INVESTIGATING THE EFFECT OF ITEM-CONTEXT ON STUDENTS’ PERFORMANCE ON MATHEMATICS ITEMS

Felipe Javier Almuna Salgado

Submitted in partial fulfilment of the requirements of the degree of Master of Education (with coursework component)

Melbourne Graduate School of Education
The University of Melbourne

December 2010
ABSTRACT

The current research in Mathematics education recommends to measure how well students are able to apply their knowledge and mathematical skills and use them to solve problems in a real-life context but an increasing body of research has been ambivalent about the effects of contextualising mathematical tasks on students’ performance.

Thus, the present piece of research aims to investigate the effect of an alteration of item-context on students’ performance, if any, on Mathematics items. Items with the same mathematical core whilst varying in objective figurative context are used in this study. The study will also examine to what extent the nature of demand of the item-context may affect students’ performance. A mixed method approach is used due to the character of the research aims.

In this study significant differences were not been found in students’ performance when they solved items across altered contexts; but evidence indicated that students’ performance on items required the second-order use of context may be more influenced by the alteration of context.

The results of this research give some insight into how the second-order use of context may influence students’ performance on Mathematics items. This outcome enhances the understanding of contextualised mathematical assessment and provides a foundation for future research into contextualised mathematical tasks.
DECLARATION

This is to certify that

(1) The thesis comprises no material which has been accepted for any other degree in any university.

(2) To the best of my knowledge and belief, this thesis contains no material previously published or written by any other person, except where due reference is given in the text.

(3) The thesis is less than 20,000 words in length, exclusive of tables, figures, references, and appendices.

___________________________
Felipe Javier Almuna Salgado
ACKNOWLEDGMENTS

I would like to thank the following people for their help, patience, and kind understanding throughout the period of my research:

- To my supervisor Professor Kaye Stacey for her guidance, patience, constant support, and the many stimulating conversations that prompted me to complete this study.

- To my wife, María Alejandra, for her support, understanding, tolerance, and assistance of my research journey; without her I could not have made this exciting journey.

- To CONICYT (Comisión Nacional de Investigación Científica y Tecnológica, Chile) that financially supported my Master of Education’s studies at The University of Melbourne.

- To my friend Mateo Caulkins for his time and help in checking my English grammar and style, correcting both and offering suggestions for improvement.

- To the Participants of my research, without whom, there would be no research.

- To my fellow postgraduate researches of the Science and Mathematics Education Cluster at The University of Melbourne, especially to Rohani, and Malcolm for their reflections about my ideas for this thesis.
CONTENTS

ABSTRACT .................................................................................................................. II
DECLARATION ............................................................................................................ III
ACKNOWLEDGMENTS ............................................................................................... IV
INDEX OF TABLES ...................................................................................................... VIII
INDEX OF FIGURES .................................................................................................. VIII

CHAPTER 1: INTRODUCTION ................................................................................. 1

1.1 INTRODUCTION .................................................................................................. 1
1.2 RESEARCH PROBLEM ...................................................................................... 3
  1.2.1 Research Questions .................................................................................. 3
1.3 DEFINITION OF KEY TERMINOLOGY .......................................................... 4
  1.3.1 Item-Context .......................................................................................... 4
  1.3.2 Types of Uses of Context ........................................................................ 7
  1.3.3 PISA Definition of Mathematical Literacy ............................................. 10
1.4 PROPOSED RESEARCH DESIGN: AN OVERVIEW ........................................ 11
1.5 IMPETUS AND SIGNIFICANCE OF THIS STUDY ........................................... 11
1.6 PLAN OF THIS STUDY .................................................................................... 12
1.7 FORMAT OF THE THESIS ............................................................................. 12

CHAPTER 2: RESEARCH ON CONTEXT ............................................................... 13

2.1 INTRODUCTION .................................................................................................. 13
2.2 AN OVERVIEW ON MATHEMATICAL TASKS EMBEDDED IN CONTEXT ........... 13
  2.2.1 Introduction .......................................................................................... 13
  2.2.2 Standard Applications and Modelling Mathematical Tasks ...................... 17
2.3 THE CONTEXT EFFECT .................................................................................. 22
  2.3.1 The Context Effect: Its Differential Nature ............................................ 23
2.4 STUDENTS’ DEALING WITH THE TASK CONTEXT ....................................... 30
2.5 IMPLICATIONS OF THE LITERATURE REVIEW FOR THIS STUDY ............ 32

CHAPTER 3: METHODOLOGY ............................................................................. 33

3.1 INTRODUCTION .................................................................................................. 33
3.2 RATIONALE FOR THE USE OF MIXED METHODOLOGY .............................. 33
  3.2.1 Research Design .................................................................................. 35
3.3 INSTRUMENTATION ......................................................................................... 38
  3.3.1 The Test Items ..................................................................................... 38
  3.3.2 Characterisation and Differences among Family of Items ....................... 40
  3.3.3 Rotated Test Design ............................................................................ 46
  3.3.4 Pilot of the Items ................................................................................ 46
3.4 THE MARKING SCHEME ................................................................................ 47
3.5 DATA COLLECTION PROCEDURE, PARTICIPANTS, AND TASK ADMINISTRATION ....... 48
  3.5.1 Procedure .......................................................................................... 48
  3.5.2 Participants ......................................................................................... 49
  3.5.3 Task Administration ........................................................................... 49
3.6 STUDENTS’ INTERVIEW .............................................................................. 50
3.6.1 Stimulus Material Provided and Structure of the Interview ................................................. 50
3.7 DATA ANALYSIS .......................................................................................................................... 51
3.8 METHODOLOGICAL ISSUES ........................................................................................................ 52
  3.8.1 Student Sample Size and Potential for Generalisation ......................................................... 52
  3.8.2 Ethical Considerations ............................................................................................................. 53
3.9 SUMMARY OF HYPOTHESES AND STATISTICAL PROCEDURES ............................................. 53

CHAPTER 4: RESULTS .......................................................................................................................... 55
  4.1 INTRODUCTION ............................................................................................................................ 55
  4.2 RESULTS OF THE QUANTITATIVE PHASE ................................................................................. 55
    4.2.1 Effect of a more Familiar Context .......................................................................................... 59
    4.2.2 Effect of Fewer Constraints in the Item .................................................................................. 61
  4.3 THE USES OF CONTEXT AND OVERALL PERFORMANCE ......................................................... 63
    4.3.1 The First-Order Use of Context and Overall Performance .................................................... 64
    4.3.2 The Second-Order Use of Context and Overall Performance ................................................ 65
    4.3.3 Effect of the Nature of the Use of Context on Overall Students’ Performance ....................... 65
  4.4 RESULTS OF THE QUALITATIVE PHASE ................................................................................... 69
    4.4.1 Introduction ............................................................................................................................. 69
    4.4.2 Description of the Data Included in this Section ..................................................................... 69
  4.5 STUDENTS DEALING WITH THE FIRST-ORDER USE OF CONTEXT ........................................... 71
  4.6 STUDENTS’ DEALING WITH THE SECOND-ORDER USE OF CONTEXT ...................................... 78
  4.7 STUDENTS CONNECTION WITH THE ITEM-CONTEXT ................................................................ 84

CHAPTER 5: CONCLUSION .................................................................................................................... 90
  5.1 MAJOR FINDINGS .......................................................................................................................... 90
  5.2 LIMITATIONS OF THIS STUDY ................................................................................................... 92
  5.3 RECOMMENDATIONS ................................................................................................................ 94
    5.4.1 Recommendations for PISA Mathematics ............................................................................. 94
    5.4.2 Further Research .................................................................................................................... 95

REFERENCES ......................................................................................................................................... 97

APPENDIX A: THE TWELVE ITEMS USED IN THE STUDY ................................................................. II
APPENDIX B: THE MARKING SCHEMES ........................................................................................... V
APPENDIX C: TRANSCRIPTS OF THE INTERVIEWS ........................................................................ IX
INDEX OF TABLES

TABLE 1
EXAMPLE OF THE OBJECTIVE AND SUBJECTIVE FIGURATIVE CONTEXTS IN TWO MATHEMATICAL TASKS ....5
TABLE 2
THE THREE USES OF CONTEXT .............................................................................................................8
TABLE 3
EXAMPLES OF STANDARD APPLICATION TASKS ...............................................................................18
TABLE 4
EXAMPLES OF MODELLING TASKS .....................................................................................................20
TABLE 5
BUSSE’S TYPOLOGY OF STUDENTS’ DEALING WITH THE CONTEXT .....................................................31
TABLE 6
PISA ITEMS SELECTED FOR THE STUDY ..............................................................................................38
TABLE 7
THE TWELVE ITEMS USED IN THE STUDY ............................................................................................41
TABLE 8
ITEMS ARRANGEMENT PER BOOKLET ..................................................................................................46
TABLE 9
MARKING SCHEMES FOR FAMILY B-ITEMS ............................................................................................48
TABLE 10
QUESTIONS USED IN THE STUDENTS’ INTERVIEW ..............................................................................51
TABLE 11
SUMMARY OF HYPOTHESES AND DATA ANALYSIS PROCEDURES FOR TESTING THEM .................54
TABLE 12
DISTRIIBUTION OF THE STUDENTS’ SCORES, IN PERCENTAGES, SAMPLE SIZE, ITEM CHARACTERISTICS, ITEM MEAN AND ITEM STANDARD DEVIATION (FROM A MAXIMUM OF 2 SCORE POINTS) WITH THE PISA MARKING SCHEME ACROSS ALL ITEMS ........................................................................................................58
TABLE 13
DISTRIIBUTION OF THE STUDENTS’ SCORES, IN PERCENTAGES, SAMPLE SIZE, ITEM CHARACTERISTICS, ITEM MEAN AND ITEM STANDARD DEVIATION (FROM A MAXIMUM OF 2 SCORE POINTS) WITH THE ADAPTED MARKING SCHEME ACROSS ALL ITEMS ........................................................................................................59
TABLE 14
INDIVIDUAL RESULTS IN RELATION TO BOOKLETS’ ITEMS AND USES OF CONTEXT .....................70

INDEX OF FIGURES

FIGURE 1
VISUAL MODEL FOR MIXED METHODS PROCEDURES (THE TWO PHASES RESEARCH DESIGN) ..........37
FIGURE 2
MEAN OF STUDENTS’ PERFORMANCE BY ITEM ON PI AND PLI1 VERSIONS ........................................60
FIGURE 3
MEAN OF STUDENTS’ PERFORMANCE BY ITEM ON PI AND PLI2 VERSIONS ........................................62
FIGURE 4
MEAN OF STUDENTS’ PERFORMANCE ON THE USES OF CONTEXT ON PI AND PLI1 ITEMS BY USED OF CONTEXT ........................................................................................................................................64
FIGURE 5
SUMMARY OF MEAN OF STUDENTS’ PERFORMANCE BY ITEM FAMILY AND USES OF CONTEXT ON PI AND PLI1 ITEMS ........................................................................................................................................66
FIGURE 6
MEAN ABSOLUTE DIFFERENCE IN MEAN SCORE BY ITEM AND USES OF CONTEXT ........................67
FIGURE 7
PARTICIPANT 1’S SOLUTION METHOD ON PI-A ................................................................. 71
FIGURE 8
PARTICIPANT 4’S WRITTEN ANSWER ON PLI1-C .......................................................... 73
FIGURE 9
PARTICIPANT 7 SOLUTION TO PI-C .................................................................................. 75
FIGURE 10
PARTICIPANT 5 SOLUTION ON PLI1-C ............................................................................ 76
FIGURE 11
P4 SOLUTION ON PI-B ........................................................................................................ 78
FIGURE 12
PARTICIPANT 10 SOLUTION ON PI-B ............................................................................... 79
FIGURE 13
PARTICIPANT 8 SOLUTION ON PLI1-B ............................................................................ 81
FIGURE 14
PARTICIPANTS DRAWINGS ON PLI1-B ........................................................................... 83
FIGURE 15
PARTICIPANT 4 SOLUTION ON PLI1-D ............................................................................ 86
FIGURE 16
PARTICIPANTS 3 AND 5 SOLUTIONS ON PLI1-D RESPECTIVELY ................................. 87
CHAPTER 1: INTRODUCTION

1.1 Introduction

In the 1980s and 1990s advances in cognition and measurement theories led researchers to consider how assessment could be developed to ensure that meaningful aspects of students’ learning and thinking are captured (Lane, 2004). In this manner, different curricula and standard documents around the globe started to develop new forms of connectedness of mathematical content by focusing on contextualisation and problem solving (Stillman, 2000). They also focussed on forms of assessments that required the application of not just procedural knowledge but also conceptual and contextual knowledge to solve contextualised mathematical tasks in order for students to apply Mathematics in a variety of circumstances. In this manner, practitioners using contextualised mathematics tasks in context commonly hoped to foster motivation for the learning of Mathematics in order to perceive the links and transfers between Mathematics concepts and methods and the real-world. Relations between mathematical tasks in context and Mathematics are particularly relevant in PISA\(^1\) Mathematics because the emphasis of this assessment is on mathematical knowledge in context.

In this vein, since the first results of PISA Mathematics were published in 2001, constant debate about the role of item-context and its relation to students’ performance has been taking place in the research community (Blum, 2002) because there are views which consider including contextual elements may diminish students’ performance by increasing the level of difficulty and

\(^1\) The Programme for International Student Assessment, conducted by the Organisation for Economic Co-operation and Development (OECD). For more information refer to www.pisa.oecd.org
confusion for some students (Clarke, 2004). Although there is an extensive body of empirical studies on students’ performance on PISA, namely: student-related factors, teacher-related factors, school-related factors, educational systems institutional-organisational related factors (Hatzinikita, Dimopoulos, & Christidou, 2008), very few studies consider the effects of item-context factors on students’ performance. But as will be shown in the literature review on research about item-context they do not satisfactorily answer the question of how item-context factors may influence students’ performance on assessments of contextualised items in Mathematics.

As a result, the present piece of research aims to examine on one hand, differences in students’ performance, if any, when they solve three kinds of items: (1) PISA Mathematics items, (2) PISA-like items embedded in contexts that are more familiar to students, and (3) items with fewer constraints. These item types will be fully explained in Chapter 3. The study will also examine to what extent the nature of demand of the item-context may affect students’ performance. This study is based on a mixed method approach (Creswell, 2003, 2005) due to the character of the research aims.

The results of this research will try to give some insight into how item-context may influence students’ performance on Mathematics items. Therefore it will contribute to a deeper understanding of assessment using contextualised items which in turn will have some implications for PISA Mathematics and for directions for teaching and learning in the Mathematics classroom.

The remainder of Chapter 1 is organised as follows: Section 1.2 addresses the research problem and outlines the research questions of this study; Section 1.3 presents a definition of key terminology. On a different vein, Section 1.4
outlines the anticipated research design for this study whilst Section 1.5 presents the impetus and significance of this study. Finally from section 1.6 to 1.7 the plan and format of this study are explained.

1.2 Research Problem

This study undertakes to investigate how an alteration of the context in which an item is embedded impacts on students’ performance. Students will be presented with items of three different types. Their responses to the items could be categorised in a number of ways but this study elects to investigate the effects of the three different types of item-context listed below on students’ performance, that is to say:

(a) A more familiar and accessible context on students’ performance.
(b) Fewer constraints in the items on students’ performance.
(c) The nature of the use of context on students’ performance.

1.2.1 Research Questions

Four research questions articulate the focus of this study. They correspond to the effects listed above. It should be noted that no particular effect is attached to Research Question 4 below, but it would illustrate and enrich the research findings of this study. The research questions are:

Research Question 1
Is the students’ performance on PISA-like items with a more familiar context to students different to that on PISA items?

Research Question 2
Is performance on items with fewer constraints than PISA items different to that on PISA items?

Later (see section 4.2.2) it was decided that this should not be regarded as an effect of item-context.
Research Question 3
Are the items which require second-order use of context (see 1.3.2) more affected by change of context than items which require first-order use of context?

Research Question 4
How do the uses of context (first and second order) affect students’ performance on Mathematics items?

Each of these research questions will be operationalised in Chapter 3, which also outlines the statistical procedures and tests for the associated research hypotheses.

1.3 Definition of Key Terminology

Key terminologies such as item-context, uses of context, PISA mathematical literacy, standard applications and modelling are indispensable to an understanding of the research topic and will be explained to clarify their intended meaning. However, terms such as standard applications and modelling are going to be fully explained section in 2.2 because of their relation to the research literature.

1.3.1 Item-Context

Context is a term that takes a number of meanings in Mathematics education; consequently, several names for context can be found in the literature. For instance, terms such as: cover story (Chapman, 2009), thematic content (Ross, McCormick, & Krisak, 1986), setting (Huang, 2004), and figurative context (Busse & Keiser, 2003; Clarke & Helme, 1996) are used as alternative names for context in the research literature.

In order to have a more refined understanding of the term context, Clarke and Helme (1996) define figurative context as “the scenario where the task is
encountered” (p. 4). Busse and Kaiser (2003) further refined the above by distinguishing between objective figurative context and subjective figurative context. According to these authors, the objective figurative context refers to “the description of the scenario given in the task” (p. 4) while subjective figurative context is associated to the “individual interpretation of the objective figurative context” (p. 4). Table 1 below presents an example of both objective and subjective figurative context for a task.

Table 1
Example of the objective and subjective figurative contexts in two mathematical tasks

<table>
<thead>
<tr>
<th>Task Example</th>
<th>Objective Figurative Context</th>
<th>Example of a Student's Subjective Figurative Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freda’s flat has five rooms. The total floor area is 60m². Draw a possible plan of her flat. Label each room and show the dimensions (length, width) of each room.</td>
<td>Flat of 5 rooms with a floor area of 60m²</td>
<td>Imagine a rock concert being held in a football stadium</td>
</tr>
<tr>
<td>For a rock concert a rectangular field of size 100 m by 50 m was reserved for the audience. The concert was completely sold out and the field was full with all the fans standing. Which one of the following is likely to be the best estimate of the total number of people attending the concert? A 2 000 B 5 000 C 20 000 D 50 000 E 100 000 Source OECD (2006, p. 94)</td>
<td>Rock concert to be held in a rectangular field of size 100 m by 50 m</td>
<td>Imagine a rock concert being held on a country farm</td>
</tr>
<tr>
<td>Imagine a rock concert being held in a public park</td>
<td>Source Clarke and Helme (1996, p. 9).</td>
<td></td>
</tr>
</tbody>
</table>

| Context Allocation | Shared amongst all students | For individual students |
In the first task example, the objective figurative context is related to a flat’s five rooms of 60m² of total floor area; nothing about Freda is important to solve the task. On the other hand, the subjective figurative context for the student takes form when the representation of the objective figurative context is drawn. For example, as can be seen in Table 1, although the image corresponds to the mathematical demands of the task, this individual students’ representation of the objective figurative context presents a flat that would not be appropriate to live in. In the second task example, the objective figurative context is related to a rock concert to be held in a rectangular field of size 100 m by 50 m; nothing about the place of the rock concert or the kind of audience attending are necessary to solve the task; in terms of the subjective figurative context, for instance, a solver might imagine that it is held in a rectangular field in a stadium (possibly with seating), farm, or a public park. These different subjective figurative contexts could make a difference to assumptions made in the solution, for example whether facilities for the crowd are inside or outside the rectangular field.

This study, however, is concerned with the objective figurative context in which a mathematical task is embedded. The importance of differentiating between aspects of context is fundamental to understand that apparently the objective figurative context is “often implicitly meant by researchers when referring to the context” (Busse & Kaiser, 2003, p. 4). For this reason the term objective figurative context will be referred to as context in this study except when otherwise stated.

In this study, the item context will be varied whilst holding the mathematical core of the item constant. The phrase mathematical core is used to describe the underlying mathematical model and solving challenges of the problem. For
instance, the second Example in Table 1, has the same mathematical core as a similar item asking about the number of cars that might fit into a car park of specified dimensions (see 3.3.2.2). Accordingly, in the rock concert example in Table 1 above, a mathematical model to solve the task is as follows:

\[
\text{Total number of people} = \text{Area of the field} \times \text{Number of people per m}^2,
\]

while on the car park example (see Appendix A, p. III) a mathematical model to solve the task is as follows:

\[
\text{Total number of cars} = \text{Area of the field} \div \text{Number of m}^2 \text{ per car}
\]

As can be seen, both items have a very similar mathematical model and solving challenges. The underlying mathematical models to solve the items above require the calculation of the area of the field and an estimation of how many people can be accommodated per square meter (for instance, the assumption needed to solve this task is that 4 people can be accommodated per m²) or how many m² are required per car in a car park (for instance, the assumption needed to solve this task is that 10m² of space are require per car); then, the mathematics involve to solve the models above consider a product or division of the figures obtained previously in order to answer the item. It should be noted that the discussion relate to the mathematical core of the items used in this study will be done in section 3.3.2.

1.3.2 Types of Uses of Context

The OECD (2009c) in its PISA Assessment Framework proposed a three-stage model in which to position different types of task context. This distinction is related to the use of context in items, namely: the zero-order use of context, the first-order use of context, and the second-order use of context (see Table 2 below). This differentiation in the use of context will be borrowed for this study.
Table 2
The three uses of context.

<table>
<thead>
<tr>
<th>Use of Context</th>
<th>Description</th>
<th>Location of the Context</th>
<th>Example</th>
</tr>
</thead>
</table>
| Zero-order use of context | No context is associated with the item or the context is used as a mere camouflage for the item. There should not be any interaction with the context. | Terms in the item are mathematical terms or the context is inside the item. | (1) Solve $3x + 2 = 5$  
(2) How much money does Mateo need to buy a blue plastic block, a red pen, and a wooden block?  
|                      |                                                                           |                          | Blue plastic block | $0.06$  
Wooden block | $0.08$  
Red pen | $0.03$ |
| First-order use of context | The item-context is needed for solving the problem and judging the answer. | Internal (inside the item) | A carpenter has 32 metres of timber and wants to make a border around a garden bed. He is considering the following designs for the garden bed.  
|                      |                                                                           |                          | Circle either “Yes” or “No” for each design to indicate whether the garden bed can be made with 32 metres of timber. |
| Second-order use of context | The item-context is really needed to move backwards and forwards between the mathematical problem and its context in order to solve the problem or to reflect on the answer within the context to judge the correctness of the answer. | Internal and External (inside and outside the item) | (1) There are many lakes in Australia that are dry for most of the time, only filling for short periods immediately after rain. Lake Eyre in South Australia is an example of one of these normally dry lakes. When a dry lake bed is filled with water, how rapidly will the lake empty?  

Adapted from OECD (2009a, p. 89)

3 According to the OECD (2009a) this item requires a zero-order use of context. However, it is argued here that this item is instead a first-order use of context. The argument against this source is presented and explained on p. 10.
Accordingly, the zero-order use of context is when the item generally involves just mathematical terms, shapes and data, and therefore there is not any interaction with the context. In the examples above, nothing related to Mateo or the stationery articles is required: they are a camouflage to practise addition.

The next level of use of context in PISA is the first-order use of context. This use takes place when the context is relevant and needed for solving the problem and judging the answer, and is already present in the item (internal presence of the context). An example of this use of context can be found in Table 2 above, where the context of a garden bed and timber are needed in order to understand and to evaluate the solutions for a task related to the geometrical concept of perimeter.

