U
29 x 12 =		1	8
		4	0
	2	9	0
	3	4	8

	H	T	U	
465 / 3 =	1			r.165
165 / 3 =		4		r. 45
45 / 3 =		1	5	
	1	5	5	

D) CHANGE NUMBERS IN PROPORTION

29 x 12 = 58 x 6 = 116 x 3 = 348.

936 / 9 = 312/ 3 = 104.

E) RETRIEVAL - JUST KNOW THE ANSWER WITHOUT CALCULATING

F) ANOTHER STRATEGY NOT GIVEN ABOVE