It should be noted that for the OECD (2009a, p. 89) this is an example of ‘zero order use of context’. Although it may look as a ‘zero order use of context’, it is argued here that the item-context (a carpenter with 32 meters of timber to make a border around a garden bed) does not act as a mere camouflage. Rather, it is needed in order to understand that the geometrical concept of perimeter is related to the four designs of a garden bed and consequently to the mathematical solution of this item. As a matter of fact, if the words ‘carpenter’, ‘border’, ‘garden’ and ‘timber bed’ were blocked out, this item could not be solved.

Finally, the second-order use of context takes place when students need to move backwards and forwards between the mathematical problem and its context in order to solve the problem or to reflect on the answer within the context to judge the correctness of the answer (external presence of the context). The second-order use of context requires a strong student\item-context
interaction (refer to 2.2.2). Assumptions relevant to the task-context are needed in the examples provided in Table 2 for this use of context; the student has to bring additional knowledge of the context into the solving process. In the Lake Eyre example above, additional knowledge such us: (a) size, shape, depth, surface area of water, (b) temperature, evaporation rate, (c) seepage rate, (d) rainfall rate, (e) salinity and (f) runoff rate are needed to solve the item. Then, a mathematical model is needed (which could be either a very flat cylinder or a flat cone). On the other hand, in the Rock Concert example, realistic knowledge relate to the number of people that can be accommodated per square metre is essential to solve correctly this item (this item will be fully discussed in 3.3.2.2).

1.3.3 PISA Definition of Mathematical Literacy

The key concept of the PISA theoretical framework is the concept of literacy. In this way, PISA is not linked to a particular curriculum but based on a definition of Mathematical Literacy by the OECD. According to the PISA 2009 assessment framework; Mathematical Literacy\(^4\)\(^5\) is:

> an individual’s capacity to identify and understand the role that Mathematics plays in the world, to make well-founded judgements and to use and engage with Mathematics in ways that meet the needs of that individual’s life as a constructive, concerned and reflective citizen (OECD, 2009c, p. 84).

As can be inferred from the definition above, relations between the item-context and Mathematics are particularly relevant in PISA, hence their importance to this study. The importance of the PISA definition of Mathematical Literacy is that it involves an interpretation and understanding of Mathematics “put into functional use in a multitude of different situations and contexts” (Blum, 2002, p. 151).

\(^4\) The same definition of mathematical literacy was used by the OECD in previous PISA surveys.

\(^5\) A major revision of the PISA Mathematics framework and definition is currently underway in preparation for the PISA 2012 assessment.
1.4 Proposed Research Design: An Overview

This research project is based on a mixed method approach (Creswell, 2003, 2005) due to the character of the research questions; Research Questions (1), (2), and (3) require a quantitative approach whilst Research Question (4) requires a qualitative interpretation. The quantitative method in this thesis will serve to identify differences in students’ performance, if any, when item context is changed. On the other hand, the qualitative component will serve to investigate and provide an understanding of how context may affect students’ accessibility to the items, and therefore students’ performance.

1.5 Impetus and Significance of this Study

The decision to undertake formal research on item-context factors affecting students’ performance on mathematical tasks in context responds, on one hand, to a personal interest in the field of contextualised Mathematics in education, and in PISA as an international assessment. Mathematical tasks in context can provide students with a bridge between the abstract role of Mathematics and their role as an active citizen in society, that is to say, Mathematics in context is considered appropriate in preparing students for the mathematical requirements they will face in their everyday lives.

On the other hand, an OECD (2010) report in PISA has also demonstrated the economic importance of improvements in Mathematics literacy, in the following way:

An improvement of one-half standard deviation in average performance on Mathematics and Science performance at the individual level implies, by historical experience, an increase in annual growth rates of GDP per capita of 0.87%. This historical impact suggests a very powerful response to improvements in educational quality (OECD, 2010, p. 17).
It is hoped that findings of this study will promote among teachers and assessment writers the importance of considering the item-context when assessing embedded items that require a link between item-context and Mathematics for their solution.

1.6 Plan of this Study

This research project will be implemented in two Year 10 classes. Hence, the mathematical work of students approximately 15 years old will be part of this study. The decision to study the mathematical work of 15 year old students is because PISA is administered at this age.

Considering the above, quantitative data (a paper and pencil test) and qualitative data (an interview) will be collected in order to answer the four research questions of this study (see 1.2.1). Section 3.5 to 3.7 supply further information about the types of data collected, description as well as analysis of the data.

1.7 Format of the Thesis

This thesis relies upon research into item-context which is developed across six chapters. The composition of these chapters is as follows. Chapter 1 introduces the research problem, definition of key terminology and purpose of this study. Chapter 2 provides a literature review on research into context. On a different vein, Chapter 3 offers the research methodology and methods which guide the data collection process of this study. Chapter 4 reveals the findings of both quantitative and qualitative instruments; finally Chapter 5 poses and formulates overall conclusions providing additionally advice for this research topic.
CHAPTER 2: RESEARCH ON CONTEXT

2.1 Introduction

This chapter reviews and outlines the literature related to this study. It gives an overview on mathematical tasks embedded in context and describes the influence of item-context on students’ performance and reports studies associated with this issue. Finally, this chapter concludes by examining an approach on students’ dealing with item-context in order to understand how it may affect the student performance in Mathematics. In this manner, an extensive literature search on item-context factors affecting students’ performance was undertaken. Key words and phrases were used to search different data bases (Academic Search Premier (EBSCO), JSTOR, Web of Science (ISI), Education Resources Information Center (ERIC), Dissertations and Thesis (Proquest), American Educational Research Association (SAGE) and the Wilson Education Full Text database). Searches included relevant terms and phrases such as: item-context Mathematics, PISA Mathematics, item-context students’ performance, and item-context factors affecting students’ performance on Mathematics, among others.

2.2 An Overview on Mathematical Tasks Embedded in Context

2.2.1 Introduction

Frames of recommendations for reforming Mathematics education in the U.S. in the 1980s and 1990s especially called for an increased emphasis on problem solving and a decreased emphasis on the repeated practice of algorithms (NCTM, 1989, 2000); not only because of their potential for “motivating students and for the meaningful development of new Mathematics concepts and skills” (Depaepea, DeCorte, & Verschaffel, 2010, p. 152), but also to develop
in students the skills of knowing when and how to apply the Mathematics they know effectively in situations encountered in everyday life and at work (Blum, Galbraith, & Niss, 2007; Boaler, 1993a; OECD, 2009a).

These recommendations also took place in different curricula and standard documents around the world. In Australia, for example, the emphasis on problem solving and on modelling has been included formally as a major thrust in the guidelines for Mathematics curricula (Stacey & Groves, 1984, 1989). In particular, the state of Victoria led the world in Mathematics modelling and problem solving at school in the 1990s. Starting in 1982, students have been introduced to meaningful experiences in Mathematics which stress the ability of students to apply their mathematical knowledge in solving new, and non-routine contextualised mathematical tasks through lessons based on problem solving and modelling the experience of doing mathematical investigations and projects (Stacey & Groves, 1984, 1989). This has been done in cooperation with teachers. In this manner, the use of contextualised mathematical tasks commonly hopes to foster motivation for the learning of Mathematics and to perceive the links and transfers between Mathematics concepts and methods and the world.

The previous can be furthered by a set of arguments for which contextualised tasks should be used in Mathematics. The main arguments found in the literature (Blum & Niss, 1991; Boaler, 1993a; Felton, 2010; Helme, 1994; Stacey & Pierce, 2006) are summarised below:

(1) **The Formative Argument**

The emphasis is put on the application of Mathematics in context as a means for developing general competences, attitudes, and skills orientated towards fostering creative and problem solving abilities as well as “open-mindedness,
self-reliance, and confidence in their [students’] own powers” (Blum & Niss, 1991, p. 42).

(2) The Critical Competence Argument
This argument highlights the importance of preparing mathematically literate students to enable them to “see and judge independently, to recognise, understand, analyse, and assess representative examples of the uses of Mathematics, including solutions to socially significant problems” (Blum & Niss, 1991, p. 43).

(3) The Utility Argument
Contextualised tasks may enhance the transfer of Mathematics to other contexts. This sort of task may increase the chance of students applying Mathematics to other areas later in studies or employments. Mathematics is seen under this argument as a service subject or as a subject of instrumental interest (Helme, 1994). This argument relies on the assumption that the ability to use Mathematics in context “does not result automatically from the mastering of pure Mathematics but requires some degree of preparation and training” (Blum & Niss, 1991, p. 43).

(4) The Picture of Mathematics Argument
This argument stresses the importance to provide students with a rich and comprehensive picture of Mathematics in all its facets, “as a science, as a field of activity in society and culture” (Blum & Niss, 1991, p. 43).

(5) Promoting Mathematical Learning Argument
This argument insists that Mathematics in context is well suited to assist students in “acquiring, learning, and keeping mathematical concepts, notions, methods, and results, by providing motivation for and relevance of mathematical studies” (Blum & Niss, 1991, p. 44); contributing to forming students who can think mathematically within and outside of Mathematics.
(6) *Making Sense of Mathematics Argument*

The use of contexts may assist in overcoming the common perception of Mathematics as a “remote body of knowledge” (Boaler, 1993, p. 13) with no connection to the real-world. It may serve to demonstrate that it has a relevant meaning in the real-world.

(7) *The Use of Mathematics in Context Argument*

Mathematical tasks in real-world contexts allow students to understand the connection between Mathematics and the real-world (Felton, 2010, p. 61).

(8) *The Halo-Effect Argument*

Last but not least, Stacey and Pierce (1994) show that some teachers use contexts that appeal to students (for example a problem about a dog) to improve students’ attitude towards learning Mathematics by associating the subject with pleasant things.

While the importance of using mathematical tasks in context seems to be well acknowledged, a March 2010 search on the ERIC (CSA), Web of Science (ISI), Education Research Complete (EBSCO), and Expanded Academic ASAP (Gale) databases indeed reveals a lack of research papers on item-context. This search resulted in 15 articles specifically on the relationship between item-context and students’ performance published over the period 1986 to 2010. As the specific research of item-context factors in Mathematics education does not seem to be, in general, part of the research agenda, the consequences of ignoring the effects may be overlooked.
2.2.2 Standard Applications and Modelling Mathematical Tasks

To begin with, Galbraith (1987) establishes that mathematical tasks embedded in contexts are often termed *applications*. Generally, applications require a translation of the problem into a suitable representation to produce the task comprehension, interpretation, and a mental representation of the problem. Then they require a formulation of a mathematical model which is linked with that representation, and the successful choice and use of relevant Mathematics involved in solving the problem.

In the examination of how the context is needed to solve mathematical tasks, two different kinds of application tasks can be distinguished, namely: *standard applications* and *modelling tasks*. These kinds of applications sometimes blur because both take into account contextual elements in their formulation, but in the examination of their nature there are fine distinctions which reveal essential differences between them.

On one hand, Blum et al. (2007) stress the term *standard application* tends to “focus on the direction: mathematics→reality” (p. 10) and therefore it generally emphasises the mathematical concepts involved. In simple but at the same time significant words “with applications we are standing inside Mathematics looking out: Where can I use this particular piece of mathematical knowledge”.

In standard applications the item-context plays a secondary role, that is to say, the item-context is treated routinely because students have been taught how to use a particular piece of Mathematics which fits into a predetermined model or technique. Hence, the item-context is not important to solve the item; all the information provided by the item-context is assumed to be known and the
A mathematical model has been taught to students for its relevance to everyday life. Hence, standard applications can be solved without any further link to the given context in which the item is embedded. Table 3 presents examples of standard application tasks.

<table>
<thead>
<tr>
<th>Task Examples</th>
<th>Use of Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>This year a ticket to the AFL (Australian Football League) was $20.00. Next year the price will be increased by 10%. What will be the new price of a ticket next year?</td>
<td>Zero-Order</td>
</tr>
<tr>
<td>Mark (from Sydney, Australia) and Hans (from Berlin, Germany) often communicate with each other using “chat” on the Internet. They have to log on to the Internet at the same time to be able to chat. To find a suitable time to chat, Mark looked up a chart of world times and found the following:</td>
<td>First-Order</td>
</tr>
<tr>
<td><strong>Question 2:</strong> Mark and Hans are not able to chat between 9:00 AM and 4:30 PM their local time, as they have to go to school. Also, from 11:00 PM till 7:00 AM their local time they won’t be able to chat because they will be sleeping. When would be a good time for Mark and Hans to chat?</td>
<td>Source (OECD, 2009a, p. 36)</td>
</tr>
</tbody>
</table>

In the first example, the mathematical concept of percentages is embedded in a real life situation; the mathematical model to solve the problem is immediately at hand (20 x 1.1 = 22, therefore the new price is $22.00), extra-information is not needed because students know how to use this particular model in a range of contexts. In the next example, Mathematics to solve the problem is also at hand adding and subtracting whole numbers with adjustment modulo 12 (9 hours difference between Berlin and Sydney for Q1) and it is assumed that students know that time difference remains constant between cities. Neither assumptions nor a new mathematical model are required from students in the examples above, hence the emphasis of standard applications is on the Mathematics involved but not on the item-context.
Taking Table 3 into account, the nature of the uses of context leads to link the zero-order and first-order use of context (see Table 2) with standard applications. In the first example nothing about the AFL’s ticket is required to determine the new price of the ticket, as a matter of fact, if the word AFL was blocked out, this task could still be solved (zero-order use of context). In the second example, the use of context is necessary to understand the way that time spent sleeping and at school constrains the times that could be suitable for communicating with each other. The solution needs to be evaluated and must fit the task context (first-order use of context).

In a different direction, modelling tasks tend to “focus on the direction: reality→mathematics” (Blum et al., 2007, p. 10) and therefore they generally emphasise an interaction process between context and mathematics. Once again, in simple words, [with this kind of problems] “we are standing outside Mathematics looking in: Where can I find some Mathematics to help me with this problem?” (Blum et al., 2007, p. 10).

The essence of modelling tasks is that students do not know either data or the mathematical model already. It has not been taught because the situation is not common or important enough in real life to teach students. Hence, in modelling tasks the context plays an important role because the information (data) to solve the task is usually found in the task-context. Table 4 below illustrates examples of modelling tasks.
### Table 4

*Examples of modelling tasks*

<table>
<thead>
<tr>
<th>Task Examples</th>
<th>Use of Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>For a rock concert a rectangular field of size 100 m by 50 m was reserved for the audience. The concert was completely sold out and the field was full with all the fans standing. Which one of the following is likely to be the best estimate of the total number of people attending the concert? A 2 000 B 5 000 C 20 000 D 50 000 E 100 000</td>
<td></td>
</tr>
<tr>
<td>In modern architecture, buildings often have unusual shapes. The picture below shows a computer model of a ‘twisted building’ and a plan of the ground floor. The compass points show the orientation of the building. The ground floor of the building contains the main entrance and has room for shops. Above the ground floor there are 20 storeys containing apartments. The plan of each storey is similar to the plan of the ground floor, but each has a slightly different orientation from the storey below. The cylinder contains the elevator shaft and a landing on each floor.</td>
<td></td>
</tr>
<tr>
<td><strong>Second-Order</strong></td>
<td></td>
</tr>
<tr>
<td>In the first task the dimensions of a hypothetical rectangular field are presented and the space occupied by a person has to be mathematised in order to estimate the number of rock fans that could be accommodated. In this item, certainly the Mathematics involved to solve the item is simple, although this task needs to be solved taking into account contextual information. An assumption not found in the task-context is needed, that is to say: the number of people that can be accommodated per square metre; then, a mathematical model can be used in order to produce a solution that addresses the problem.</td>
<td></td>
</tr>
</tbody>
</table>

*Source OECD (2006, pp. 86, 94)*
The next example, Question 1 requires an estimation of the height of a twisted building. Here there is not data at hand to solve the task straightforwardly. Thus, these modelling tasks can be linked with the second-order use of context.

Consequently, items which stress the second-order use of context require an “explicit reference to the context in order to produce a solution that addresses the problem” (OECD, 2009a, p. 31). This involves a purposeful interpretation of contexts in order to produce a relevant mathematical representation of the underlying problem as exemplified in Table 4. Hence, the second-order use of context can be linked to modelling tasks.

In PISA, second-order use of context tasks are where students have to provide assumptions (data) even if the mathematical relationships are relatively easy⁶. For instance, in the rock concert example in Table 4 above a mathematical model to solve the task is as follows:

\[ \text{Total number of people} = \text{Area of the field} \times \text{Number of people per m}^2 \]

The assumption needed to solve this task is that approximately 4 people can be accommodated per m².

In the twisted building example (Question 1) in Table 4 above, a mathematical model is as follow:

\[ \text{Height of the building} = \text{Number of floors} \times \text{Height of every floor} \]

The assumption needed to solve this task is that a floor has a height of approximately 3.5 metres.

---

⁶ In most modelling tasks the mathematical relationships have to be supplied by the students. See Lake Eyre example in Table 2.
2.3 The Context Effect

An increasing body of research has been ambivalent about the effects of contextualising mathematical tasks (Boaler, 1994; Busse, 2005; Vappula & Clausen-May, 2006). On one hand, some of this research has pointed to the possibilities of misinterpretation of the task when embedded in a context (Boaler, 1993b; Shannon, 2007; Verschaffel, Corte, & Lasure, 1994). Others have pointed out the beneficial effect of context by providing students a model of interaction between Mathematics and the real-world (Blum & Niss, 1991; Stillman, 2002).

A commonly held belief, on one hand, is that contextualised tasks are harder than symbolic computation ones. This belief can be corroborated in part by assessment data in the U.S. which reveal that elementary–school children generally perform less well on contextualised arithmetic tasks contrasted to equivalent symbolic computation problems. For instance, a Carpenter, Corbitt, Kepner, and Reys’s study (as cited in Rittle-Johnson & Koedinger, 2005) as well as a Kouba, Carpenter, and Swafford’s study (as cited in Rittle-Johnson & Koedinger, 2005) indicate that elementary–school children in the U.S. do better on symbolic computation problems than on contextualised ones.

On the other hand, contextualised tasks can be easier to solve than non-contextualised ones. Carraher, Carraher, and Schielman (1987) presented three sets of arithmetic items to 16 Brazilian third graders. The items were presented in three different contexts: (a) in a simulated store situation in which the participant played either the role of the store owner or the customer, (b) embedded in a standard application word exercise, and (c) in symbolic computation exercises. In that way, Carraher, et al. (1987) found that Brazilian
children showed significant differences in performance when they solved simulated store contexts (outside school contexts, usually presented in verbal form), than problems inside school contexts presented in written form, and symbolic computation exercises. Students performed better in solving store problems than in solving symbolic computation exercises; the average difference in facility being about 20% between store problems and symbolic computation ones.

Also differences in the way students approached the altered exercise versions were detected (Carraher et al., 1987) because in the stimulated store contexts students had to deal with money, which changed the arithmetic demand of the tasks. In that case, Carraher et al. showed that embedding problems in contextualised real-world contexts can be meaningful for students and can lead them to a greater performance.

As can be noticed, context may play a dual role capable of affecting students’ performance either negatively or positively. The next section provides insight into this issue.

2.3.1 The Context Effect: Its Differential Nature

Research studies in the field of cognition (Fiddick, Cosmides, & Tooby, 2000; Marsh, Todd, & Gigerenzer, 2004) show that in general the contexts in which tasks are presented influence the strategies that “individuals choose to solve problems and the success of those strategies” (Leighton & Gokiert, 2005, p. 2). Commonly, reasons given for the above include factors such as particular words, phrases, pictures, and item layout or format. These can be factors to which students react or respond (Girotto, 2004; Johnson-Laird, 1983; Leighton & Gokiert, 2005; Sternberg & Ben-Zeev, 2001).
In school Mathematics, in formulating mathematical representations of contextualised tasks, factors such as poor text comprehension, inability to transform a valid situational model\textsuperscript{7} into a formal mathematical model, and inability to construct appropriate links between the data affect students’ performance. These are also found to be important factors affecting accessibility to students and hence performance on these sorts of tasks (Johnson-Laird, 1983; Nathan, Kintsch, & Young, 1992; Stacey & MacGregor, 1998).

To begin with, research findings show that contextualised mathematical tasks can distract students from an item’s mathematical structure. For example, they may use commonplace ideas and argue in terms of context rather than mathematically. Boaler (1994) gives an example of the situation above. In a study of 50 students, she analysed performance on two sets of questions intended to assess the same mathematical content (equivalence of fractions) but placed in different contexts, namely: soccer season, plants planted, cutting pieces of wood, and a fashion workshop. This study revealed interestingly that females underachieved more in the context related to them (fashion) than they did on isomorphic items that were thought to be less appealing to them (such as soccer season). It should be noted that this study was carried out to understand the transfer of Mathematics from the classroom to the real-world (Boaler, 1994).

According to Boaler (1994), female students’ underachievement on the fashion item was caused by the problem context presumably distracting them from its mathematical structure. One way of interpreting Boaler’s results is that students attempted to integrate real world variables, which were not required to solve the fashion item, with the Mathematics of the task. Murphy (1990) cited in Boaler (1994), reaffirms the preceding by claiming that the context of an

\textsuperscript{7} Representation of the events presented in the problem. See Nathan, Kintsch, & Young, 1992.
assessment task may disadvantage students because they may value the circumstances that a task is presented in, having difficulty subsequently to abstract the right issues from the context.

Boaler’s (1994) results contradicted Leder’s (1974) study about sex differences in Mathematics problem appeal as a function of problem context. Four sets of parallel tasks with the same mathematical core but differing by being written in a male or female biased context were applied to 143 males and 129 females in ten coeducational high schools in Melbourne, Australia. Leder (1974) demonstrated that sex-differentiated interest and appeal in contexts could be translated to a mathematical task and be retained for male and female students; that is male students performed better in male orientated contexts and female students performed better in female orientated contexts.

Shannon (2007), cited a study from De Bock et al. (2003) which also revealed that contexts for linear equations could have a negative impact on students’ performance. These researchers also concluded that students’ close involvement with the task context may have had a negative rather than a positive influence on their performance (Shannon, 2007, p. 178).

In another study reported by Shannon (2007), she tested the same mathematical task (a linear function task) embedded in three different contexts, namely: supermarket trolleys, shopping baskets, and paper cups. She analysed how students abstracted salient features of geometry of the contexts above into variables required to solve the items but not in terms of familiarity. She claims that the mathematical modelling of everyday objects as a motivational tool is not important in these contexts, compared to the opportunities they provide for mathematical abstraction and justification. This state of opportunities “effectively means that students may not perceive the connections between
mathematical situations presented in different contexts” (Boaler, 1994, p. 6) and therefore affect general performance.

In the examination of the effect of context on students’ performance, on the other hand, better performance by students can be correlated with an involvement with the item context. Stillman (2002) in her dissertation investigated exhaustively how forty-three Years 11 and 12 students used the task context during a modelling task in selecting the Mathematics they needed initially, in keeping on track throughout the particular task they were attempting and generally in their task solving, and as a final check at the end of a task. In general, Stillman (2002) found that mathematical modelling tasks result in more understanding and thus involvement with the context than standard application tasks, and that better performance is correlated with medium to high involvement with the context; nevertheless students who perform well in Mathematics are an exception from this tendency. They also do well with low involvement.

In terms of PISA Mathematics, one specific study related to the research problem in this thesis could be found in the databases mentioned previously. Sáenz (2009) studied the role of contextual, conceptual and procedural knowledge in activating mathematical competences in PISA items with particular interest in the function of contextual knowledge. The subjects were student teachers from Spain. It should be noted that the definition of contextual knowledge in Sáenz’s study is related to the definition of context used in this thesis, hence the importance of this study to this piece of research.

One-hundred and forty pre-service teachers answered a paper and pencil test of thirty nine PISA items which were released by the OECD in 2003. The results
showed the importance of contextual knowledge in the activation of competencies to solve the most difficult items in Sáenz’s study “those which demand something more than a simple reproduction of algorithms and formulae” (Sáenz, 2009, p. 135). As can be noticed in the quote above, by the most difficult items, Sáenz is making reference to those PISA items which require multiple step algorithm use or those PISA items which require the incorporation of contextual knowledge to solve the item\(^8\).

The results of Sáenz’s study on the difficulties with contextual knowledge can be classified into two categories, namely: (1) using commonplace ideas suggested by the context, and (2) creating and communicating mathematical arguments.

On one hand, Sáenz’s study of PISA items found that PISA pictures and diagrams in the items may have an impact on students’ performance. Performance was influenced by contextual and superficial characteristics of the items (mainly visual) hindering the comprehension of mathematical concepts involved. For example, in the Robberies PISA item (item Pi-D also used in this study, see 3.3.2.4) there appears to be a big increase in the number of robberies between the years 1998 to 1999. In the item’s text there is a claim about the number of robberies made by the reporter who has been taken in by the visual appearance of the chart. Sáenz (2009) concluded that “many students’ reasoning is based on visual aspects such as the appearance of the chart, [as a result] they do not argue in terms of the percentage increase in robberies or in terms of trends” (Sáenz, 2009, p. 132). Students were also taken in by the visual appearance of the chart. The chart is shown in Appendix A, page ii, in Pi-D.

\(^8\) For more details see Table 1 on page 130 on Sáenz’s paper.
These difficulties are already highlighted by Girotto (2004), Leighton and Gokiert (2005), Johnson-Laird (1983), and Sternberg and Ben-Zeev (2001) who argued that irrelevant features not intended to give solid information such as illustrations can mislead students in their way of understanding the item context; which, according to Sáenz, are linked with insufficiencies in contextual knowledge.

Regarding category (2) above, Sáenz found that participants tended to use natural (generic) language in expressing themselves without the mathematical detail required (category 1). This issue deserves a more detailed examination because of its specific degree of variations among participants. He concluded that it is more difficult to express mathematical arguments when only using common language than when using numbers or formulae. To illustrate it, he used the following participant’s quote in the Robberies PISA item “the increase in the actual number of robberies is not a large increase” suggesting it would be a much clearer answer when expressed as “the increase is, approximately, 10/500 which is small”.

However, it is worth bearing in mind that to make answers well-matched with those that the experts have in mind it is necessary to assume that the answer of a task is shared with everyone else who knows Mathematics. In other words, the possibility to have valid answers should be a consequence of taken-as-shared interpretations which represent a basis for communication and subsequently understanding. As Johnson (1987, p. 203) claimed:

…some statements will correspond to the world more accurately, for our purposes, than others –some of them will be obviously true, others will be clearly false, and many will be problematic borderline cases. But in every case, this “correspondence” will always be relative to our understanding of our world (or present situation) and of the words we use to describe it.
Hence, Sáenz’ quote above may be a borderline case where the relative understanding of the item-context played a role. In other words, when the participant stated that the increase is not a large increase he or she might be thinking of 10/500 when communicating the answer; therefore, he or she may have assumed that it was not necessary to give more details because when everyone is exposed to this amount they should know that it is small. In terms of communicating mathematical arguments, what is trying to be explained is that interpretation of a task context is not self-evident and therefore neither are the answers. The student may have taken the numerical amount 10/500 as shared. As Cobb, Yackel, and Wood (1992, p. 10) explained:

If we assumed without question that the relationships we have in mind are in the students’ environment waiting to be perceived, our only recourse when our initial attempts to bring the relationships to their attention are unsuccessful is to be increasingly explicit and spell it out for them. In doing so, we open ourselves to the possibility that students will take form for substance and merely learn to behave in ways that convince us that they see what is considered self-evident.

To date, the unpredictable differential nature of the context effect, which may be positive or negative in contextualised items is clear. The research literature above suggests that task context can hinder students’ accessibility to the relevant mathematical model, although the results so far are variable, which leads to infer that context needs further study.

Furthermore, there are a number of related issues which should be addressed to fully understand the role of context on students’ performance. These issues are mainly related to students’ cultural background, gender (presumably because groups present particular types of interests), and individual differences in dealing with the task context. Although all of these are important, the last
preceding issue now will be discussed for its significance to the hypotheses of this study.

2.4 Students’ Dealing with the Task Context

In researching the effects of context on students’ performance, it is necessary to consider the influence of individual differences of understanding a task context. As pointed out before, the interpretation of a task context is not self-evident, nor interpreted in some universal fashion (Clarke & Helme, 1996).

Busse (2001, 2005) started a series of exploratory qualitative studies in order to systematically understand the effects of context, focussing on questions such as how an individual deals with the context and how the context given in a task is internalised. His studies centred on how upper secondary students deal individually with context aspects. The importance of these studies is that their findings establish clearly that “the context is neither an objective nor an invariable feature of the task. Students may deal very individually with the context, and it can be an object of change during the solving process” (Busse, 2005, p. 354). The importance of Busse’s studies is that he went beyond his findings to attempt to distinguish between different ideal forms of dealing with the context. According to his studies, in the process of formulating mathematical representations of contextualised tasks there are distinctive differences in dealing with the task context and mathematical world. The four types of students’ dealing with the context he distinguished are namely: Reality bound, Mathematics bound, Integrating and Ambivalent (Busse, 2005). Table 5 below summarises this approach of dealing with the context.
This typology can be a powerful instrument to understand the effects of the item-context when students solve contextualised items, and therefore will be used in this study. Additionally, this typology serves to identify what aspects students focus on when they solve the task. In other words, when students solve a contextualised task they have to deal with sometimes conflicting aspects of the item-context. By doing this, students may deal with the preceding demands in a variety of ways. Therefore, as stated before, the interpretation of the context is not self-evident, as a result, impacting on the performance of students.

Table 5
Busse’s Typology of students’ dealing with the context.

<table>
<thead>
<tr>
<th>Reality Bound</th>
<th>Mathematics Bound</th>
<th>Integrating</th>
<th>Ambivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representatives of this type consider a realistic task as fully characterised by the real problem which is described in the task. During the solving processes they use extramathematical concepts and methods. They do not mathematise the real problem and they do not apply mathematical methods in the solution process.</td>
<td>Representatives of this type regard the context of a realistic task as a mere decoration. They translate contextual expressions that are used in the task text immediately into mathematical expressions. They use only that amount of contextual information which is given in the task test. Additional personal contextual knowledge is not applied. The task must be solved exclusively by mathematical methods.</td>
<td>Representatives of this type perceive the realistic task in its real context, but they also apply mathematical methods. They use personal knowledge about the context which exceeds the contextual information given in the tasks’ text in order to mathematise the problem and validate the solution. During the solution process they apply mathematical methods.</td>
<td>Representatives of this type perceive the realistic task with its two aspects: Mathematics and Reality. But they feel ambivalent concerning the legitimacy of the way they are supposed to solve the task: internally they prefer contextually accentuated reasoning while externally they prefer a mathematical reasoning. These two ways of reasoning just coexist, they are not synthesised to form a whole.</td>
</tr>
</tbody>
</table>

Source Busse (2005, p. 355)
2.5 Implications of the Literature Review for this Study

The previous discussion has raised several issues concerning contextual influences on students’ performance which this study aims to examine in relation to PISA Mathematics. The literature review suggests that context produces a differential impact on students’ performance. The results are mixed on students’ performance with reference to the effects of contextualised tasks; it can have a positive, a negative or a neutral impact depending on what the contextual item is being compared with. Its effects on PISA Mathematics are undetermined, due to the lack of studies addressing this issue. Therefore, it is expected that this study will help to clarify the role of the task’s context on students’ uses of context.

As the literature also has suggested, the task’s context is not self-evident; that is to say, individual differences in interpreting the context may influence students’ performance. It is also expected that individual differences in responses to the task may occur in this study. As a result, the above typology of students’ dealing with the context seems to be an adequate instrument not just to understand the effect of context on students, but it is expected that it will also be linked with their subsequent performance.
CHAPTER 3: METHODOLOGY

3.1 Introduction

This section presents the research design and it is divided into three subsections. First, the rationale for the choice of the methodology will be outlined. Second, the instrumentation, data collection procedure, sample, task administration, and data analysis will be described. Finally, methodological issues which may influence this study will be delineated.

3.2 Rationale for the Use of Mixed Methodology

As was stated in 1.4, this research project is based on a mixed method approach (Creswell, 2003, 2005) due to the character of the research questions which require both quantitative (Research Questions 1, 2, and 3) and qualitative (Research Question 4) interpretations. The quantitative method in this thesis will serve to identify differences in students’ performance across PISA items and variations, while the qualitative component will serve to examine how the objective figurative context may affect students’ access to the solution process in PISA items, and therefore students’ performance.

From the paragraph above it can be inferred that the independent variable will be item-context. It is anticipated that the alteration of this independent variable may influence students’ performance on items. Therefore, the dependent variable consists of students’ performance on the items in this study.

This Research is situated within certain epistemological paradigms. As a result, an epistemology of knowledge governs the research planning and implementation process in this study. According to Cohen, Manion, and Morrison (2003) paradigms consist of belief systems and how the researchers
view and understand the world. In this regard, mixed methods researchers commonly take a pragmatist approach to research (Tashakkori & Teddlie, 1998; Teddlie & Tashakkori, 2009), emphasising that the nature of the research question should dictate methodology. It can be inferred that researchers ought to make decisions regarding data collection, analysis, and interpretation methods based on “what works” (Howe, 1988, p. 15).

Strengths and weaknesses of mixed method designs have been widely discussed in the literature (Creswell, 2003, 2005; Feilzer, 2010; Howe, 1988; Tashakkori & Teddlie, 1998; Teddlie & Tashakkori, 2009). A major tenet of pragmatism is that quantitative and qualitative methods are compatible. As a result, for this study both numerical scores and students’ responses can help to better understand the research problem in a single study. Consequently, the rationale for mixing methods is that quantitative and qualitative methods complement each other and allow for more complete analysis when used in combination (Feilzer, 2010; Tashakkori & Teddlie, 1998). Nevertheless, the most important criticism of the pragmatic approach in research is related to the adoption of what works in making decisions. Critics, such as Tashakkori and Teddlie (1998), have warned of the lack of sufficient theoretical articulation and the presence of certain imprecise descriptions in a number of pragmatist guided studies.

Accordingly, a number of basic advantages of a mixed method approach (Ivankova, Creswell, & Stick, 2006) are that its designs make it an appropriate method for this research. Advantages include:

1. Easy to implement for a single researcher, as it sequentially proceeds from one stage to another.
2. Sequential explanatory mixed method design is useful for exploring quantitative results in more detail (refer to 3.2.1), and
3. This design is especially useful when unexpected results arise from a quantitative study.

Although some limitations of this design include:
1. As any mixed method design, it requires lengthy time to complete.
2. It requires feasibility of resources to collect and analyse both type of data (Ivankova et al., 2006).

Hence, taking this into account, the researcher will attempt to define his theoretical stance as follows. For this researcher:
1. Combinations of quantitative and qualitative methods complement each other to provide a richer description of the research problem under study.
2. Data are embedded in specific settings and time frames.
3. Hypotheses can be generated and tested quantitatively, as well as qualitatively through analysis of textual data.

3.2.1 Research Design

The effect of item-context on students’ performance will be studied by examining responses to four items from the PISA survey and variations of them (described in 3.3.1). There will be two types of variations: (1) a change of item-context and (2) a diminution of item-constraints. These variations will require the same mathematical core to solve the items but altered in constraints and context.

The aim is to compare performance on items that have the same mathematical core whilst varying in the objective figurative context. Two research phases shape the design of this study.
In the first phase, the quantitative data will be collected. Potential participants will be required to complete one paper and pencil test (4 items and 30 minutes long) as part of their normal class work. As a result, Phase 1 of this thesis will serve to identify differences in students’ performance, if any, as well as recognise item-context factors that may affect students’ performance when participants solve items embedded in different contexts. It should be noted that it was not judged satisfactory to give one participant several items with the same mathematical core and different contexts. This is because it is expected that many or most students would either recognise the underplaying mathematical structure, or would not recognise it in any instance. As a result, a rotated design of three booklets was used (refer to 3.3.2) so that different students will solve problems with the same mathematical core but different contexts.

In Phase 2, a qualitative semi structured interview (15 minutes long) will be carried out and recorded with those who agree to participate. Participants will view their previous mathematical work using stimulated recall. This interview will be conducted to help explain why factors identified in the first phase may be significant factors affecting students’ performance.

It should be noted that before the two research phases mentioned above took place, a pilot study (described in 3.3.4) was needed, partly because it was necessary to observe how students reacted to the variations of context. The major outcome of this pilot study was that it showed the difficulty in getting good items.

The visual model of procedures for the design of this study is presented in Figure 1.
### Figure 1
**Visual Model for Mixed Methods Procedures (The Two Phases Research Design)**

<table>
<thead>
<tr>
<th>Phase</th>
<th>Procedure</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>PILOT STUDY</td>
<td>Non Applicable</td>
<td>Numeric Data (N=20) and students’ written comments to the items (Students’ feedback).</td>
</tr>
<tr>
<td></td>
<td>One paper and pencil test (applied to 20 volunteer students from a Year 10 class)</td>
<td></td>
</tr>
<tr>
<td>PHASE 1</td>
<td>Quantitative Data Collection</td>
<td>Numeric data (N=30) and students’ written solutions. Methodology Chapter 3.</td>
</tr>
<tr>
<td></td>
<td>SPSS quantitative software analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Students are invited to participate in phase 2</td>
<td>Descriptive statistics. Chapter 4.</td>
</tr>
<tr>
<td></td>
<td>Individual semi-structured interviews. Stimulated recall of their previous mathematical work</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Qualitative Data Collection</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thematic analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Explanation of the meaning of quantitative results. Interpretation of the meaning of the interviews.</td>
<td></td>
</tr>
<tr>
<td>PHASE 2</td>
<td>Quantitative Data Analysis of Participant responses</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Participants Invitation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Qualitative Data Collection</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interpretation of the Entire Analysis</td>
<td></td>
</tr>
</tbody>
</table>

To sum up, this subsection provided an overview of the research design and the rationale for selecting mixed methodology in addressing the research questions of the study. The next section will explain the instrumentation of each phase of this study.
3.3 Instrumentation

3.3.1 The Test Items

Three versions of items form part of this study. Four items from the PISA Mathematics assessment (Pi), four items with the same mathematical core as the PISA items but with changed context (PLi1), and four items which require the student to engage with the same mathematical core but with fewer constraints (PLi2). Hence, two PISA-like items (PLi1 and PLi2) have been written for each of the four PISA items above. These items are shown in Appendix A in full.

Four PISA items (Pi) released by the OECD in 2006\(^9\) have been selected for this study. Table 6 below summarises item names, PISA item codes, item difficulty, OECD average percentage correct (full marks), and item’s use of context. The consideration of these particular four PISA items used in this study responds to two criteria, namely: (1) the low OECD average percentage of correct (full credit) answers in these items and (2) use of context. The four PISA items selected for this study are presented in Table 6 below.

Table 6

<table>
<thead>
<tr>
<th>Items</th>
<th>PISA Item’s name</th>
<th>PISA item’s code</th>
<th>OECD average percentage correct (Full Credit)</th>
<th>Use of Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pi-A</td>
<td>Internet Chat Relay</td>
<td>M402Q02</td>
<td>29%</td>
<td>First</td>
</tr>
<tr>
<td>Pi-B</td>
<td>Rock Concert</td>
<td>M552Q01</td>
<td>35(^{10})</td>
<td>Second</td>
</tr>
<tr>
<td>Pi-C</td>
<td>Support for the President</td>
<td>M702Q01</td>
<td>36%</td>
<td>First</td>
</tr>
<tr>
<td>Pi-D</td>
<td>Robberies</td>
<td>M179Q01</td>
<td>30%</td>
<td>Second</td>
</tr>
</tbody>
</table>

Source OECD (2009)

---

\(^9\) For further details visit: www.mypisa.org

\(^{10}\) This percentage corresponds to the OECD average percentage of students obtaining full credit of this item on the PISA field trial.
The PLi1 items are characterised by the same mathematical core of the corresponding PISA item but they differ in the context in which the item is embedded maintaining the same use of context as Pi (first or second order). The context in which the PLi1 are embedded is intended to be more familiar to the students. A more familiar context in this study means that the PLi1 items are drawn from the students’ everyday experiences. For instance, PLi1-A (see Appendix A, page II) was set in the context of the 2010 Commonwealth Games. This was a well and widely advertised context through the media before and during the data collection. In the same vein, PLi1-B, the Car Park item, PLi1-C, the Shopping Centre Poll, PLi1-D, the Train Ticket Prices, are expected to be contexts that are better known than PISA items contexts. There is a public car park at the train station near the secondary school where the data were collected, as well as a shopping centre. It is also expected that PLi1-D would be a more familiar context to students as it is related to the price of train tickets.

PLi2 items were introduced to observe reported effects of the alteration of the contextual demand of the items on students’ performance. The PLi2 items are characterised by a mathematical core which is a subset of the mathematical core of the corresponding PISA and PLi-1 items but they vary in the context (with the exception of PLi2-A, refer to Table 7 for more details).

However, compared to the variation of context in PLi1 items, in PLi2 items the variation of context (and certainly in the mathematical core) produces differences in the contextual constraints (extra conditions) of the item. For instance, on PLi2-B (Tiles problems) students are presented with the dimensions of a rectangular table and tiles, and have to calculate the total number of tiles needed to tile the table. In this item, students will not have to deal only with the procedural demand of this item but also with some minor
contextual demand (space between tiles). Hence, in PLi2 items, an explicit reference to the item-context is not central because there are fewer constraints for the solution which makes the solution space smaller.

It should be noted that the inclusion of PLi1 and PLi2 was made to observe whether or not the introduction of alterations (contexts, constraints) may increase, decrease or maintain the same level of the contextual demand of the item and as a result its difficulty. It is predicted that PISA-like items would elicit different contextual demands for PLi1 and PLi2 in students which would demonstrate that item-context is a crucial dimension to take into consideration in Mathematics assessments. Finally, later reflection revealed that the actual use of context was different to the intended use (see PLi2-B and PLi2-C). As a result, all PLi2 items were not considered neither for hypothesis testing in Research Question 3 nor explored through qualitative data in Chapter 4.

To sum up, every version has four items; therefore twelve items are used in this study. It should be noted that every item corresponds to a version (Pi, PLi1, or PLi2) and to a family of items; that it is to say, each is a sibling of two others. To exemplify this, Pi-A, PLi1-A, and PLi2-A are siblings.

### 3.3.2 Characterisation and Differences among Family of Items

In this section the twelve items used in this study will be described explaining differences in item context, item complexity, item solution as well as alterations. A summary is shown for each item version in Table 7. Information from this table will be referred to throughout 3.3.2.1 to 3.3.2.4; as a result information in this table will be not described in detail at this point. Appendix A presents the twelve items used in this study.
### Table 7
*The twelve items used in the study*

<table>
<thead>
<tr>
<th>PISA Items</th>
<th>PLi1-A</th>
<th>PLi1-B</th>
<th>PLi1-C</th>
<th>PLi1-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pi-A</td>
<td>Internet chat relay</td>
<td>2010 Commonwealth Games</td>
<td>Support for a new shopping centre poll</td>
<td>Train ticket prices</td>
</tr>
<tr>
<td>Pi-B</td>
<td>Rock concert</td>
<td>Car park</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pi-C</td>
<td>Presidential poll</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pi-D</td>
<td>Robberies</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| PLi2 | | | |
| Pi2-A | Internet chat relay | | |
| Pi2-B | Tiles | | |
| Pi2-C | Polls | | |
| Pi2-D | Sea surface temperature | | |

<table>
<thead>
<tr>
<th>Difference from PISA Items</th>
<th>Use of Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>First</td>
</tr>
<tr>
<td>None</td>
<td>Second</td>
</tr>
<tr>
<td>None</td>
<td>First</td>
</tr>
<tr>
<td>None</td>
<td>Second</td>
</tr>
<tr>
<td>None</td>
<td>First</td>
</tr>
<tr>
<td>None</td>
<td>Second</td>
</tr>
<tr>
<td>None</td>
<td>First</td>
</tr>
<tr>
<td>None</td>
<td>First</td>
</tr>
<tr>
<td>None</td>
<td>Zero</td>
</tr>
<tr>
<td>Context</td>
<td>Second</td>
</tr>
</tbody>
</table>

### 3.3.2.1 Family-A Items

Pi-A requires students to identify the relevant Mathematics to solve a non-routine, but simple problem and use different representations (OECD, 2009a) using time differences between two cities, namely Sydney and Berlin. This item highlights the importance of taking into account contextual information found in the item to check the final answer. This item emphasises a “first-order” use of context because the context is relevant and needed for solving the problem and judging the answer. It is not “second-order” because students do not have to deal with any external contextual information. The students have to understand the way that time spent sleeping and at school constrains the times that could be suitable for two students to communicate with each other (OECD, 2009a).

On the other hand, PLi1-A has the same mathematical core of Pi-A but it is embedded in a different context, namely the 2010 Commonwealth Games. This change of context is made to observe difference in performance, if any, when students solve this more current item. In PLi1-A students have to understand the way that the time difference between Melbourne and Delhi constrains the
times that could be suitable for students to communicate with each other. As with Pi-A, this item also highlights a “first-order” use of context because the context is relevant and needed for solving the problem and judging the answer, that is to say students can check and verify appropriately whether their solutions fit the task’s requirements. It was judged that for these students, the mathematical core of Pi-A and PLi1-A was the same. It is assumed that 15 year old students can work with a time difference of 9 hours or 4 hours and 30 minutes equally well.

Finally, PLi2-A is embedded in the same context as Pi-A but it has fewer explicit context constraints. Therefore, students may take into account by themselves restrictions to the solution such as sleep time but they may not. Thus, it is expected that students do go back to the problem to verify appropriateness of the solution stressing in that way the importance of the first-order use of context in this item.

3.3.2.2 Family-B Items

On Pi-B, participants are presented with the dimensions of a rectangular field, and have to mathematise the space occupied by a person in order to estimate the number of rock fans that could be accommodated (OECD, 2009a). This item emphasises the “second-order” use of context because the answer does depend on contextual assumptions made by students and as a result, this item could have more than one possible correct answer if realistic assumptions about the number of people that can be accommodated per square meter are not made. The item’s written statement does not include all data; it has to be brought in from context by the student.

On PiL1-B, students are presented with the dimensions of a Cark Park, and have to mathematise the space occupied by a car in order to estimate the total
number of cars that could be accommodated. The reference to the context that students need to solve this item is hypothesised to be more accessible to them than the rock concert item. Here, students may have a clear idea of what a full car park looks like. It is expected that students will be involved with an interpretation of the context in order to produce a relevant mathematical representation of the item. Consequently, the second-order use of context is hypothesised to be less demanding because of students' greater familiarity with the context. This would be observable in their performance.

Before continuing, it should be noted that the familiarity to students of the PISA-like items used in this study is a 'supposed familiarity', that is to say, there was not considered a questionnaire for students to respond about if the context of the PISA-like items used in this study are in fact more familiar contexts for the students after all. It can have an effect in the conclusions reached for this thesis (see 5.2).

It should be noted that both Pi-B and PLi1-B have a very similar mathematical model, namely: the area of a rectangular shape and the requirement to use a rate (for instance number of people per square metre) to convert area to a number. Mathematically they are equivalent but they have different contextual demands mathematising the space occupied by a person and a car respectively. Hence, they have the same mathematical core.

Lastly, in PLi2-B students are presented with the dimensions of a rectangular table as well as the tiles, and have to calculate the total number of tiles needed to tile the table; students may not have any difficulty in solving it correctly because they will not have to deal with any contextual demand but only with the procedural demand of this item. Hence, this item requires a first-order use
of context. Second-order use of context is only relevant if students allow space between tiles for grouting.

3.3.2.3 Family-C Items

Pi-C \(^{11}\) requires students to: understand the text, conceptually understand different aspects of sampling, produce and explain the reasons for choosing the answer given (OECD, 2009a). Successful students will answer justifying their conclusion giving the correct answer with two valid reasons. There are three important characteristics to evaluate the samples in the question: the survey should be taken from a large sample, it should be a random sample, and of course only respondents who are eligible to vote should be considered. This last aspect reinforces the interaction between participants and the “first-order” use of context.

PLi1-C varies from Pi-C only in the change of context; this item is embedded in a support for a shopping centre poll to observe difference in performance, if any, when students solve this item. It is hypothesised to be a context that is more familiar to students than elections. Finally, on PLi2-C there are two important characteristics to evaluate: the survey should be taken from a large sample; and it should be a random sample. Hence, these items have the same mathematical core (students in both items have to evaluate the survey should be taken from a large sample, it should be a random sample, and of course only respondents who are eligible to vote should be considered). This is a no context constraints-item; students will not have to deal with any contextual demand. That is to say, the context given for this item is stated simply as a set of opinion polls which is an implicit context to randomness and sample size. Thus, there are not context

---

\(^{11}\) Pi-C is a modified version of the original PISA item. In the modified version, the date at which the poll was made is not included.
constraints, and the students need to know that only two variables are relevant (sample size and sample method) in order to answer the item. Taking into account the above, the context given here is only a camouflage for answer the item. As a result, students only have to deal with the conceptual demand of this item, hence this item can be classified as zero-order use of context.

3.3.2.4 Family-D Items
On Pi-D and PLi1-D, students are required to analyse a graph and interpret the data involved. The items require students to understand and decode a graphical representation in a critical way, make judgments and find appropriate argumentation based on mathematical thinking and reasoning (interpretation of data), use some proportional reasoning in a statistical context, and communicate their reasoning process effectively (OECD, 2009a). Thus, these items have the same mathematical core. The difference between Pi-D and PLi1-D is the context in which the item is embedded (see Table 7 and Appendix A). The change of context was made to observe if the same judgements are present when students are exposed to a money context. This was hypothesised to be more familiar to students than the robberies context.

Finally, PLi2-D presents a modified version of Pi-D to observe if the same judgements are present when the item context is a Sea Surface Level Temperature context. Some students may answer this question based on their contextual knowledge (climate change for instance) rather than on the increase of temperature over one year period. The regularity with which climate change issues are debated in recent years means that this context is surely to be very familiar to students.
3.3.3 Rotated Test Design

As was discussed in 3.2.1 and 3.3.1 twelve items form part of this study. It should be noted that no student did both an item and its sibling items, so that direct comparison of performance at the level of the student will be not possible. Instead a rotated design of three booklets was used. Each participant dealt with a third of the total items. Table 8 presents the items’ arrangement in the three booklets issued to participants.

Table 8

<table>
<thead>
<tr>
<th>Items arrangement per booklet</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Booklet A</strong></td>
<td>Internet Chat Relay (Pi-A)</td>
</tr>
<tr>
<td></td>
<td>Car Park (PLi1-B)</td>
</tr>
<tr>
<td></td>
<td>Polls (PLi2-C)</td>
</tr>
<tr>
<td></td>
<td>Sea Surface Temperature (PLi2-D)</td>
</tr>
<tr>
<td><strong>Booklet B</strong></td>
<td>Internet Chat Relay (PLi2-A)</td>
</tr>
<tr>
<td></td>
<td>Support for a new Construction (PLi1-C)</td>
</tr>
<tr>
<td></td>
<td>Robberies (Pi-D)</td>
</tr>
<tr>
<td></td>
<td>Tiles (PLi2-B)</td>
</tr>
<tr>
<td><strong>Booklet C</strong></td>
<td>2010 Commonwealth Games (PLi1-A)</td>
</tr>
<tr>
<td></td>
<td>Rock Concert (Pi-B)</td>
</tr>
<tr>
<td></td>
<td>Price of Train Tickets (PLi1-D)</td>
</tr>
<tr>
<td></td>
<td>Support for the President (Pi-C)</td>
</tr>
</tbody>
</table>

3.3.4 Pilot of the Items

A group of 20 volunteer students from a Year 10 class in Melbourne, Australia, agreed to pilot the early versions of the twelve items in June and July 2010. The main purpose of the pilot was to gain feedback on how students may react to versions embedded in different contexts as well as to identify difficulties with items’ text and layout. The volunteer students answered the items using the rotated test design above. In general, the piloting of the items revealed, on one hand, that students were on average more successful in PLi2 items compared to Pi and PLi1 items. The mean of full credit answers was 1.1 on PLi2, 0.92 on PLi1, and 0.85 on Pi from a maximum of 2 point scores. On the other hand, from the point of view of Research Question 3, the second-order use of context seemed to
be more affected in the mean performance when an alteration of context was introduced on items versions ($\bar{X}_{\text{first-order}} = 0.98$ and $\bar{X}_{\text{second-order}} = 0.66$).

After the piloting, items were checked and amended. A final check by the supervisor of this study was made in order to ensure that the twelve items were mathematically isomorphic where required. It should be noted that although the tasks were in a strict one to one mathematical correspondence, considering the same amount of information and visual aspects (when needed). Sources of variation rather than context could not be entirely controlled. Obvious sources of variation include wording of each item as well as different figures on items and/or graphs. For example, in items of family-D different scales are used across different contexts in which the item is embedded.

### 3.4 The Marking Scheme

The participants’ responses to the paper and pencil test will be marked in two different forms: (1) according to the relevant PISA marking guidelines (OECD, 2009c) and (2) according to an adapted marking scheme specially designed to provide more discrimination between responses, taking into account the possible effect of context on the students’ answers. In other words, if the students’ responses are related to contextual misconceptions suggested by the item-context, the item will be marked according to the students’ commonplace ideas. Both marking schemes consider a full credit score of 2 points, a partial credit of 1 point, and a no credit score of 0 points. The PISA marking scheme allocated a code of 1 to these full credit answers which do not have partial credit available. In order to be able to combine scores, scores on these items were doubled.
Finally, it should be noted that the marking scheme was not piloted. The complete scales and descriptions are included in Appendix B, but excerpts from the scales are shown below in Table 9 for the purpose of illustration.

Table 9  
Marking schemes for Family B-items

<table>
<thead>
<tr>
<th>Items</th>
<th>PISA Guidelines</th>
<th>Adapted marking scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full Credit (2 points)</td>
<td>Partial Credit (1 point)</td>
</tr>
<tr>
<td>Pi-B</td>
<td>Successful participants chose alternative C</td>
<td>Non applicable</td>
</tr>
<tr>
<td>PLi1-B</td>
<td>Successful participants chose alternative B</td>
<td>Non applicable</td>
</tr>
<tr>
<td>PLi2-B</td>
<td>Successful participants chose alternative D.</td>
<td>Non applicable</td>
</tr>
</tbody>
</table>

3.5 Data Collection Procedure, Participants, and Task Administration

3.5.1 Procedure

In terms of data collection procedures, Public, Catholic, and Private Schools in Melbourne, Australia were invited to participate in this study. They were contacted via email. The Ethic Approval resolution and the Plain Language Statement were attached to the email. As was established before, the paper and
pencil test was administered to a whole class. Results for all students in the class were given to the teacher regardless of whether students agreed to participate in the study’s data collection. A letter was sent to the home of potential participants and their parents/guardians from selected classes, inviting the students to participate in all the data collection procedures (Phase 1 and 2 of the data collection).

3.5.2 Participants

The target population in this study was students approximately 15 years old, embedded in a Year 10 Mathematics class. The decision to study the mathematical work of 15 year old students is because PISA is administered at this age. It is fundamental to note that only participants with adequate English language skills were invited to take part for the whole study. Participants with cognitive impairment, an intellectual disability, a mental illness or Aboriginal Torres Strait Islander Peoples could be a possibility in the sample but not an audience this study was actively targeting.

3.5.3 Task Administration

The testing of participants was undertaken between September 10 and September 17, 2010. Thirty participants in two Year 10 classes agreed to take the paper and pencil test. Participants were tested in a class period. Three booklets (A, B, and C) were issued, so each participant dealt with a third of the total items. The booklets were ordered by the researcher as follow: Booklet A, Booklet B, and Booklet C. Once ordered in that configuration, they were distributed to participants in that particular order. Participants had 30 minutes to solve the items.
3.6 Students’ Interview

One of the major interests of this study is to understand the connection between item-context and students’ performance. While this connection can be directly observable through numerical scores, clarification for the students’ written responses needs to be executed if the research problem in this study is to be fully acknowledged. It is believed by the researcher of this thesis that the interview would offer students the opportunity to comment on contextual factors, if any, which may have affected their answers. The stimulus material for students who agreed to participate in the interview and the structure of the interview are detailed below.

3.6.1 Stimulus Material Provided and Structure of the Interview

On one hand, students’ verbal data was stimulated by the provision of their paper and pencil test. Each paper and pencil test was provided to students five minutes before the interview to consider their written answers to the items to which they would be required to respond. The interview took place 2 days after students answered their paper and pencil test.

On the other hand, the interview was conducted working with individual students. A section of their previous mathematical work was re-shown to the students. The interview was structured in relation to three major aspects, namely: (a) each item’s mathematical procedure, (b) each item’s nature of complexity (contextual), and (c) each item’s influence of the item-context in the solution process. Questions pertaining to the process of solving items in the paper and pencil test were asked to the participant. Table 10 below presents the questions used in the interview.
Table 10

Questions used in the Students’ Interview

- What do you notice about your explanation on your previous work?
- How useful was the context to your solution in this particular question?
- Would that have made a difference?
- Did you consider the context of the question when solving the question?
- Could you relate to it?
- What were you thinking at this stage?
- Were you confident that this method would work?
- Have you ever dealt with this context in another way? In newspapers, TV, etc?
- Do you consider that previous exposure to a particular context has influenced the way you answered this item? How? Why?
- If the same thing happened to you, would you behave in the same way in real life as you did in solving the problem? Why?

3.7 Data Analysis

Research Questions 1, 2, and 3 determined the choice of statistical analysis to be used in this study. The data collected was subjected to a quantitative analysis in order to compare students’ performance. The quantitative component aimed to characterise students’ responses “along traits of interest related to the research question” (Creswell, 2003, p. 227) to inform how item-context factors may influence students’ performance.

The purpose of the analysis in the first phase was determined by the difference in the mean of the students’ performance using Pi, PLi1, and PLi2 items as well as in uses of context. As a result, descriptive statistics, ANOVA tests as well as an analysis of absolute mean difference, was conducted. All statistical analysis of the quantitative phase was conducted with the help of the Statistical Package for Social Sciences Software (SPSS™). The results of this analysis will be reported in the next chapter.

Using a different approach, the second qualitative phase will focus on explaining the results of the first phase for Research Question 4. All the interview transcripts were analysed for either common or different themes in
order to illustrate the extent to which the nature of contextualised items have a similar or dissimilar effect on students’ performance (see 3.6).

Finally, it is expected that both the quantitative and qualitative analysis of the written students’ work as well as the ten interviews will provide a new body of data for the study of students’ performance on contextualised tasks.

### 3.8 Methodological Issues

This study has methodological limitations which may adversely influence the extent to which generalisable conclusions can be drawn. The purpose of this section is to address limitations as well as to state ethical considerations for this study.

#### 3.8.1 Student Sample Size and Potential for Generalisation

The student sample size in this study is moderate. Because of this, advantages and disadvantages can impact the methodology that governs this study. On the one hand, statistical analysis can be performed on the students’ numerical scores because the number of students’ answers is large enough for the quantitative data collection ($N_1=30$). Additionally, the sample size for interviewing students is small enough ($N_2=10$) to allow a detailed analysis of the interview transcripts, an important requirement for the examination of item-context factors affecting students’ performance.

On the other hand, clearly it is not possible or anticipated that the results and conclusions related to this sample will be definitive in terms of the effect of item-context on students’ performance. Nevertheless, a small study like this can demonstrate the extent to which this differential nature applies to a particular sample. In other words, this sample size has the capability to identify trends on students’ performance. The recognition and explanation of counter-examples to
a particular trend, which this study may be able to achieve, can raise valuable questions about the differential nature of item-context as well as stimulate further research involving this issue.

3.8.2 Ethical Considerations

Ethical issues were addressed at each phase of this study. The permission for conducting research has been obtained from The Melbourne Graduate School of Education Human Ethics Advisory Group (MGSE HEAG) on the 28th of April 2010 (Ethics ID: 1033514). The anonymity of all participants has been protected by numerically coding each paper and pencil test and keeping names confidential. Additionally, while conducting the individual semi-structured interviews, participants will be assigned pseudonyms for use in the description and reporting the results. Participants will be told summary data may be disseminated to the academic community; however in no manner will it be possible to trace participants.

Another important aspect to take into consideration in this section is the role of the researcher in the data collection phases. In the first, quantitative phase, the researcher will collect and mark the data using the collaboration of the classroom teacher. However, in the second, the qualitative phase, due to the nature of the data collection method the researcher will assume a more participatory role. Thus, the possibility for subjective interpretations and potential bias for interpreting the transcripts will have a minor impact on this study.

3.9 Summary of Hypotheses and Statistical Procedures

Table 11 lists the hypotheses and summarises the data analysis procedures for each research question in this study.
Table 11

Summary of hypotheses and data analysis procedures for testing them

<table>
<thead>
<tr>
<th>Effect of a more familiar and accessible context on students’ performance</th>
<th>Data Source</th>
<th>Analysis Type</th>
</tr>
</thead>
</table>
| (a) Is the students’ performance on PLi112 different to that on Pi13? | • Students’ scores on the four versions of PLi1.  
• Students’ scores on the four versions of Pi. | • Comparison of mean scores |

H₀: Performance on PLi1 = Performance on Pi  
H₁: Performance PLi1 ≠ Performance on Pi

<table>
<thead>
<tr>
<th>Effect of fewer constraints in the items-context on students’ performance</th>
<th>Data Source</th>
<th>Analysis Type</th>
</tr>
</thead>
</table>
| (b) Is performance on PLi214 different to that on Pi? | • Students’ scores on the four versions of PLi2  
• Students’ scores on the four versions of PISA items. | • Comparison of mean scores |

H₀: Performance on PLi2 = Performance on Pi  
H₁: Performance PLi2 ≠ Performance on Pi

<table>
<thead>
<tr>
<th>Effect of the nature of the use of context on students’ performance</th>
<th>Data Source</th>
<th>Analysis Type</th>
</tr>
</thead>
</table>
| (c) Are the items which highlight the second-order use of context more affected by change of context than items which highlight the first-order use of context? | • Students’ scores on the two versions of PLi1 which highlight the first and second order use of context.  
• Students’ scores on the two versions of Pi which highlight the first and second order use of context. | • Absolute difference on the means of items which highlight the first and second order use of context. |

H₀: first order | PLi1(Av)–Pi(Av) | = second order | PLi1(Av)-Pi(Av) |
H₁: first order | PLi1(Av)-Pi(Av) | ≠ second order | PLi1(Av)-Pi(Av) |

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Data Source</th>
<th>Analysis Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d) To what extent does the nature of the item-context affect students’ performance on items which highlight the first and second order use of context?</td>
<td>• Written records of the items’ solutions and students’ interviews.</td>
<td>• The focus will be put on the students’ written and verbal records of the items’ solutions.</td>
</tr>
</tbody>
</table>

No formal hypotheses, although it was expected that the nature of item-context would affect students’ performance on those items which highlight the second-order use of context.

---

12 PLI1: PISA-Like items embedded in a more appealing context for students  
13 Pi: PISA items  
14 PLI2: PISA-Like items embedded in an altered context with fewer constraints
CHAPTER 4: RESULTS

4.1 Introduction

The results of this study will be considered in relation to the research questions as well as to the hypotheses in Table 11. In this manner, Section 4.2 presents the results of the quantitative phase of this study; that is to say, overall performance on Pi, PLi1, and PLi2 is provided in this section. On top of this, Section 4.3 provides the average performance corresponding to the uses of context (first and second order use of context). It should be noted that in Section 4.2 to 4.3 both PISA marking guidelines and its corresponding adapted marking scheme will be used to assess students’ performance across items and the uses of context. Finally, Section 4.4 presents the results of the qualitative phase of this study.

4.2 Results of the Quantitative Phase

Tables 12 and 13 present summaries of the distribution of the students’ average scores in each item, in percentages, with both the official PISA marking scheme (Table 12) and the adapted marking scheme (Table 13) which was designed to provide more discrimination between responses (see section 3.3). Information from these tables will be referred to throughout 4.2 to 4.4.

Both Tables show the three versions of items discussed in section 3.3: items from the PISA Mathematics assessment (Pi), items which are as similar as possible to the PISA items but with changed context (PLi1), and items which require the student to engage with fewer constraints and also with an altered context (PLi2); therefore twelve items form part of these tables. It should be noted that every item is a version (Pi, PLi1, or PLi2) of a similar item; that is to
say, each item is a sibling of two others (for instance, Pi-A, PLi1-A, and PLi2-A are all siblings and make up a family of items).

Each table reports the quantitative differences between the items, its sibling items, and the type of context use as well as the number of student written responses (10 per item version). Note that in the PISA marking scheme, partial credit was not available for Pi-A and Pi-B and their siblings. For these two items, the full credit PISA code of 1 score point was re-scaled to be 2 score points. Lastly, it should be noticed that in order to make straight comparisons of the figures and graphs to be obtained in the statistical analysis, in the PISA marking scheme the full credit was scored as 2 on each item.

In Table 12 is presented the distribution of students’ scores (in percentages) sample size, items versions, items characteristics, items means and items standard deviations with the PISA marking scheme across all the item versions. Before continuing is imperative to inform the reader how the figures in Table 12 were calculated. Firstly, the ‘Item Mean’ column presents the items means which were calculated taking into account the full and partial credit answers (from a maximum of 2 score points\textsuperscript{15}). In the same manner, the ‘Item Standard Deviation’ column shows the items standards deviations which were calculated considering the full and partial credit answers (from a maximum of 2 score points). Secondly, in Table 12 are incorporated ‘Average Performance’ rows. These rows are, in one hand, associated to the mean (in percentages) of full credit, partial credit, and no credit answer across each item category; hence, that these rows do not all sum to 100%. In the other hand, these rows are also linked to the mean of item means and item standard deviations (which were calculated from a maximum of 2 score points) across each item category. It should be

\textsuperscript{15} See the marking scheme of the items in section 3.4.
noticed this way of computation of figures in Table 12 is also applicable for Table 13.

To set the performance of these students in perspective, on the items from PISA Mathematics assessment used in this study (Pi-A, Pi-B, Pi-C, and Pi-D), these students showed better average performance compared with the average performance observed in the participating countries on the PISA 2003 survey (see Table 6). To be precise, the OECD average percentage correct (full marks) on the items above was 29%, 35%, 36%, and 30% respectively (OECD, 2009b) whereas in this study these percentages were 50%, 50%, 60%, and 20% correspondingly (see Table 12).

In general, when clustering all the item versions employed in this study into three groups: PISA items (Pi), PISA-Like items 1 (PLi1), and PISA-Like items 2 (PLi2), performance was superior on PLi2. This was predicted for Research Question 2 but will be analysed more carefully later. When the adapted marking scheme was used, the overall performance across Pi, PLi1, and PLi2 does not show large changes from the PISA marking scheme, but the percentage of no credit answers decreased between the PISA and the adapted marking scheme (see Tables 12 and 13). This was because the adapted marking scheme provided more discrimination between responses which required contextual information for their solution, and also because in the adapted marking scheme, partial credit was available for all Pi and PLi1 items.

\[16\text{ This percentage corresponds to the OECD average percentage of students obtaining full credit percentage of this item on the PISA field trial.}\]
Table 12
Distribution of the students’ scores, in percentages, sample size, item characteristics, item mean and item standard deviation (from a maximum of 2 score points) with the PISA marking scheme across all items

<table>
<thead>
<tr>
<th>Items</th>
<th>N</th>
<th>Differences from PISA items</th>
<th>Use of Context</th>
<th>Full Credit (2)</th>
<th>Partial Credit (1)</th>
<th>No Credit (0)</th>
<th>Item Mean</th>
<th>Item Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pi-A</td>
<td>10</td>
<td>None</td>
<td>First</td>
<td>50%</td>
<td>-</td>
<td>50%</td>
<td>1.00</td>
<td>1.05</td>
</tr>
<tr>
<td>Pi-B</td>
<td>10</td>
<td>None</td>
<td>Second</td>
<td>50%</td>
<td>-</td>
<td>50%</td>
<td>1.00</td>
<td>1.05</td>
</tr>
<tr>
<td>Pi-C</td>
<td>10</td>
<td>None</td>
<td>First</td>
<td>60%</td>
<td>0%</td>
<td>40%</td>
<td>1.20</td>
<td>1.03</td>
</tr>
<tr>
<td>Pi-D</td>
<td>10</td>
<td>None</td>
<td>Second</td>
<td>20%</td>
<td>40%</td>
<td>40%</td>
<td>0.80</td>
<td>0.78</td>
</tr>
<tr>
<td>Total Pi</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average Performance in Pi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>45%</td>
<td>20%</td>
<td>45%</td>
<td>1.00</td>
<td>0.96</td>
</tr>
<tr>
<td>PLi1-A</td>
<td>10</td>
<td>Context</td>
<td>First</td>
<td>50%</td>
<td>-</td>
<td>50%</td>
<td>1.00</td>
<td>1.05</td>
</tr>
<tr>
<td>PLi1-B</td>
<td>10</td>
<td>Context</td>
<td>Second</td>
<td>20%</td>
<td>-</td>
<td>80%</td>
<td>0.40</td>
<td>0.84</td>
</tr>
<tr>
<td>PLi1-C</td>
<td>10</td>
<td>Context</td>
<td>First</td>
<td>50%</td>
<td>40%</td>
<td>10%</td>
<td>1.40</td>
<td>0.69</td>
</tr>
<tr>
<td>PLi1-D</td>
<td>10</td>
<td>Context</td>
<td>Second</td>
<td>0%</td>
<td>30%</td>
<td>70%</td>
<td>0.30</td>
<td>0.48</td>
</tr>
<tr>
<td>Total PLi1</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average Performance in PLi1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30%</td>
<td>35%</td>
<td>52.5%</td>
<td>0.77</td>
<td>0.89</td>
</tr>
<tr>
<td>PLi2-A</td>
<td>10</td>
<td>Fewer constraints</td>
<td>First</td>
<td>50%</td>
<td>-</td>
<td>50%</td>
<td>1.00</td>
<td>1.05</td>
</tr>
<tr>
<td>PLi2-B</td>
<td>10</td>
<td>Context and fewer constraints</td>
<td>First</td>
<td>90%</td>
<td>-</td>
<td>10%</td>
<td>1.80</td>
<td>0.63</td>
</tr>
<tr>
<td>PLi2-C</td>
<td>10</td>
<td>Context and fewer constraints</td>
<td>Zero</td>
<td>60%</td>
<td>0%</td>
<td>40%</td>
<td>1.20</td>
<td>1.03</td>
</tr>
<tr>
<td>PLi2-D</td>
<td>10</td>
<td>Context and fewer constraints</td>
<td>Second</td>
<td>40%</td>
<td>0%</td>
<td>60%</td>
<td>0.80</td>
<td>1.03</td>
</tr>
<tr>
<td>Total PLi2</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average Performance in PLi2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>60%</td>
<td>0%</td>
<td>40%</td>
<td>1.20</td>
<td>0.99</td>
</tr>
</tbody>
</table>

- Partial credit not available on Pi-A and Pi-B and its sibling items on the PISA marking scheme.
Table 13
Distribution of the students’ scores, in percentages, sample size, item characteristics, item mean and item standard deviation (from a maximum of 2 score points) with the Adapted marking scheme across all items

<table>
<thead>
<tr>
<th>Items</th>
<th>N</th>
<th>Differences from PISA items</th>
<th>Use of Context</th>
<th>Full Credit (2)</th>
<th>Partial Credit (1)</th>
<th>No Credit (0)</th>
<th>Item Mean</th>
<th>Item Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pi-A</td>
<td>10</td>
<td>None</td>
<td>First</td>
<td>50%</td>
<td>10%</td>
<td>40%</td>
<td>1.10</td>
<td>0.99</td>
</tr>
<tr>
<td>Pi-B</td>
<td>10</td>
<td>None</td>
<td>Second</td>
<td>50%</td>
<td>50%</td>
<td>0%</td>
<td>1.50</td>
<td>0.52</td>
</tr>
<tr>
<td>Pi-C</td>
<td>10</td>
<td>None</td>
<td>First</td>
<td>60%</td>
<td>20%</td>
<td>20%</td>
<td>1.40</td>
<td>0.84</td>
</tr>
<tr>
<td>Pi-D</td>
<td>10</td>
<td>None</td>
<td>Second</td>
<td>20%</td>
<td>60%</td>
<td>20%</td>
<td>1.00</td>
<td>0.66</td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.25</td>
</tr>
</tbody>
</table>

| PLi1  |     |                             |                |                |                   |              |           |                         |
| PLi1-A| 10  | Context                     | First          | 60%            | 20%               | 20%          | 1.40      | 0.84                   |
| PLi1-B| 10  | Context                     | Second         | 30%            | 10%               | 60%          | 0.70      | 0.94                   |
| PLi1-C| 10  | Context                     | First          | 50%            | 50%               | -            | 1.50      | 0.52                   |
| PLi1-D| 10  | Context                     | Second         | 40%            | 40%               | 20%          | 1.20      | 0.78                   |
| Total | 40  |                             |                |                |                   |              |           | 1.20                   |

| PLi2  |     |                             |                |                |                   |              |           |                         |
| PLi2-A| 10  | Fewer constraints           | First          | 50%            | 40%               | 10%          | 1.40      | 0.69                   |
| PLi2-B| 10  | Context and fewer constraints| Second         | 90%            | 0%                | 10%          | 1.80      | 0.63                   |
| PLi2-C| 10  | Context and fewer constraints| First          | 60%            | 40%               | -            | 1.60      | 0.51                   |
| Pli2-D| 10  | Context and fewer constraints| Second         | 30%            | 30%               | 40%          | 0.90      | 0.87                   |
| Total | 40  |                             |                |                |                   |              |           | 1.42                   |

- Partial credit not available on PLi1-C and PLi2-C items on the adapted marking scheme.

4.2.1 Effect of a more Familiar Context

This section examines the effect of embedding items in a more appealing context. It compares overall performance on PISA Items (Pi) to overall performance on items embedded in a more appealing context for students (PLi1). Figure 2 below shows how an alteration of context made easier Pi-C and its sibling version PLi1-C for students, whereas an alteration of context made difficult Pi-B and Pi-D for students across their corresponding sibling versions, namely: PLi1-B and PLi-D. When the adapted marking scheme was used; an alteration of context made easier Pi-A, C, and D for students but this alteration of context made difficult Pi-B for students. Recall from section 3.5.3 that no student did both an item and its sibling items, so that direct comparison of performance at the level of the student is not possible. Hence, this can only be
through means. When framing Research Question 1, it was expected that the alternative context would lead to higher success rate.

**Figure 2**
Mean of students' performance by item on Pi and PLi1 versions

Table 12 shows that the overall mean score was better on Pi items than on PLi1 items under the PISA marking guidelines; nevertheless no statistical significance was found in the analysis of the ANOVA test ($F(1,78) = 1.179, p = 0.281$); thus there is insufficient support for the experimental hypothesis and therefore the experimental hypothesis is rejected and the null hypothesis that performance was equivalent on Pi and PLi1 is accepted. In Table 13 it can be seen similar differences were observed between Pi and PLi1 items under the adapted marking scheme; but when comparing their means also no statistical significance was observable using a one-way ANOVA.
(F (1, 78) = 0.078, p = 0.781), which means there is insufficient support for the experimental hypothesis and therefore the experimental hypothesis is rejected and the null hypothesis that performance was equivalent on Pi and PLi1 is accepted under this marking scheme.

It seems from Figure 2 above that the statistical tests have been strongly influenced by the large difference in performance between item Pi-B and its sibling PLi1-B. With the adapted marking scheme, performance on the other three items was improved by the change of context, as had been hypothesised. Possible reasons for the negative effect of change for PLi1-B will be discussed in section 4.4 when the qualitative data is discussed.

4.2.2 Effect of Fewer Constraints in the Item
This section presents the effect of fewer constraints in the items. It compares overall performance on PISA items (Pi) contrasted to overall performance on items embedded in an altered context with fewer constraints (PLi2). Figure 3 below presents the mean of students’ performance on Pi and PLi2 items graphically; the mean of students’ performance remained constant on Pi-A, Pi-C, and Pi-D, and their sibling versions whilst the mean was influenced positively on Pi-B and its sibling version.

When students’ responses were marked under the adapted marking scheme, the trend above did not hold; that is to say, Pi-A, Pi-B, and Pi-C, and their sibling version were influenced positively by the change of context with fewer constraints whereas with Pi-D and its sibling version the influence was the opposite.
As noted above, and as will be discussed in Chapter 5, the decision to alter both the context and number of constraints in these items has proved to be a methodological error. However, the time limitation of this research project did not enable data collection with new items.

Figure 3
Mean of students’ performance by item on Pi and PLi2 versions

As was predicted in Chapter 3 and is shown in Tables 12 and 13, students performed better overall on PLi2 than on Pi items on both the PISA and the adapted marking scheme. However, this does not mean that significance of results was observable; in fact, the ANOVA test for the overall performance of students on these items under the two marking schemes showed no significant difference in students’ performance (PISA marking scheme:
\( F(1, 78) = 0.839, \quad p = 0.363 \); Adapted marking scheme \( F(1, 78) = 1.005, \quad p = 0.307 \) respectively). Hence, there is insufficient support for the experimental hypothesis and therefore the experimental hypothesis is rejected and the null hypothesis stands under this marking scheme.

The Null hypothesis that performance was equivalent on Pi and PLi2 cannot be rejected on this sample.

**4.3 The Uses of Context and Overall Performance**

Research Question 3 attempts to explore whether changes of context in items requiring second order use of context have more effect on students’ performance than in first order use items. Therefore, analysis of mean absolute difference in mean score by item and uses of context will be executed to explore this possible relationship.

As was discussed in Chapter 3, PLi2 differs from Pi and PLi1 not just in context but also in constraints; PLi2 items have fewer constraints and therefore it was expected that the rate of Full Credit answers would be increased in relation to Pi and PLi1 because the mathematical demand is less for these items. Taking into account the above, only Pi and PLi1 will be considered in this analysis; both Pi and PLi1 have the same mathematical demand, they vary just in the objective figurative context in which the item is embedded.

The mean of students’ performance followed particular trends in the uses of context. Figure 4 below shows how the mean performance was influenced positively on the first-order use of context when the alteration of context was made on PLi1 on both the PISA and the adapted marking scheme. Nevertheless, on the second-order use of context the trend was negative when an alteration of context was introduced; students performed lower on PiL1 on both the PISA
and the adapted marking scheme. However, for this research question, it is the magnitude of the change, and not its direction that is of interest.

Figure 4
Mean of students’ performance on the uses of context on Pi and PLi1 items by used of context

4.3.1 The First-Order Use of Context and Overall Performance

The overall performance on PISA items (Pi) compared to overall performance on items embedded in a more familiar context for students (PLi1) on the first-order use of context is shown in Figure 4. From Tables 12 and 13, similar differences were observed on Pi and PLi1. When these items were marked according to the adapted marking scheme the above trend was also observed, although the difference was a little larger.
4.3.2 The Second-Order Use of Context and Overall Performance

Tables 12 and 13 as well as Figure 4 suggest that the average performance on the second order use of context varies between Pi and PLi1. Students performed better in Pi than PLi1 when the PISA marking scheme was used. It should be noted that as stated in 4.2, it seems from Figure 3, that the second-order use of context has been strongly influenced by the large difference in performance between item Pi-B and its sibling PLi1-B. Possible reasons for the negative effect of change for PLi1-B will be also discussed with the qualitative data collected.

4.3.3 Effect of the Nature of the Use of Context on Overall Students’ Performance

The hypothesis related to Research Question 3 in sub-section 3.9 proposed that changing context would make more change in items’ families where there was second-order use of context than the first-order use of context (refer to 1.3.2 for definitions). Since this hypothesis predicts the magnitude of change, rather than whether one context of a pair makes an item easier or harder than another, the hypothesis will be examined by consideration of the absolute values of the differences in performance.

The above can be observed in Figure 5 and Figure 6 below. Figure 5 provides the students’ performance on the four items families grouped by type of use of context (as noted above). The performance on the PISA items was sometimes better and sometimes worse than performance on PLi versions. However, bigger changes can be seen in the second order use of context. It is therefore important to use absolute difference in scores. Figure 6 therefore shows the mean absolute difference in mean score by item family and uses of context correspondingly. Under the PISA marking scheme, the mean absolute difference of means for the two families requiring the second-order use of
context (families B and D) was 0.6 and 0.5. These two absolute differences are considerably larger than the mean absolute differences of means for the two families requiring the first-order use of context (families A and C) which are 0.0 and 0.2 respectively (see Figure 6).

Figure 5
Summary of mean of students’ performance by item family and uses of context on Pi and PLi1 items
Figure 6
Mean Absolute Difference in Mean Score by item and uses of context

When those items were marked under the adapted marking scheme the mean absolute differences in the second-order items were 0.8 and 0.2 respectively (see Figure 6); whereas on the items which involved the first-order use of context the absolute mean differences were 0.1 and 0.3.

With the experiment only using two families of items for each type of context use this analysis provides evidence to support the claim that the items involving second-order use of context seem to be more affected by change of context than items involving the first-order use of context. The Research
Question 3 therefore appears to be supported, but the data is not sufficiently extensive to justify statistical analysis.
4.4 Results of the Qualitative Phase

4.4.1 Introduction
Section 4.2 has indicated that students on average were no more successful statistically on items embedded in an altered context (PLi1) or in an altered context with fewer constraints (PLi2) compared to PISA items (Pi). However, it seemed that a change of context is more likely to influence students’ performance on the second-order use of context but not on the first-order use of context (see 4.3). Therefore Sections 4.4 to 4.7 will provide evidence relate to the above claim using data obtained from the students’ written records of the items’ solutions as well as from the students’ interviews.

4.4.2 Description of the Data Included in this Section
The data presented in Section 4.4 to 4.7 corresponds to the ten students who agreed to participate in the whole data collection process. Hence, the sample was reduced from the initial thirty to ten students. These students each completed one booklet of items and then were interviewed by the researcher of this study using methods that were fully described in sections 3.5 and 3.6. These students are referred to as Participant 1, Participant 2, … , and Participant 10 abbreviated as P1, P2, …, P10 respectively while the researcher is referred to as Rs.

Considering the above, the analysis in this chapter will be undertaken through the following data sources: (a) the ten student’s written answers to the paper and pencil tests with items having the same mathematical demand but altered in contexts and constraints (see Tables 7 and 8), and (b) the ten students’ interviews which are related to their paper and pencil test solutions. Table 14 informs about the data, participants, and individual students’ results for the
It should be noted that only Pi and PLi1 items will be considered for the analysis in this chapter for the reasons stated in section 4.4.

Table 14
Individual results in relation to booklets’ items and uses of context

<table>
<thead>
<tr>
<th>Student</th>
<th>Booklet</th>
<th>Pi-A</th>
<th>PLi1-A</th>
<th>Pi-C</th>
<th>PLi1-C</th>
<th>Pi-B</th>
<th>PLi1-B</th>
<th>Pi-D</th>
<th>PLi1-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>A</td>
<td>C</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>I</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P2</td>
<td>A</td>
<td>I</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>I</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P3</td>
<td>B</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>C</td>
<td>-</td>
<td>-</td>
<td>P</td>
<td>-</td>
</tr>
<tr>
<td>P4</td>
<td>C</td>
<td>-</td>
<td>C</td>
<td>C</td>
<td>-</td>
<td>I</td>
<td>-</td>
<td>-</td>
<td>I</td>
</tr>
<tr>
<td>P5</td>
<td>B</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>P</td>
<td>-</td>
<td>-</td>
<td>P</td>
<td>-</td>
</tr>
<tr>
<td>P6</td>
<td>A</td>
<td>C</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>M</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P7</td>
<td>C</td>
<td>-</td>
<td>C</td>
<td>C</td>
<td>-</td>
<td>C</td>
<td>-</td>
<td>M</td>
<td>-</td>
</tr>
<tr>
<td>P8</td>
<td>A</td>
<td>C</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>C</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P9</td>
<td>A</td>
<td>I</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>I</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P10</td>
<td>C</td>
<td>I</td>
<td>I</td>
<td>-</td>
<td>I</td>
<td>-</td>
<td>-</td>
<td>P</td>
<td>-</td>
</tr>
</tbody>
</table>

• Partial credit not available on Pi-A and Pi-B and its sibling items on the PISA marking scheme.
• C: Correct (Full Credit)
• I: Incorrect (No Credit)
• P: Partial (Partial Credit)
• M: Missing answer
• Booklet A: Pi-A, PLi1-B, PLi2-C, PLi2-D
• Booklet B: PLi2-A, PLi1-C, Pi-D, PLi2-B
• Booklet C: PLi1-A, Pi-B, PLi1-D, Pi-C

Through these data sources, the role of item-context on students’ performance is explored; that is to say, patterns in students’ responses to the paper and pencil test as well as to the interviews will be identified in relation to the uses of item-context and their consequences, if any, for students’ performance. All the interviews transcripts can be found in Appendix C but extracts of the interviews will be used in this chapter.

It should be noted that students’ data has been divided into two categories: on students’ dealing with the first-order use of context and on students’ dealing with the second-order use of context. Possible patterns will be identified by examining similarities and differences in the students’ written solution and interview. It is thus hypothesised that this analysis will support the claim that a change of context is more likely to influence students’ performance for the
second-order use than the first-order use. Now qualitative evidence is presented.

4.5 Students Dealing with the First-Order Use of Context

As Table 14 illustrates, the six students (P1, P3, P4, P6, P7, and P8) who dealt with first-order use of context items correctly. On the basis of the analysis of the students’ written answers it appears that students who solved the items correctly appear understood precisely the requirements of the questions in the items. For example, Participant 1’s written solution on Pi-A revealed how she used the item-context to understand the item as well as to check if the solution fitted the item’s requirements.

Figure 7
Participant 1’s solution method on Pi-A

<table>
<thead>
<tr>
<th>Place</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sydney</td>
<td>5:00 pm</td>
</tr>
<tr>
<td>Berlin</td>
<td>8:00 am</td>
</tr>
</tbody>
</table>

Question 1.1:
At 7:00 PM in Sydney, what time is in Berlin?
10:00 AM
7:00 pm - 10:00 am
2 1/2 hours
10:00 am + 2 1/2 hours = 10:00 am

Question 1.2:
Mark and Hans are not able to chat between 9:00 AM and 4:30 PM their local time, as they have to go to school. Also, from 11:00 PM till 7:00 AM their local time they won’t be able to chat because they will be sleeping. When would be a good time for Mark and Hans to chat?

Show your working here and explain your reasoning.

This time would be suitable because neither of the people would be at school or asleep. I found this answer by choosing a suitable time for one of the boys then seeing if it would be suitable for the other.
Participant 1’s comments on Pi-A item revealed how the item-context assisted her in finding a point of connection between the requirement of the item and the information provided by the item-context. In referring to Pi-A, she commented:

P1: I started a process of elimination, a sort of trial and error but it happened that the first one I used worked. So I just found a suitable time in Sydney first by minusing 2 hours from question 1.1 and then I used the time difference from question 1 to figure the time in Berlin which also fit a reasonable hour, and yeah.
Rs: At this stage, was there anything that stopped your thinking process, anything that was difficult?
P1: The only difficulty was it had to be a specific time in both time zones.
Rs: Anything else?
P1: Not really.

It seems that a first-order use of context did not hinder her making sense of the item.

Participant 4 answered the PLi1-A item correctly which required the first-order use of context in her booklet. Participant 4’s written solution to PLi1-A (2010 Commonwealth Games, see Figure 8 below) had a different objective figurative context than its parallel Pi-A item (Internet Chat Relay). Her comments also showed that when the first-order use of context is needed to solve the item an alteration of context was not a complication to extract the information necessary for solving and judging the answer. In commenting about this item, Participant 4 said that she could not figure out what a good time would be, whereas from the item-context she felt that she could relate more easily to the requirements of the item:

Rs: What about in 1.2?
P4: Well, I looked at the time they can’t broadcast, between 9 and 4:30, and between 11 and 7, and I thought of the time that in between they can broadcast so I just picked a random number and worked out the other time in the other place, yeah, and if the two worked in the same... I got the answer.
Rs: Did you have a concern about what a good time would be?
P4: Not really, I didn’t … no, like exactly what time is a good time, so that it would be tricky, so I guess...
**Rs:** Did you go back to check the solution?

**P4:** No, I didn't.

**Rs:** Did the fact that this question is about the 2010 Commonwealth Games produce any impact in order for you to check the solution?

**P4:** No. no.

**Rs:** Did you see it as a probable real situation?

**P4:** Yes, sort of.

**Rs:** Was there anything external that impacted your solution process?

**P4:** No

---

**Figure 8**

*Participant 4’s written answer on PLi1-C*

Mark and Hans are sport journalists who are going to broadcast the 2010 Commonwealth Games. Hans is reporting in Delhi, India and Mark is in Melbourne. They have to broadcast live to their Melbourne audience daily. To find a suitable time to broadcast live, producers looked at a chart of world times and found the following:

- Greenwich: 12 Midnight
- Delhi: 5:30 AM
- Melbourne: 10:00 AM

**Question 1.1:**
At 7:00 PM in Melbourne, what time is in Delhi?

4 30min difference
2 30pm Delhi

**Question 1.2:**
Due to the television scheduling, Mark and Hans are not able to broadcast live between 9:00 AM and 4:30 PM their local time, and also from 11:00 PM until 7:00 AM their local time. When would be a good time for Mark and Hans to do it? Write the local times in the table.

<table>
<thead>
<tr>
<th>Place</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delhi</td>
<td>5:30 pm</td>
</tr>
<tr>
<td>Melbourne</td>
<td>9:30 pm</td>
</tr>
</tbody>
</table>

**Show your working here and explain your reasoning.**

I chose a time in between 4:30 and 11pm, which was 5pm I added 4.30 hours to it to receive the Melbourne time. Both were in the limit so they are acceptable.

**Additionally, what made this question easy or hard for you to answer?**

The different times when they can broadcast can vary and the definition of “a good time” is not clear in its meaning.
It is informative that she did not report the 2010 Commonwealth Games context as an extra motivation factor to solve this item. It should be noted that the Commonwealth Games were advertised amply on national television at the time of the data collection. Although she saw the item-context as a possible real scenario, no evidence of specific motivation can be seen in the solution or in the interview’s excerpt. It is possible that the 2010 Commonwealth Games as well as the protagonists (Mark and Hans) of this item are less appealing to girls, hence the fact that no variation in performance was found between Pi-A and PLi-A could be explained by the fact that the data was obtained from a girls’ school (see Table 12).

Participant 7 performed correctly on Pi-C (Support for the President). From her written solution to this item (see Figure 9) she showed an understanding of the important characteristics to evaluate the samples in this item: the survey should be taken from a large sample, it should be a random sample, and of course only respondents who are eligible to vote should be considered.
Participant 7’s comments towards the eligibility of the participants in the poll reinforce the interaction between the item’s requirement and her involvement with the item-context to solve and judge her answer:

**Rs:** What can you tell me about this item?

**P7:** I picked newspaper 3 because it seems that with these... people phoning in you don’t know exactly how phoning in is, but with these ones [referring to alternative A and C] they are people who can vote, they will all be totally random, whereas these people [referring to alternative D] could be in the same group, so obviously they’re gonna show their support for what they believe in.
Participant 5 solved PLi1-C (Support for a New Shopping Centre) which is the parallel item to Pi-C. She got partial credit on this item because she gave just one valid reason (eligibility) to support her answer (see Figure 10). It should be noted that Participant 5 in her written response was not explicit to communicate her second argument (note the underlining in her response). Perhaps, participant 5 meant randomness but selection does not mean randomness always. For this reason, the Participant 5’s comments on this item are necessary to clarify the underlining on her written answer.

Figure 10
Participant 5 solution on PLi1-C

**Question 3:**
In Zedland opinion polls were conducted to find out the level of support for the construction of either a new Shopping Centre or a Swimming Pool. Four polls were conducted separately. The percentages of people preferring the shopping centre are shown below:

- **Poll 1:** 36.5\% (poll with a sample of 1000 **volunteer Zedland residents**).
- **Poll 2:** 41.0\% (poll with a sample of 500 **women volunteers**).
- **Poll 3:** 39.0\% (poll with a sample of 1000 **randomly selected Zedland residents**).
- **Poll 4:** 44.5\% (poll with a sample of 1000 **randomly selected Zedland women**).

Which poll’s result is likely to be the most reliable for predicting the level of preference for the new Shopping Centre? Give two reasons to support your answer.

*Show your working here and explain your reasoning.*

| Poll 3 is going to be the most reliable as it selected the residents (not based on gender) |

When commenting on her answer she referred to the two valid reasons (randomness and eligibility) needed to support a correct answer:

**P5:** I chose my answer by reading through the polls and the words that they used, volunteers or randomly selected, and then women or residents; so obviously more women could want the shopping centre or if you are a volunteer, probably if it is someone out in the street doing something you are more active or you want the shopping centre whereas if you are randomly
selected you may not used, things like that, so the randomly selected residents would be the most accurate.

Participant 5’s comments show how she recognised contextual aspects of the item on the eligibility of the people to be polled taking into account aspects such as people who would support the construction of a new shopping centre (volunteers, women, residents); this information seemed to help her make sense of the information provided in the item’s text rather than helping her validate the answer.

From Participant 7 and Participant 5’s written solutions and comments to items in family-C it can be noted that when this change of context was introduced it did not play a role when students answered the items.

At this point, it seems that students who answered items requiring the first-order use of context integrated the item-context with metacognitive processes such as checking that their results fitted the contextual demands of the items; the item-context assisted them in making sense of the items, and in finding points of connection between what they were required to do and the context in which the item was embedded. There was no evidence of a motivational effect of familiar context for these students.

The role of familiarity of task context and its subsequent impact on performance was not observable. This illustrates the possibility that performance on items linked to this use of context is less likely to vary when changing the objective figurative context.
4.6 Students’ Dealing with the Second-Order Use of Context

As Table 14 illustrates, ten students dealt with an item which required the second-order use of context. Evidence suggests a dual role of context in items that emphasise a second-order use of context; for some items the item-context helped but for others it hindered the solution process.

Participant 4, who answered incorrectly Pi-B (Rock Concert item, see Figure 11) commented on the difficulty of extracting relevant information from this context although she understood the mathematical procedure involved to solve the item:

**Figure 11**
P4 solution on Pi-B

**Question 2:**
For a rock concert a rectangular field of size 100 m by 50 m was reserved for the audience. The concert was completely sold out and the field was full with all the fans standing. Which one of the following is likely to be the best estimate of the total number of people attending the concert?

A About 2 000
B About 5 000
C About 20 000
D About 50 000
E About 100 000

*Show your working here and explain your reasoning.*

P4: Yes, if you think about what a rock concert looks like you have a better chance of solving the problem correctly; it is more like real life I guess.
Rs: So, would it be fair to ask this item if you haven’t seen or experienced one before?
P4: Yes and no. If it is yes, it is a common sense that there would be a lot of people there very squished and no because if you haven’t seen probably you wouldn’t know how to start it because you have to think about it.
**Rs:** Lastly, was the Mathematics to solve this item difficult?
**P4:** No, but it was important as the assumptions to solve it.

From Participant 4’s comments it can be inferred that she considered what a rock concert could look like but she did not take into account contextual aspects such as a realistic number of people per square metre to integrate it into the solution. She thought that 1 m$^2$ per person would be reasonable. That is to say, the interaction between the item and context did not take place because she did not consider this item as a real life situation (although she mentioned it); hence the second-order use of context was not activated.

Similar to Participant 4, Participant 10 also answered Pi-B incorrectly (see Figure 12).

**Figure 12**

*Participant 10 solution on Pi-B*

**Question 2:**
For a rock concert a rectangular field of size 100 m by 50 m was reserved for the audience. The concert was completely sold out and the field was full with all the fans standing. Which one of the following is likely to be the best estimate of the total number of people attending the concert?

A About 2,000  
B About 5,000  
C About 20,000  
D About 50,000  
E About 100,000

*Show your working here and explain your reasoning.*

\[100 \times 50 = 5000\]

I believe that if all the seats reserved were full and people where standing, the estimated people attending the concert would be about 5,000.

For Participant 10, the difficulty of extracting relevant information from the item-context (number of people that can be accommodated per square metre) is crucial to give a fair estimate:
Rs: What can you tell me about this rock concert item, please?
P10: With this one I timesed this together [the length and width of the field] because I needed to find the total number of people attending the concert
Rs: Were there any assumptions you made when you solved this item?
P10: well, I said this is the area for the amount of people you can sit in there and also some standing as well.

It is interesting to note she considered fans were seated, despite the information given in the item’s text.
From the analysis of both written solutions, students clearly understood the mathematical model required, but finally choose a conservative estimate when assuming that only one person can be accommodated per square metre.

These participants seem to be confused by the item-context; they could not detect that the item was referring to a real situation and expecting that students should use contextual cues (such as the space among people in a rock concert context) to solve the item. Hence, the second-order use of context required for this item did not take place properly. These students could not link contextual cues with the item as they seemed to be confused by the kind of question they were asked to solve. Thus, although the item was asking for estimation, students did not consider it as a real context rather students who performed less well considered it as a disconnected context from real life as can be noted in P10’s comments:

Rs: Why do you think you couldn’t relate it before?
P10: I don’t know. I just didn’t think logically. I didn’t think to put it in a real way. I just thought of working mathematically.
Rs: What do you mean by working mathematically?
P10: Like working doing lengths times widths. I thought I had to do a formula to work it out. I didn’t think in a real way, so I didn’t relate them

This confusion seemed to block the students’ access to contextual cues provided by the item; and as a result, the item-context rather than assisting them was an additional obstacle to deal with.
Participant 8 answered PLi1-B (Car Park) correctly, the sibling item for Pi-B (Rock Concert). As can be seen in Figure 13 below, she made an appropriate use of contextual cues provided by the item-context; that is to say, item-context assisted performance.

Participant 8 realised that this item were referring to a real-life situation. The item-context was the best source to retrieve information necessary to make assumptions in order to solve it:

**P8:** Yes, it (the context) did help me but also it kind of confused me because I didn’t know quite if the question was asking like how many cars all pack in together or if it was a car park situation when there are spaces between cars.

**Rs:** For this item, did you consider that you need more external factors than mathematics to solve it?

**P8:** I thought I needed a bit of logic because I know some people could have taken it more literally how many cars could be packed in, but then I thought you know that it doesn’t really work because you can’t get them out.

**Figure 13**

*Participant 8 solution on PLi1-B*
The second-order use of context assisted in the process of interpreting and relating the context to its mathematical elements. In this particular solution, it was considered as a real car park situation; therefore, in the solution process space allowance for exit roads, space between cars, and car slots needed to be considered.

Surprisingly when the researcher asked Participants 1, 2, 6, and 9 (the participants who answered PLi1-B item incorrectly) to consider contextual aspects of the item in their solution by drawing their mental picture of how a car park would look, all of them agreed that they could solve the item correctly.
after the drawing. They approached the item as a real situation and therefore made use of contextual information such as car slots (rather than estimation of the dimensions for each car), exit roads, and drive ways. It can be noticed in the students’ drawings below:

Figure 14
Participants drawings on PLi1-B

![Participant 1](image1)
![Participant 2](image2)

![Participant 6](image3)
![Participant 9](image4)

When Participants 1, 2, 6, and 9 were asked about the connection of their mental picture with their solution they clearly pointed out that their mental picture helped them to understand contextual cues (car slots, exit roads) necessary to solve the item; they realised that many factors are involved and they have to make assumptions about them.

When Participant 1 was asked about the role her drawing played in solving this item again, she commented:

**P2:** Ummm... I think it helped me. I realised there are many things. I don't know the lengths and measures involved, so it overwhelmed me, and I didn't know what to do with the question.
As Participant 2, Participant 1 also pointed out it was difficult to link the item-context with the solution, she said:

Rs: So why couldn’t you connect your picture with the problem?
P1: Ummm, when I first did it, the problem, I was... I couldn’t think of things that I see now, but when I did it, when it crosses my mind again I can pick up these little things, because I forgot about it and now I can go back to them.
Rs: Was the Mathematics to solve this problem difficult?
P1: There was just the area and the times and dividing on that and a bit of logic in it I guess.
Rs: So what was the difficulty of this problem?
P1: It was to connect these little things [the picture of the car park] with the question.

For Pi-B and PLi1-B, the context of the rock concert and the car park impaired performance in the following way. Students who answered these items incorrectly exhibited difficulties in extracting relevant information for their solution from the item-context. This difficulty was most likely because participants perceived the items in two aspects: Mathematics and Reality. They certainly felt ambivalent concerning the legitimacy of the way they were supposed to retrieve information in form of assumptions from the context rather than the written problem text: on one hand, they were unsure if they should take a contextually oriented approach while, on the other hand, they seemed to prefer a more straightforward calculation approach. The above has already been highlighted by Busse (2005, 2003). It should be noted that these participants understood the mathematical model which applied to the items but the greater contextual demands of these items clearly made them more difficult for participants than PLi2 items (refer to Table 11 and Figure 4).

4.7 Students Connection with the Item-Context

In this study was that excessive connection with the item context may allow some students to be reality bound impacting on performance (see Chapter 2). In this manner, certain contexts encouraged some students to think and argue in contextual terms rather than mathematically. This is observable on Pi-D and PLi1-D items set in contexts of robberies and train ticket prices.
Recalling from section 3.3.2.4, these items required students to understand and decode a graphical representation in a critical way. They had to make a judgment and find appropriate argumentation based on interpretation of data. Correct responses require students to think and argue proportionally in Pi-D and PLi1-D.

A familiar context close to students such as money (PLi1-D) was interpreted and judged as personal rather than from a mathematical point of view. This was evident when some students were susceptible to external aspects found in the item-context which did not form part of the task solution to communicate their arguments.

Participant 4 and Participant 10’s solutions to PLi1-D are examples of the above. These participants were reality bound; they considered this item real using Melbourne metropolitan ticket prices to judge their answers which produced a large scenario for a narrow task. Indeed, considering Melbourne metropolitan ticket prices different kinds of tickets and prices can be found depending on the travel zone (zone 1, zone 2, or zone 1 and 2) and the commuter status (student, pensioner, or adult). Evidence from Participant 4 is now presented to exemplify the above, discussing the solution in Figure 15.

P4: Well when I looked at it, it is monthly train ticket prices and it was like between $300 and $400, umm... I just thought that a train ticket can’t cost that much money and the title wasn’t very correct. It was confusing at the start.
Rs: So you are telling me that your answer was based on your knowledge about the ticket prices?
P4: Yes.
Rs: But did you consider other kinds of monthly ticket prices? Melbourne to Ballarat for instance...
P4: No, I don’t know that price I just knew the Metropolitan one.
Rs: So, based on that information you solved the item?
P4: Yes
When she was asked to focus on the graph in order to answer this item, she could not separate her personal knowledge of this context from her answer. She applied what she knew about this context. She argued that for money-context a context is needed within which to think. As Participant 4 commented:

Rs: Well, if you based your answer now just by looking at the graph, no title, would you give the same answer?
P4: I wouldn't know what it is about; I would be really confused because there is no information telling me what the graph is about.
Rs: Let’s imagine now that instead of saying “A TV reporter showed this graph and said” I would say “An overseas TV reporter showed this graph and said” and the same information is provided, would your answer change?
P4: Ummm… I don’t think… maybe it’s not that correct because he hasn’t actually… if it is in Australia he hasn’t lived in Australia to see exactly how everything is… but it may also be correct because he needs to research well. At the end you need a context for prices.
Rs: Have you considered anything else rather than prices?
P4: No, just prices.
Interestingly, when another item-context (robberies) was used to test the same mathematical core, Participants 3 and 5 focussed their arguments on absolute difference rather than using proportional reasoning or graph trends. As can be seen in Figure 16, these participants only focussed on the actual difference of robberies from 1998 to 1999.

Figure 16
Participants 3 and 5 solutions on PLi1-D respectively

Although their answer would be classified as partial on the PISA marking scheme (see Appendix B), their interviews revealed that these participants communicate their answers in a taken-as-shared way which may have an impact on their performance on items; Participant 3 commented on this item:

**P3:** Well, according to this the reporter says that was a huge increase, but here the increase is just only 10 robberies, so it depends on the people who have been robbed, but we are looking at 500 and what is it? ... anyway, 10 is not a huge difference, so it probably is a little exaggeration.

**Rs:** When you say 10, why did you conclude it is not a big increase?

**P3:** Why? Well because the number of robberies is 500, and it is pretty high and 10 is a small part of that.

**Rs:** So, when you give that answer, are you assuming that you don’t need to give further details in order for me to understand it?

**P3:** No, because 10 out of 500 is not much.
Participant 5 also commented in a similar way on her solution to Pi-D:

**P5:** To figure it out I just looked at the scale that was used and then I realised that the difference is only 8, so in one year there were 10 more robberies, but in considering that numbers are so high, they are in the 500s, so it really doesn't make much difference, that was what I thought.

**Rs:** I see. But you have written that 10 robberies don't represent a huge increase. So you are assuming that people who may read your answer understand your thinking straight away? I mean that 10 is not much?

**P5:** Yeah, it is not much.

**Rs:** What would happen if you present this argument to your family or classmates? Would you need to explain it to them?

**P5:** I wouldn't need to explain this, not much; you would know that 10 out of 500 is not much.

From these excerpts clearly participants communicated their arguments with a mathematical component implicit in it, namely proportional reasoning. Nevertheless, the PISA marking scheme (see Appendix B) seems to be inappropriate for this item. This marking scheme did not consider a full credit answer with implicit proportional reasoning which may underestimate the students’ mathematical literacy.

For Pi-D and PLi1-D, item-context also served to impair performance in the following way. On one hand, participants felt reality bound when an item was embedded in a close personal context. As a result, on the items which required the second-order use of context the familiarity of task context may have an impact on performance. This illustrates the possibility that performance on items linked to this use of context is more likely to vary when changing the objective figurative context.

On the other hand, Sáenz (2009) pointed out students who answered these items incorrectly exhibited difficulties in extracting visual information from the item-context. This difficulty was not observable in these items. From the evidence presented here, students did consider visual aspects of the item-context to communicate their arguments but students are more likely to communicate it in a taken-as-shared way which may impact their performance.
CHAPTER 5: CONCLUSION

Recalling from Chapter 2, results are mixed on students’ performance with reference to the effects of contextualised tasks; the contexts can have a positive, a negative or a neutral impact depending on what the contextual item is being compared with. One previous study (Sáenz, 2009) involving PISA Mathematics showed that participants of this study were influenced by contextual and superficial features of the items.

This present study was organised around three research questions, with a fourth research question relating to a qualitative explanation of the findings. Data derived from written solutions to items and students’ interviews were analysed to address the four research questions. In this chapter the major findings, limitations, and recommendations of this study will be considered.

5.1 Major Findings

Research Question 1 aimed to investigate the role of an alteration of context on students’ performance. Contrary to the original hypothesis, in this study students did not perform better on PISA-like items (PLi1) with a more familiar context than on PISA items (Pi). Qualitative evidence from Chapter 5 revealed that in a more familiar context (money context), students tended to bring personal information into arguments rather than using a mathematical argument.

Research Question 2 also aimed to investigate the role of an alteration of context on students’ performance but on items with fewer contextual constraints (PLi2). Contrary to the original hypothesis, students did not perform better on these items (PLi2). As discussed in 4.2.2 the decision to alter context in PLi2 items
produced a reduction of contextual constraints in these items which has proved to be a methodological error. However, in this study, the lack of support for the hypothesis in Research Question 1 and 2 does not mean that alteration of item-context had no effect on performance.

Research Question 3 investigated the relation between changing context and changes in performance in items’ families requiring the first and second-order use of context. The data analysed in Chapter 4 showed that the students’ performance on the second-order use of context seemed to be more influenced by the alteration of context than with first-order use of context. It should be noted that the analysis for Research Question 3 provides evidence to support the claim that the items involving second-order use of context seem to be more affected by change of context than items involving the first-order use of context. The Research Question 3 therefore appears to be supported, but the data is not sufficiently extensive to justify statistical analysis.

Research Question 4 sought qualitative data to explain findings related to Research Question 3. On one hand, evidence from Chapter 4 revealed that in terms of students dealing with the first order use of context, successful students appeared to understand precisely the requirements of the questions in the items whereas students who performed less well had difficulty to extract information from the item-context. On the other hand, when students dealt with items which require the second order use of context on one hand, the differential nature on performance seemed to be related more to whether students perceived each task as a real task or as a standard application task that is related to the item-context. That is to say, from the text data it seemed that when some students related the item to a real situation contextual cues provided by the context are used but when some students categorised the task as a standard
application task they seemed confused by item-context and by the lack of data needed to answer the item. Hence they did not realise the potential of the context as a resource to solve the item; in fact, some students were able to provide the mathematical model related to the item but they provided unrealistic answers to the item.

Finally, from the students’ interviews it can be seen that certain contexts encouraged some students to think and argue in contextual terms rather than mathematically which leads to point out that item-context may guide the selection of arguments to communicate students’ answers.

The results of this study confirmed Murphy’s (1990) findings that the context of an assessment task may disadvantage students because they may value the context that a task is presented in, having difficulty subsequently to abstract the right issues from the context (see 4.4). The findings reported here also provided evidence to support Sáenz’ (2009) study where he found that visual aspects of graphs may distract students from proportional reasoning or arguing graph trends (see Chapter 2). However, here this has been linked to the students’ connection to the item-context and how students used the context to argue.

### 5.2 Limitations of this Study

The results of this study tried to give insights into the four research questions stated in 1.2. Nevertheless, this study within a minor thesis is only a starting point. Due to the scope of this minor thesis, three sources formed part of the limitations in this study, namely: sample restrictions, number of items, and types of task investigated.
First of all, the number of participants may not be an ample sample size to observe moderate variation in performance across items. Also, a sufficient sample is needed because in this study the students could not do parallel items which impacted for direct student comparison. This is because it is expected that many or most students would either recognise the underlaying mathematical structure, or would not recognise it in any instance. As a result, a rotated design of three booklets was used and, hence comparisons at student level were not possible in this study. Overcoming this issue requires a larger sample size.

Secondly, when only one school and two classes are selected for the main study, the results in performance might be a result of the type of instruction provided to students; therefore it might not allow the observation of differences between different types of instruction, and may be quite atypical. The fact that the data was collected in a girls’ school also may affect the reaction to the contexts.

Thirdly, only PISA items and PISA-like items with altered contexts, and items with fewer constraints were used in this study; therefore the impact of item-context will be restricted to those types of changes. Merely assessing student performance based on PISA and PISA-Like items was restrictive not only by the type of items used but also in the number of items used. On one hand, this study only looked at the results of two versions of items. It would be interesting to investigate how students might perform on other sorts of mathematical tasks (for instance in a collaborative work project) which required the uses of context. On the other hand, the number of items used (twelve in total) was relatively small in terms of the ranges of contexts examined and hence, the type of change was restricted. Time and methodological constraints did not allow for this additional research component but it will be required for further research.
Nevertheless, it should be noted that the finding of no difference in students’ performance could be attributable to the fact that the more familiar contexts chosen for students were not more familiar for students after all. In a further study, it could be possible to find out if in reality certain contexts are more familiar to students. For example, students could be asked to vote on the familiarity, appeal, and importance of each of the contexts before solving.

In addition, there may be different demands on the mathematical core from items which were supposed to be similar. The researcher acknowledges that more comprehensive and finer methods are required to analyse the mathematical core between items.

Having established this, further research is encouraged to test the generality of the conclusions reached considering the factors above.

5.3 Recommendations

The outcomes of this study have recommendations in two areas. First, there is an issue of action in the PISA marking scheme which should be considered as a result of the findings of this study. Second, suggestions for further research are presented in order to reach a more refined understanding of the complex interaction between item-context and performance.

5.4.1 Recommendations for PISA Mathematics

Evidence from section 4.4 revealed that some participants communicated their arguments with a mathematical component implicit in it, namely proportional reasoning. Nevertheless, the PISA marking scheme (see Appendix B) seems to be inappropriate for this item. This marking scheme did not consider as full credit when an answer with implicit proportional reasoning which may
underestimate the students’ mathematical literacy. From the evidence presented in Chapter 4, some students tended to communicate their arguments in a more likely taken-as-shared way rather than mathematically. This personal way of communicating arguments may impact on their performance. Thus, it is recommended for PISA Mathematics to take a broader marking perspective when coding students’ arguments for a broader interpretation of these arguments considering taken-as-shared answers in its marking scheme. For example, as discussed in section 4.7 the PISA marking scheme did not consider an answer with implicit proportional reasoning which may underestimate the students’ mathematical literacy. In the interview, both Participant 3 and 5 (their answers were coded as partial according to the PISA marking scheme) were able to explain their implicit proportional reasoning explicitly.

5.4.2 Further Research

The general thrust of the results would suggest that changing context can make more change on performance on items that involve second-order use of context than first-order use. Recall from sub-section 1.3.2 where it was proposed that changing context would make more change in items’ families where there was second-order use of context than the first-order use of context. Since this hypothesis predicts the magnitude of change, rather than whether one context makes an item easier or harder than another, the hypothesis was examined by consideration of the absolute values of the differences in performance.

Results of this study did not show a clear picture of the relationship between alteration of context and students’ performance. This suggests a large scale quantitative study might be needed to investigate differential results on the uses of context taking into account a clearer conceptualisation of contextual constraints, different types and numbers of items, and students at different ages.
Taking the above into account, this study leads to the elaboration of a more detailed hypothesis for this research problem at the level of alteration of context only. As a result, the following hypothesis is projected for further research: The students’ performance on items which require the second-order use of context are more influenced (either positively or negatively) by a change in context on items which have the same mathematical demand compared to the first order use of context.
REFERENCES


APPENDICES

APPENDIX A: THE TWELVE ITEMS USED IN THE STUDY .................................................. II
APPENDIX B: THE MARKING SCHEMES ........................................................................ V
APPENDIX C: TRANSCRIPTS OF THE INTERVIEWS ..................................................... IX
Appendix A: The Twelve Items used in the Study

PISA Items

• Pi-A
Mark (from Sydney, Australia) and Hans (from Berlin, Germany) often communicate with each other using “chat” on the Internet. They have to log on to the Internet at the same time to be able to chat. To find a suitable time to chat, Mark looked up a chart of world times and found the following:

![Chart of World Times]

**Question 1.1:**
At 7:00 PM in Sydney, what time is in Berlin?

**Question 1.2:**
Mark and Hans are not able to chat between 9:00 AM and 4:30 PM their local time, as they have to go to school. Also, from 11:00 PM till 7:00 AM their local time they won’t be able to chat because they will be sleeping. When would be a good time for Mark and Hans to chat?

<table>
<thead>
<tr>
<th>Place</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sydney</td>
<td></td>
</tr>
<tr>
<td>Berlin</td>
<td></td>
</tr>
</tbody>
</table>

• Pi-B
For a rock concert a rectangular field of size 100 m by 50 m was reserved for the audience. The concert was completely sold out and the field was full with all the fans standing. Which one of the following is likely to be the best estimate of the total number of people attending the concert?

A About 2 000  
B About 5 000  
C About 20 000  
D About 50 000  
E About 100 000

• Pi-C
In Zedland, opinion polls were conducted to find out the level of support for the President in the forthcoming election. Four newspaper publishers did separate nationwide polls. The results for the four polls are shown below:

- Newspaper 1: 36.5% (poll with a sample of 500 randomly selected citizens with voting rights)
- Newspaper 2: 41.0% (poll with 500 readers phoning in)
- Newspaper 3: 39.0% (poll with a sample of 1000 randomly selected citizens with voting rights)
- Newspaper 4: 44.5% (poll with 1000 readers phoning in).

Which newspaper’s result is likely to be the most reliable for predicting the level of support for the President? Give two reasons to support your answer.

• Pi-D
A TV reporter showed this graph and said:
“The graph shows that there was a huge increase in the number of robberies from 1998 to 1999.”

Do you consider the reporter’s statement to be a reasonable interpretation of the graph? Give an explanation to support your answer.
PISA-Like items 1

• **PLi1-A**
  Mark and Hans are sport journalists who are going to broadcast the 2010 Commonwealth Games. Hans is reporting in Delhi, India and Mark is in Melbourne. They have to broadcast live to their Melbourne audience daily. To find a suitable time to broadcast live, producers looked at a chart of world times and found the following:

![Time Zones]

**Question 1.1:**
At 7:00 PM in Melbourne, what time is in Delhi?

**Question 1.2:**
Due to the television scheduling, Mark and Hans are not able to broadcast live between 9:00 AM and 4:30 PM their local time, and also from 11:00 PM until 7:00 AM their local time. When would be a good time for Mark and Hans to do it? Write the local times in the table.

<table>
<thead>
<tr>
<th>Place</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delhi</td>
<td></td>
</tr>
<tr>
<td>Melbourne</td>
<td></td>
</tr>
</tbody>
</table>

• **PLi1-B**
  For an exhibition, a rectangular field of 100 m by 50 m was reserved for a car park. The cark park was completely full. Which one of the following is likely to be the best estimate of the total number of cars in the car park?
  A About 200
  B About 500
  C About 1 000
  D About 2 000
  E About 3 000

• **PLi1-C**
  In Zedland opinion polls were conducted to find out the level of support for the construction of either a new Shopping Centre or a Swimming Pool. Four polls were conducted separately. The percentages of people preferring the shopping centre are shown below:

  Poll 1: 36.5% (poll with a sample of 1000 volunteer Zedland residents).
  Poll 2: 41.0% (poll with a sample of 500 Zedland women volunteers).
  Poll 3: 39.0% (poll with a sample of 1000 randomly selected Zedland residents).
  Poll 4: 44.5% (poll with a sample of 1000 randomly selected Zedland women).

  Which poll’s result is likely to be the most reliable for predicting the level of preference for the new Shopping Centre? Give two reasons to support your answer.

• **PLi1-D**
  A TV reporter showed this graph and said:

  “The graph shows that there was a huge increase in the monthly train ticket prices from 1998 to 1999.”

  ![Monthly Train Ticket prices]

  Do you consider the reporter’s statement to be a reasonable interpretation of the graph? Give an explanation to support your answer.
PISA-Like items 2

• PLi2-A
Mark (from Sydney, Australia) and Hans (from Berlin, Germany) often communicate with each other using “chat” on the Internet. They have to log on to the Internet at the same time to be able to chat. To find a suitable time to chat, Mark looked up a chart of world times and found the following:

Question 1.1:
At 7:00 PM in Sydney, what time is in Berlin?

Question 1.2:
When would be a good time for Mark and Hans to chat? Write the local times in the table.

<table>
<thead>
<tr>
<th>Place</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sydney</td>
<td></td>
</tr>
<tr>
<td>Berlin</td>
<td></td>
</tr>
</tbody>
</table>

• PLi2-B
A 100cm by 50cm table is to be tiled by 5cm by 2cm tiles. How many tiles are needed to cover this table?
A About 150
B About 200
C About 350
D About 500
E About 650

• PLi2-C
Here are the results of four opinion polls. Which poll listed below is likely to be the most reliable? Give two reasons to support your answer.

Poll 1: 36.5% (poll with a sample of 500 randomly selected participants).
Poll 2: 41.0% (poll with a sample of 500 volunteer participants).
Poll 3: 39.0% (poll with a sample of 1000 randomly selected participants).
Poll 4: 44.5% (poll with a sample of 1000 volunteer selected participants).

• PLi2-D
A TV reporter showed this graph and said:
“The graph shows that there was a huge increase in the average of the sea surface temperature from 1998 to 1999.”

Do you consider the reporter’s statement to be a reasonable interpretation of the graph? Give an explanation to support your answer.
## Appendix B: The Marking Schemes

### Marking schemes for Family A-items

<table>
<thead>
<tr>
<th>Items</th>
<th>PISA Guidelines</th>
<th>Adapted marking scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pi-A</strong></td>
<td><strong>Credit (2 points)</strong></td>
<td><strong>Partial Credit (1 point)</strong></td>
</tr>
<tr>
<td>Participants gave an answer with any time or interval of time satisfying the 9 hour time difference. These could be taken from one of the following intervals: Sydney: 4:30 pm – 6:00 pm; Berlin: 7:30 am – 9:00 am, or Sydney: 7:00 am – 8:00 am; Berlin: 10:00 pm – 11:00 pm</td>
<td>Participants gave an incorrect answer considering just one correct local time</td>
<td>Participants not responding the item or having an incorrect answer</td>
</tr>
<tr>
<td><strong>PLi1-A</strong></td>
<td>Participants gave an answer with any time or interval of time satisfying the 4.5 hours difference. These could be taken from the following intervals: Melbourne: 7:30 pm – 10:00 pm; Delhi: 15:00 pm – 5:30 pm.</td>
<td>Participants gave an incorrect answer considering just one correct local time</td>
</tr>
<tr>
<td><strong>PLi2-A</strong></td>
<td>Participants have to appreciate the fact that it is 9 hours later in Sydney. Then they have to apply this difference to the new situation.</td>
<td>Non applicable</td>
</tr>
<tr>
<td>Items</td>
<td>PISA Guidelines</td>
<td>Adapted marking scheme</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------</td>
<td>------------------------</td>
</tr>
<tr>
<td></td>
<td>Full Credit (2 points)</td>
<td>Partial Credit (1 point)</td>
</tr>
<tr>
<td><strong>Pi-B</strong></td>
<td>Successful participants chose alternative C</td>
<td>Non applicable</td>
</tr>
<tr>
<td><strong>PLi1-B</strong></td>
<td>Successful participants chose alternative B</td>
<td>Non applicable</td>
</tr>
<tr>
<td><strong>PLi2-B</strong></td>
<td>Successful participants chose alternative D.</td>
<td>Non applicable</td>
</tr>
</tbody>
</table>
## Marking schemes for Family C-items

<table>
<thead>
<tr>
<th>Items</th>
<th>PISA Guidelines</th>
<th>Adapted marking scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full Credit (2 points)</td>
<td>Partial Credit (1 point)</td>
</tr>
<tr>
<td>Pi-C</td>
<td>Participants chose alternative C, that is, they gave 2 valid reasons to support their choice</td>
<td>Participants chose alternative C, but they gave 1 valid reason to support their choice</td>
</tr>
<tr>
<td>PLi1-C</td>
<td>Successful participants chose alternative C, that is, they gave 2 valid reasons to support their choice</td>
<td>Participants chose alternative C, but they gave 1 valid reason to support their choice</td>
</tr>
<tr>
<td>PLi2-C</td>
<td>Successful participants chose alternative C that is, they gave 2 valid reasons to support their choice</td>
<td>Participants chose alternative C, but they gave 1 valid reason to support their choice</td>
</tr>
</tbody>
</table>
## Marking schemes for Family D-items

<table>
<thead>
<tr>
<th>Items</th>
<th>PISA Guidelines</th>
<th>Adapted marking scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full Credit (2 points)</td>
<td>Partial Credit (1 point)</td>
</tr>
<tr>
<td>Pi-D</td>
<td>Participants gave an answer supported by a mathematical argument (proportional reasoning)</td>
<td>Participants reasoning only focuses on an increase given by an exact number of robberies in absolute terms, but not in relative terms.</td>
</tr>
<tr>
<td>PLi1-D</td>
<td>Participants gave an answer supported by a mathematical argument (proportional reasoning)</td>
<td>Participants reasoning only focuses on an increase given by an exact amount of money in absolute terms, but not in relative terms.</td>
</tr>
<tr>
<td>PLi2-D</td>
<td>Participants gave an answer which focuses on an increase given by an estimate of the absolute increase in the temperature</td>
<td>Non applicable</td>
</tr>
</tbody>
</table>
Appendix C: Transcripts of the Interviews

- Participant 1

[Question 1]
Rs: This is your first and second question. I’d like to know what you were thinking when you answered these questions, please.
P1: Right, I looked at the different times and first I tried to figure out the time difference between Berlin and Sydney and then I used the time difference and also the graph.
Rs: Did you have any problem with the time difference? Was there anything that particularly focused your attention?
P1: No. I used the diagram really.
Rs: What about in question 1.2?
P1: I started a process of elimination, a sort of trial and error but it happened that the first one I used worked. So I just found a suitable time in Sydney first by minusing 2 hours from question 1.1 and then I used the time difference from question 1 to figure the time in Berlin which also fit a reasonable hour, and yeah. Rs: At this stage, was there anything that stopped your thinking process, anything that was difficult?
P1: The only difficulty was it had to be a specific time in both time zones.
Rs: Anything else?
P1: Not really.

[Question 2]
Rs: Now what about with question 2?
P1: Well, I had myself a problem with that, a car, I guessed it. I thought it would be maybe 2 meters by about 3 or 4 meters so I timesed the total area of the car park and then divided it by 6, like 2 by 3 but I don’t know what a total car would be, and I came up with 833.33 and rounded it to 833.
Rs: Was there anything that you noted in particular in this question in terms of difficulty or easiness?
P1: It is harder than I thought because they didn’t give the dimension of each car park; it would be very easy to figure out if I had the dimensions of each car park.
Rs: So, what factors did you consider to solve this problem? You’ve already mentioned the lengths…
P1: Yeah, I really thought on it but now hopefully other things could open to do.
Rs: I see, could you now draw a mental picture of a car park?
P1: OK, car spaces.. there must be exit roads [while drawing]
Rs: So you have drawn car slots and exit roads, haven’t you?
P1: Yes, but I didn’t do the roads for that [to test the solution].
Rs: So did you consider these factors when solving the problem?
P1: No, I didn’t honestly. If I did consider probably it would be nearly 500 cars.
Rs: Could you explain it to me?
P1: I’ve sort of done that. So these 833, now probably go down to 500 because if you take out some of the cars for the roads, there would be fewer car parks because of the roads, I think, so I think it would be 500.
Rs: Did you consider the context of the item when you solved it?
P1: None of those [referring to car spaces and exit roads], ummm as they sort of wanna get as many as possible I think.
Rs: With this information, with this mental picture and the information that the context has provided you, do you consider that it would be easier now?
P1: Yes, knowing the exit roads and putting them into the question would make it easier.
Rs: Have you been in a car park before?
P1: Yes, I have.
Rs: And the mental representation of it?
P1: Yes.
Rs: So why couldn’t you connect your picture with the problem?
P1: Ummm, when I first did it, the problem, I was… I couldn’t think of things that I see now, but when I did it, when it crosses my mind again I can pick up these little things, because I forgot about it and now I can go back to them.
Rs: Was the Mathematics to solve this problem difficult?
P1: There was just the area and the times and dividing on that and a bit of logic in it I guess.
Rs: So what was the difficulty of this problem?
P1: It was to connect these little things [the picture of the car park] with the question.

[Question 3]
Rs: What can you tell me about your solution process in this question?
P1: With the poll, you want as many people as possible to have an accurate answer, so I focused on 3 and 4 because they had more people, and then I thought that if people volunteer for it they could be the same sort of people who volunteer for the poll, so if it is randomly selected there is more chance of having different opinions though.
Rs: Have you been questioned for a poll?
P1: No, but we did statistics in Maths, and I know that I need to get rid off the biased population at least and you have to randomly select your sample, as random as possible.

[Question 4]
Rs: Can you explain your interpretation of the graph?
P1: Well I thought on the graph, the 1999 doubled the 1998 when you just looked on the bars, but if you looked on the left, on the actual degrees, you think that 1 degree isn’t that much.
Rs: Your conclusion is supported by the facts you just mention, but why do you believe that 1 Celsius degree is not much?
P1: With temperature, I mean for my experience in Melbourne, we have days of 44 degrees and also days
with less than 10 degrees, so with the water temperature it would be able to change the same as the temperature weather.

Rs: Has your previous experience with temperature influence the way you have argued your answer?
P1: Yes, it has.

Rs: Well, now if you put your experience away and just looked at this graph, would you conclude the same?
P1: Ummm, I think I would, because 1 is like the lowest whole number there is, and if it is less than a whole number it is not very much. But I don’t know how water temperature changes I linked it with the weather temperature.

Rs: You have answered in two ways, focussed on visual aspects and supporting your answer by the fact that 1 degree is not much, but would your answer change if you could avoid your exposure to a temperature context?
P1: I think I would have the same conclusion. It is temperature! I need to know how it changes.

---

**Participant 2**

Rs: Could you please explain to me what your thinking was when you solved this question?
P2: Ok, well, I needed, I got the actual time in Sydney, 7pm, and in Berlin it was 4 hours, no hang on, yes, 4 hours ahead of Sydney, so I just counted 4 hours ahead from 7pm.

Rs: I see, was there anything in particular that drew your attention? The charts, clocks, the context?
P2: Not really, I understood everything well, everything that I needed to know was there, ummmm…

Rs: did you consider the Greenwich time?
P2: No I didn’t actually.

Rs: Ok, what about question 1.2?
P2: For this one, I set out the time that Mark and Hans could talk, and then I just worked out that blocks of time that they could talk, and then just worked out the 6pm in Sydney, is 9pm in Berlin, both fall in the same free time block.

Rs: Were you confident with your solution process?
P2: Yes for this one, yes.

Rs: Did you go back to the question in order to check it?
P2: Yes, I did.

[Question 2]

Rs: Could you explain to me your solution process in this question?
P2: Ok, well I had a lot of trouble with question 2, because I found that this question tells you about the perimeter of the field but it didn’t tell you anything about like how big the car park is gonna be, the driving area in the car park, so I didn’t actually know how many cars could actually fit, because I didn’t know that.

Rs: But could you be more precise, please?
P2: Well, basically, it is extremely rough, I worked out the area of the car park, and then I only guessed the average a car length would be, and then I just divided the area by the lengths of the car, and then I got 500, so.

Rs: Did you divide the area by the length of the car and that was your estimation?
P2: Yes.

Rs: You have already mentioned that you consider the length of a car as estimation, did you consider anything else?
P2: Ummm, I didn’t know what to do with this.

Rs: you also mentioned exit roads, didn’t you?
P2: Yes, like the drive ways between the two cars, and like even the space that each car park provides for every single car.

Rs: Could you drawn for me your mental picture of the car park?
P2: Ok, ummm…, the drive ways, the space. This is not very accurate.

Rs: I understand it, no problem. This mental picture played a role in your solution process?
P2: Yes, it helped me a lot. But I think I could work it out a little bit more, like I could divide it up and make the spaces even. It probably would help me, it would be more accurate. Ummm but it is like an estimate.

Rs: Anything else to add about it?
P2: Ummm… I think it helped me. I realised there are many things. I don’t know the lengths and measures involved, so it overwhelmed me, and I didn’t know what to do with the question.

Rs: Let’s imagine that this picture would be present in the problem; can you speculate of an effect on your solution process?
P2: I think it would be a lot easier.

Rs: So what it was difficult?
P2: I think to relate the picture of the car park with the problem, yes… it was difficult.

Rs: And what about the mathematical component of this question?
P2: The concept of measurement, area?

Rs: [I nodded positively]
P2: No, I found that all the measurement stuff is kind of easy for me, but it is just I was limited on details and stuff I needed. They weren’t provided.

[Question 3]

Rs: What can you tell me about your thinking in question 3?
P2: I said that poll 1 and poll 3, because the more you’ve got the better, and randomly selected means that it wouldn’t be biased or you can’t know what is gonna happen with the results, then you get more accurate results.

Rs: When you mentioned the more you got the better, what are you referring to specifically?
Participant 3

**[Question 1.1 and 1.2]**

Rs: Could you explain to me your solving method for these questions?
P3: Well, for the first one, for 1.1, I figured out what the Greenwich time was, which is 12, and then the difference between Greenwich and Sydney and Greenwich and Berlin, so it is 1 hour between Greenwich and Berlin, and 10 hours between Greenwich and Sydney, and I just subtracted to find the difference between Sydney and Berlin.

Rs: Was it difficult for you?
P3: No, it wasn’t, it was fine with the clocks.

Rs: What about in question 1.2?
P3: Oh well, the time, well it wasn’t really much for this one, because it didn’t say about the actual people what would be a good time for them, so really the only one they really gave you is the suggested time, because the hours are reasonable...

Rs: When you say reasonable, what do you really mean?
P3: Well, it is not really late in the evening and really early in the morning, so yep

Rs: Well, at this stage was there anything in particular that stopped you or helped to solve the problem?
P3: Only the previous question I guess, it was reasonable.

Rs: Anything else?
P3: No.

**[Question 2]**

Rs: What about here, what was your thinking to solve it?
P3: Oh well, I found the area of the table, and the area of one tile, and then I just divided to find how many tiles would cover the whole table.

Rs: The Mathematics involved in this question...
P3: It was fine.

Rs: Did you notice anything in particular when solving the question?
P3: Yes, the table may not necessarily be like a rectangle, it could be a different shape, so it doesn’t actually say the tiles and the table are rectangular or square.

Rs: What can you tell me about the word about in the answers?
P3: It didn’t mean exactly that, so it wasn’t exactly accurate, like it would give you the feeling that the answer was accurate ...

Rs: Did it have an impact on your final answer?
P3: Ohh, not really, but I didn’t think about it, I think it wouldn’t change my answer...

Rs: Why do you think this word was in the answers?
P3: When it is about, it means probably it is not exactly that answer, and because I got the exact area, that was I thought maybe the table wasn’t exactly or the tiles that shape [rectangular shape].

Rs: Did you consider the spaces between tiles in real life?
P3: No.

Rs: And why do you think you didn’t?
P3: Probably I didn’t think about it, it doesn’t say anything about considering it, but if I did, it probably still would be about 500 anyway, because the space wouldn’t have really much impact because it is not really big space.

Rs: Ok I see, but from where do you know that the space is not really much?
P3: Well, the differences between B,D and E, like about 150 in D and E, so that would have an impact so...

Rs: Ok, but your conclusion about the space was based...
P3: I didn’t think about the space between tiles at all

Rs: But now?
P3: Now? Well, because generally with tables and tiles there isn’t a great lot of space between so, thinking about that it wouldn’t make much impact I guess.

Rs: So far, you have established that you didn’t consider the space among tiles as a factor...
P3: No, I didn’t.

Rs: Why didn’t you consider it?
P3: Probably because the question, I was so sure, you didn’t get much information and you have to work with what you’re given.

[Question 3]
Rs: Now, what about with this question, Support for ...?
P3: Umm, I thought if you’re… going to get to an accurate answer, you need to have a randomly selected people. To the ones who volunteer you can’t really use them, because they are probably… if they are volunteering probably they all have a positive opinion generally, and the ones who are all women, you couldn’t use because you need both, because the swimming pool is not just going to be for women, and I chose poll 3 I think. Oh and the bigger the sample size the more accurate it’s gonna be too.
Rs: I see what you are saying, thanks. Have you been polled before?

[Question 4]
P3: Well, according to this the reporter says that was a huge increase, but here the increase is just only 10 robberies, so it depends on the people who have been robbed, but we are looking at 500 and what is it?... anyway, 10 is not a huge difference, so it probably is a little exaggeration.
Rs: When you say 10, why did you conclude it is not a big increase?
P3: Why? Well because the number of robberies is 500, and it is pretty high and 10 is a small part of that.
Rs: So, when you give that answer, are you assuming that you don’t need to give further details in order for me to understand it?
P3: No, because 10 out of 500 is not much.

•  Participant 4

[Question 1.1 and 1.2]
Rs: What can you tell me about your solution process in these items?
P4: I thought that in Melbourne it was 10am and in Delhi 5:30, so I just wanted the difference between these two to find how for apart they are.
Rs: Did the clocks help you to find the difference?
P4: No, not really.
Rs: What about in 1.2?
P4: Well, I looked at the time they can’t broadcast, between 9 and 4:30, and between 11 and 7, and I thought of the time that in between they can broadcast so I just picked a random number and worked out the other time in the other place, yeah, and if the two worked in the same… I got the answer.
Rs: Did you have a concern about what a good time would be?
P4: Not really, I didn’t … no, like exactly what time is a good time, so that it would be tricky, so I guess…
Rs: Did you go back to check the solution?
P4: No, I didn’t.
Rs: Did the fact that this question is about the 2010 Commonwealth Games produce any impact in order for you to check the solution?
P4: No, no.
Rs: Did you see it as a probable real situation?
P4: Yes, sort of.
Rs: Was there anything external that impacted your solution process?
P4: No.

[Question 2]
Rs: What about your process in this question?
P4: I just took the fact of the field which is 100 by 50, and that area is 5000, and I just assumed that each person just takes around 1m², so if I was standing up it would take 5000.
Rs: Have you been in a rock concert before?
P4: No.
Rs: Have you seen one?
P4: Yes, on TV.
Rs: And how does it look?
P4: Very squished.
Rs: Did you consider other external aspects in the solution?
P4: Yes, people may be in a wheelchair or something so it may take more [space] ummm
Rs: Does it mean that you considered exit roads?
P4: Yea, that is it. But they are squished together so it wasn’t that much space between everyone…
Rs: Could you make a guess now with that information?
P4: What do you mean, do it again?
Rs: Yep.
P4: There will probably be more. 10 000 I guess. Or maybe C.
Rs: Why C?
P4: I don’t know… there would be more people per square meter so it looks like a rock concert I’ve seen.
Rs: When you solved the item did you think on it?
P4: No
Rs: What about now?
P4: Yes, if you think about what a rock concert looks like you have a better chance of solving the problem correctly; it is more like real life I guess.
Rs: So, would it be fair to ask this item if you haven’t seen or experienced one before?
P4: Yes and no. If it is yes, it is a common sense that there would be a lot of people there very squished and no because if you haven’t seen probably you wouldn’t know how to start it because you have to think about it.
Rs: Lastly, was the Mathematics to solve this item difficult?
P4: No, but it was important as the assumptions to solve it.
[Question 3]
Rs: Now with this question, what can you tell me about it?
P4: Well, when I looked at it I looked at all the percentages and the information in the brackets because some of them have the readers phoning in, so I just instantly excluded them because they are more likely to be biased, they are more willing to... ummm yeah, so then I just took newspaper 1 and newspaper 4, and then I took newspaper 4 because it has a bigger sample of people that they chose, so it is more likely to be accurate.
Rs: I see; but in your answer you chose newspaper 3, didn't you?
P4: Ohh yes, I meant newspaper 3, sorry.
Rs: That’s fine. When you were reading the information provided there was anything that drew your attention?
P4: No not really, I can’t remember.
Rs: Ok, if I ask you now the difference between 1000 readers phoning in and 1000 randomly selected citizens with voting rights, can you see any difference?
P4: Yes, because those people who want to phone in are more likely to be supporting the president because they are more into it...
Rs: Have you been polled before?
P4: No.
Rs: And your notion of polls, where have you got it from?
P4: News, TV, a bit of everything...
Rs: Has that previous knowledge helped you? If yes, in what way?
P4: Well, we have also been talking in class if you have a larger random people it is more likely to be actually correct so I just took that and I used in my answer.
Rs: Finally, was there anything external that influenced your solution?
P4: Not really.

[Question 4]
Rs: Can you explain to me your solving method please?
P4: Well when I looked at it, it is monthly train ticket prices and it was like between $300 and $400, umm... I just thought that a train ticket can’t cost that much money and the title wasn’t very correct. It was confusing at the start.
Rs: So you are telling me that your answer was based on your knowledge about the ticket prices?
P4: Yes.
Rs: But did you consider other kinds of monthly ticket prices? Melbourne to Ballarat for instance...
P4: No, I don’t know that price I just knew the Metropolitan one.
Rs: So, based on that information you solved the item?
P4: Yes
Rs: Well, if you based your answer now just by looking at the graph, no title, would you give the same answer?
P4: I wouldn’t know what it is about, I would be really confused because there is no information telling me what the graph is about.
Rs: Let’s imagine now that instead of saying “A TV reporter showed this graph and said” I would say “An overseas TV reporter showed this graph and said” and the same information is provided, would your answer change?
P4: Ummm... I don’t think... maybe it’s not that correct because he hasn’t actually... if it is in Australia he hasn’t lived in Australia to see exactly how everything is... but it may also be correct because he needs to research well. At the end you need a context for prices.
Rs: Have you considered anything else rather than prices?
P4: No, just prices.

- Participant 5

[Question 1]
Rs: Could you explain me to your solution process for questions 1.1 and 1.2?
P5: What I did was, I just counted the difference in hours between Berlin and Sydney and then I just, what worked from that, so it was 7 in Sydney and added 17 hours to Berlin time to find that it was 10AM.
Rs: Did the clocks impact your answer?
P5: Yes because to count it I thought, you know, from that spot around so I counted the hours from the clock.
Rs: And for question 1.2?
P5: I don't know. I chose these times because they seemed like normal hours for both to be able to get on the internet and so like that. You know, you couldn’t go on like late night or too early in the morning, so I tried to choose a time in between.
Rs: I understand, but what do you mean by normal hour?
P5: Not too late, not too early. You know it might be a good time in Sydney like 7, but it would be too late too much during the day in Berlin.
Rs: Let’s imagine that this question would have time constraints for the communication between Sydney and Berlin; would it change the accessibility to answer? Why?
P5: Yes, it would be easier to choose time with those constraints on it, because you know about what would be a good time to choose.
Rs: Regarding your previous definition of a good time to chat. What aspects did you consider to support it?
P5: I chose a good time from like, for people would go to work or to bed something like that. That was how I chose the time.
Rs: Was it based on any contextual aspect?
P5: From my personal experiences, probably.
[Question 2]

Rs: What about this question, the tiles. Could you explain your solution process?
P5: I found the area of the table, and then the area of the tiles, so then I... I think there is 100cm in length for the table, so you only can fit 20 tiles across or you could fit 50 tiles in any case, and then the table was 50cm in width so you better fit 25 or 10 tiles, so then I timed 20 by 25 and 50 by 10 in both cases the answer was 500, and that was my answer.
Rs: Have you noticed or thought of anything in particular when you solved this item?
P5: No... why?
Rs: What can you tell me with reference to the word about in the answers?
P5: What do you mean?
Rs: According to your solving process your answer was 500 tiles but you chose an alternative which stated that you would need about 500 tiles, so did the word about influence your answer? Did you think about this word?
P5: I see, I didn’t think about this until now but I guess I could change a bit, you need space between tiles.
P5: and this space is...?
Rs: Not much space.
P5: Probably’d stay the same.
Rs: Are you thinking about the space between tiles?
P5: Yes, now I am.
Rs: Your assumption about the space is based on?
P5: Tiled surfaces that I’ve seen.
Rs: I see. Was the Mathematics involved in this item difficult?
P5: No, it wasn’t but the space between tiles would make it a little more difficult.
Rs: Why do you think you didn’t pick up the space factor?
P5: Because you were told about the dimensions of the tables and the tiles. The question doesn’t say anything about space between tiles so I didn’t pick it up until now.
Rs: So it looks like a classical Mathematics question?
P5: It looks like something routine, something that doesn’t make you think about space and stuff like that.

[Question 3]

Rs: What about your solution process in this question?
P5: I chose my answer by reading through the polls and the words that they used, volunteers or randomly selected, and then women or residents; so obviously more women could want the shopping centre or if you are a volunteer, probably if it is someone out in the street doing something you are more active or you want the shopping centre whereas if you are randomly selected you may not used, things like that, so the randomly selected residents would be the most accurate.
Rs: Have you been polled before?
P5: No.
Rs: So, from where have you got this knowledge?
P5: Well, in class at the moment we are doing Statistics so concepts like selected versus volunteer and stuff like that, but also just you know that if you randomly select someone rather than people who offer to do a poll results would be different.
Rs: That view comes from?
P5: Common sense.

[Question 4]

Rs: What can you tell me from this graph which represents the number of robberies?
P5: To figure it out I just looked at the scale that was used and then I realised that the difference is only 8, so in one year there were 10 more robberies, but in considering that numbers are so high, they are in the 500s, so it really doesn’t make much difference, that was what I thought.
Rs: I see. But you have written that 10 robberies don’t represent a hug increase. So you are assuming that people who may read your answer understand your thinking straight away? I mean that 10 is not much?
P5: Yeah, it is not much.
Rs: What would happen if you present this argument to your family or classmates? Would you need to explain it to them?
P5: I wouldn’t need to explain this, not much; you would know that 10 out 500 is not much.

• Participant 6

[Question 1]

Rs: Could you explain to me what was your solution process in these questions, please?
P5: Ok, what I tried to do was work it out, I thought about the difference between Berlin and Sydney and I answered it.
Rs: Did the clock help you?
P5: Yes, yes I am very visual. So to be able to work on it was a big help for me.
Rs: What about in 1.2?
P5: I wrote down the times next to each other what time was in both cities, again very visual, so I had to figure out what time it is in both places at the same time, and that’s pretty much what I did.
Rs: Now, what can you tell me about what a good time for Mark and Hans is? What does it mean for you?
P5: Basically when they are able to chat.
Rs: And that information you have got from?
P5: I got it from here.
Rs: Was there anything external that you brought into the solution?
P5: No, to be honest. Just finding a guess time when they were available to chat, not really thinking about anything else.
[Question 1]
 Rs: This was your first question, this was related to a time difference between two cities, can you explain your solution process, please? What did you take into consideration to solve it?
 P7: Well, because the time difference is due by Delhi, sorry it is after Melbourne, so 4:30 seems like a logical time for them.
 Rs: That time difference where did you get it from?
 P7: From the chart.
 Rs: Was the chart difficult for you?
 P7: No, it’s pretty all right. It seemed logical for the answer. So, then for this one I just added the time difference onto the 4:30 and I got the 9PM, it seems like a similar time because if it really was the Commonwealth Games they would be finishing when the evening would be starting as well.
 Rs: You mentioned the Commonwealth Games, have you considered them in any way when you solved this item?
 P7: Well, this situation makes sense. It’s gonna happen in Delhi on October I guess so you have to work out time differences between Delhi and Melbourne.
 Rs: Have you linked the question context and the usual time of broadcasting into the solution?
 P7: Yes, because the night sport would be starting in Delhi and in Melbourne it’s around 9PM and it’s the time when people decide to go to bed, or maybe they can tomorrow morning see the results on the morning news too.

[Question 2]
 Rs: Are these spaces [pointing to the car slots previously drawn by the student] equal to each other in a car park?
 P5: Umm, no, not usually. Ohhh yes, yes, they are. And also the drive-ways are.
 Rs: With this new information, can you make a roughly estimation of how many cars would fit in a car park?
 P5: Yes, but I don’t know how big a space is.
 Rs: We don’t know the size of a car slot of course, but with all the new information that you’ve got now, can you make a fair guess?
 P5: I don’t know… probably it is wider than 2 metres I guess and probably … maybe 5 metres in length, is that all right?
 Rs: With this information, can you proceed?
 P5: Yes, I can. I didn’t think about it before [while working]… there are 20 [car slots] across, that’s not right, what am I doing?
 Rs: Take your time [the bell rang]
 Rs: Ok, you have made an estimation of how many cars would fit in a car park. Quickly, what stopped you from solving this question?
 P5: I don’t know. I didn’t think about it before. I guess not many people would assume it. But if I had a picture it would be easier.

• Participant 7

 Rs: What did you think when you solved this item?
 P7: Because with the rock concert when people go they are all different sizes, so you need to be able to seat everyone in not just … I just doubled both sizes and I came up with 20 000. There would still be room but, you know, it wouldn’t be enough room.
 Rs: Can you tell me why you doubled the lengths of the field?
 P7: Because it’s 100m and no one person is like 1m long, so that would work, it would be stupid to fit every meter, so I doubled it up.
 Rs: Did you bring any external factor into the solution? You have already mentioned the space among people…
 P7: Yes, and the sizes [referring to people] whether they would be like small, tall, big and things like that.
 Rs: Have you been in a Rock Concert before?
 P7: Not much, because I was sitting. I went to a concert and I was sitting.
 Rs: Have you seen one before?
 P7: Yes I’ve seen some.
 Rs: Can you tell me the picture of a rock concert you have seen before?
 P7: Ummmm
 Rs: In terms of space between people?
 P7: There seems to be enough space for people to sit down and walk to the back or up to the front.
[Question 3]

RS: What can you tell me about this item?
P7: I picked newspaper 3 because it seems that with these... people phoning in you don’t know exactly how phoning in is, but with these ones [referring to alternative A and C] they are people who can vote, they will all be totally random, whereas these people [referring to alternative D] could be in the same group, so obviously they’re gonna show their support for what they believe in. And I thought 39% is like high enough but not too high either, so then 1000 is enough people polled.
RS: have you considered that phoning in could be a way of voting?

[Question 1]

RS: This is the world chat times item, could you tell me what was your solution method please?
P8: I thought Greenwich is like the main time line, so it is like zero, and because Berlin was 1AM, so it was 1+ hour, and in Sydney was 10AM, so it was 10+ hours. So, to find the time I just minused in between the two cities, and that gave me the answer.
RS: For you, was Greenwich important to solve the item?
P8: Yes, I know that Greenwich is like the main ummm... the centre, the zero hour.
RS: If you know that… would this change affect you?
P8: I probably would handle it a lot more difficult, and I wouldn’t know how to go with the other question.
RS: Ok, I see, so now let’s focus on question 1.2, what can you tell me about it?
P8: It was a difference of 9 hours between the two cities, so I just basically worked out... I used that answer [referring to 1.1] and took half an hour off on both times, that’s right.
RS: What was for you a good time when you solved this item?
P8: So, I thought when I did my answer, half an hour is probably not like that much time for them to be able to talk but I think ... from like 8 o’clock they still would be able to talk anyway, so I just put it there as an answer...
RS: you haven’t answered this one, can we have a quick look at it now?
P7: Yes, sure.... I couldn’t support the statement because the graph more or less shows the ticket sales more than the prices, because this is the way it looks to me, so I am assuming that the amount of tickets bought is due to the prices...
RS: So, are you relating this graph to the amount of tickets sold rather than the prices?
P7: Yes, because it’s the same month and the same amount of people have bought the tickets, so the prices went up, some more money was coming in due to the tickets
RS: Have you considered this graph is about years rather than months?
P7: No.

• Participant 8

RS: And your knowledge about polls where have you got it from?
P7: Experience, my personal experience.
RS: Now, what can you tell me about the Mathematics involved to solve this item?
P7: Pretty simple, I think ,and also you have multiple choices to choose the answer.
RS: Were the context and the Mathematics in this item difficult to connect?
P7: No, it wasn’t too difficult.
[Question 2]
Rs: What about with this question?
P8: I thought it was a bit of a trick question, because I talked with some of the other girls about this, and they said they had tiles and other things, I thought the main thing was an accurate measurement for each car, so I estimated... I said 4 by 2 metres roughly.
Rs: And that estimation where did you get it from?
P8: Standard, you know, like a mid-size car sort of probably about 4 or 5 metres in length and 2 metres in width, but then I thought like if it is a car park there is gonna be room between each car, because the cars can't be packed in, so...
Rs: And that picture you've got, did it help you?
P8: Yes, it did help me but also it kind of confused me because I didn't know quite if the question was asking like how many cars all pack in together or if it was a car park situation when there are spaces between cars.
Rs: Could you identify the Mathematics involved to solve the question?
P8: The area, I found the total area of the field then how many cars of a certain size would be able to fit in that area.
Rs: Could you draw for me a picture of your car park?
P8: [while drawing] Basically it is like a rectangle with roads like... that and gaps between cars...
Rs: when you solved the question did you think of car slots?
P8: I just thought of an average size of a car...
Rs: With that estimation that you've made?...
P8: My mental picture, my estimations did help me but also kind of confused me because it wasn't exactly information, like all I needed wasn't there I needed measurements and everything.
Rs: Even though when you were asking for an estimation?
P8: Yes, I just thought basically there is gonna be a few metres gap, so I answered 500, because I thought 200... the area of the field is 5000 metres and then my calculation was 625 which was closer to 500 than 2000, so that is why I chose that answer.
Rs: For this item, did you consider that you need more external factors than mathematics to solve it?
P8: I thought I needed a bit of logic because I know some people could have taken it more literally how many cars could be packed in, but then I thought you know that it doesn't really work because you can't get them out.
Rs: What would you add to this problem to solve it more comfortably?
P8: If the question had said it needs to be that much space for the cars to be able to get out or whatever you know that would help. When I was doing it not enough information was there... but I found it easy to do.

[Question 3]
Rs: Now, what can you tell me about question number 3, the poll one?
P8: I said the third one would be more reliable because there is more of the sample like a larger sample and I also thought... umm... depending on what the topic of the poll would be... if people were to volunteer they could be biased... like if they say yes we want to do it, they may have more opinion than from someone who's been chosen randomly.
Rs: According to you, what kind of poll would be appropriate to be answered by volunteers?
P8: Umm... maybe something like... based in schools... something where the group of people are... you know... where the poll was on something like who is your favourite politician... and people were to volunteer... you know they may be biased whereas people who are chosen randomly... like may not have that much an opinion on it anyway.
Rs: Have you experienced a poll before? Have you been polled before?
P8: Not really, I've just done like surveys and things but like school based things.
Rs: From where have you got the necessary knowledge to solve this item?
P8: From Math classes.
Rs: When you were answering the question did you think of external factors that may influence your final answer?
P8: Yes, well I thought... my main thought you know what would this poll be on, you know, they would wanna have volunteers or people to be randomly selected... and I just thought this [alternative C] would be the most appropriate answer in a general sense.

[Question 4]
Rs: Finally, what can you tell me about this item?
P8: I thought it was reasonable but I thought it doesn't really tell you much really. It doesn't tell you, you know, it could be difference in seasons... umm... things like global warming and stuff that you know...can affect the answer. And also they haven't got like a source or anything like that, and say that it is a proper scientific research... umm
Rs: Do you consider that the information you mention above would be important to give an appropriate answer in this item?
P8: I basically thought like... you know... temperatures changes all the time... you know... umm... I don't know... Well I thought of global warming the earth is warm and everything is heating up anyway... you know... so, it's not like a real big rise anyway, it's not 10°... it's around 1.5°, so it's not a huge difference.
Rs: So you are trying to say that to answer this question you need to bring external information? Aren't you?
P8: In this one you need to think outside the square... what could be making the temperature change, and also it doesn't really have like a topic like global warming... you know... getting worse... you know... factors that may increase it
[Question 1]
Re: Can you tell me what was your thinking when you solved these items?
P9: I took the time back from Sydney I counted the time between them and I took it back.
Re: What about with 1.2?
P9: I put the time they cannot talk with each other and then I put it in there and I worked out between them, the time that would be appropriate, so after 4:30 obviously there.
Re: Did you think of anything external when you solved 1.2?
P9: Not really.

[Question 2]
Re: What about now with the car park?
P9: I didn’t get it really. But I timesed the length and the width and …umm… it was equal to 5000 and this was my final answer.
Re: Why did you answer that?
P9: Because I was trying to find the area but I noticed I was almost wrong in my answer. I drew just a rectangle to find the area of the car park.
Re: Oh ‘right, you mentioned your answer was wrong, so what did stop you or hinder you to solve this item correctly?
P9: I knew 5000 wasn’t the answer because this is the area of the car park…but I didn’t think of anything else.
Re: What about this picture? [A picture of a car park was drawn by her on her paper and pencil test when she did the test]
P9: I wasn’t thinking really I just drew lines.
Re: Do you know what a car park looks like?
P9: Yes
Re: Can you now draw your picture of a car park, please?
P9: Yes sure, just one like that [the original picture mention above] but …with exit roads and car spaces.
Re: What can you tell me about the size of the car slots you’ve drawn? Are they similar, different, etcetera?
P9: They are the same.
Re: Well with this new information brought by your picture could you make a guess now?
P9: Ok…[while working]
Re: Take your time.
P9: It would be about 500.
Re: Can you tell me why, please?
P9: Because… a hundred meters would be less than a hundred cars so that’s 200… 250 … and then it makes now 500, is this right?
Re: Yes it is!!!
P9: When I was doing it I didn’t think of the original one I did but my common sense tells me that it’s better.
Re: why?
P9: Because I related the question to my picture of the car park.
Re: Why do you think you couldn’t relate it before?
P9: I don’t know. I just didn’t think logically. I didn’t think to put it in a real way. I just thought of working mathematically.
Re: What do you mean by working mathematically?
P9: Like working doing lengths times widths. I thought I had to do a formula to work it out. I didn’t think in a real way, so I didn’t relate them.

[Question 3]
Re: What about with the polls, which of these would be the more reliable one?
P9: So, I chose this one because… umm… because if that was volunteered they may have been in the same group of people but if they were picked up at random they could be anyone. And 500… umm… I thought 1000 would be better because there is more variety, it’s more than 500, because with 500 you can have more of the same things, but with 1000 there is more variety.
Re: When you were solving this item did you think of anything external in particular?
P9: No, I just only thought of the number [size of the sample] and of random volunteers.
Re: From where have you got the knowledge necessary to answer this item?
P9: I didn’t relate this one to Maths really. I just thought the higher the poll the more… like chance of being random…
Re: You’ve mentioned you didn’t relate this question to Maths, so where have you got this knowledge from?
P9: Umm… common sense, I don’t know. It’s not common sense. We are doing Statistics at the moment, so from there.

[Question 4]
Re: What can you tell me about this graph?
P9: Ok, I took the year 1999 I also took the degrees in between them, from 1998 and I got 1.75° approximately, this was the rise in one year.
Re: According to your response it was a reasonable increase over one year period. So, why did you think that almost 2° was a huge increase?
P9: I don’t really know I just said it was a lot because 2° in water must be a lot… but I don’t really know about temperature.
Re: I understand you, but could you tell me more about what you thought of when you judged the 2° as a huge increase, please? Where have you got this knowledge from?
P9: Well, 2° it’s not a lot but maybe with temperature you can feel it more…
Re: I see, let’s suppose that instead of temperature you had a money context, would you judge the answer in the same way as in this question?
P9: No it wouldn’t be much
Re: So what’s the difference?
P9: Well, because a 100° may be a lot more but $100 it’s still a lot but it wouldn’t be as much because I am more familiar with money.
Participant 10

[Question 1]
Rs: Could you tell me what your solution method was in these questions, please?
P10: I looked at the difference between the times like how many hours between Melbourne and Delhi here and then I worked out however much was the difference...4.5, and that was pretty much what I thought.
Rs: What about with question 1.2?
P10: With this one I wasn't 100% sure but I was looking at the local times so I thought like similar to...in Melbourne we have the news around 6 or 7 o'clock so I kind...I really thought realistic...umm I wasn't too sure. It may be hard to know the local time in Delhi compared to Melbourne...
Rs: Why was that?
P10: Because we are in Melbourne compared to someone else in the world or another area so I wasn't sure.
Rs: Was there anything in the question that could relate Melbourne times to Delhi times?
P10: Probably... maybe... you know... they left out; they didn't give much information to help you, kind of...
Rs: Did you consider the time chart when you worked on 1.2?
P10: Umm... honestly... probably not. I was kind of... no I didn't. Probably, my knowledge about sport times made more common sense to me then maybe this [pointing to the world time chart].
Rs: Why do you think it makes more common sense?
P10: Just common knowledge what makes more sense, realistic... yeah.
Rs: in this case your knowledge did help you?
P10: Yeah... this helped me a bit.

[Question 2]
Rs: What can you tell me about this rock concert item, please?
P10: With this one I timesed this together [the length and width of the field] because I needed to find the total number of people attending the concert.
Rs: Were there any assumptions you made when you solved this item?
P10: well, I said this is the area for the amount of people you can sit in there and also some standing as well.
Rs: Have you been in a rock concert before? If yes, can you tell me what it looked like?
P10: Yes I have, I sat but there was also an area where you could stand. There was a big space between chairs, rows and things like that.
Rs: Was it important for you to have been in a rock concert before in order to solve this item?
P10: umm... probably understanding how people are sat helped me.

[Question 3]
Rs: Finally, what can you tell me about this item, please?
P10: Yes, I just had a look at the diagram, so I said that the graph shows an increase of more than $150 in the last year, so...
Rs: Why did you choose $150 as a limit for a huge increase?
P10: Just because it was talking about between 1998 and 1999, so I just said pretty much what it was showing, I mean the increase in the last year $150...
Rs: And that amount is a huge or small increase?
P10: A huge increase in a year.
Rs: So what aspects did you consider in supporting your conclusion?
P10: Well, I think that $150 in a year is quite a lot for them to go up... umm... probably considering Melbourne prices. To go up $150 in tickets in a year is a lot.
Author/s:
Almuna Salgado, Felipe Javier

Title:
Investigating the effect of item-context on students' performance on mathematics items

Date:
2010

Citation:

Persistent Link:
http://hdl.handle.net/11343/36236

File Description:
Investigating the effect of item-context on students’ performance on mathematics items

Terms and Conditions:
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.