Using example generation to explore undergraduates' conceptions of real sequences: A phenomenographic study

by

Antony W. Edwards

A Doctoral Thesis

Submitted in partial fulfilment of the requirements for the award of
Doctor of Philosophy of Loughborough University

in the
Mathematics Education Centre
School of Mathematics

May 2011

Declaration of Authorship

I, Antony Edwards, declare that this thesis titled, "Using example generation to explore undergraduates' conceptions of real sequences: A phenomenographic study" and the work presented in it are my own. I confirm that:

- This work was done wholly while in candidature for a research degree at this University.
- Neither the thesis nor the original work contained therein has been submitted to this or any other institution for a degree.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.

Signed: ______________________________

Date: ______________________________

“I can’t understand anything in general unless I’m carrying along in my mind a specific example and watching it go.”

Richard Phillips Feynman

Abstract

Using example generation to explore undergraduates' conceptions of real sequences: A phenomenographic study

by Antony W. Edwards

This thesis uses an example generation task to explore undergraduate students' understanding of basic sequence properties in Real Analysis. First, a review of the literature looks at three areas of research: the transition to studying mathematics at the tertiary level, examples and the process of example generation, and the learning of Real Analysis. It notes a lack of research on how students interact with simpler definitions in Analysis, and suggests that an example generation task is an ideal research tool for this purpose.

Then, two pilot studies are reported. The first gave 101 students an example generation task during a lecture. In this task, students were asked to generate examples of sequences that satisfied certain combinations of properties. In the second pilot study a similar task was given to six students in an interview setting with a 'think-aloud' protocol. These pilot studies found that many students gave sequences that did not satisfy the requested properties, whilst other students gave examples that were not sequences.

The thesis then reports on a main study in which the example generation task was completed by 15 students during an interview, and 147 students during classes. The interview data is analysed phenomenographically, with results presented along four dimensions of variation, where each dimension describes different ways of experiencing an aspect of sequence example generation: Using Definitions, Representation of Sequences, Sequence Construction Strategies, and Justifications. The larger-scale class data is then analysed by Rasch Analysis to objectively rank the questions in order of their difficulty, and to show that the interview-based responses reflect those in the wider cohort.

By asking students to generate their own examples of sequences, this thesis has furthered what is known about student understanding in two areas. The first area is how students understand content related to sequences in Analysis. The thesis considers students' understanding of how sequences can be represented, how sequence property definitions can be combined and how definitions affect sequences in different ways. The second area is how students interact with example generation tasks, the approaches that are effective when students are trying to generate examples, and the ways students justify or check their answers.

Acknowledgements

I would like to thank everyone who helped spark my interest in mathematics as a youngster. In particular, I would like to thank Mum for demonstrating the existence of negative numbers to me by typing '$1 - 2$' on a calculator, Pete Gledhill for showing me how π can leap from nowhere, Carol Smith for helping me find the underlying structure when all I wanted to do was complete calculations as quickly as possible, and Alison Smith for always pushing me mathematically that little further than I was comfortable to go.

At Loughborough, there cannot be enough thanks to go to my supervisor, Lara Alcock. I'm sure lots of glowing things have been said about supervisors by Ph.D. candidates over the years, and so these should be taken as said: she always makes time to see her students, her advice is sensible, fair and always good-humoured, etc. What I would like to thank her for, above and beyond these platitudes, is for her (possibly unique) quality of always being able to produce the perfect book to get me out of a rut. Whether it be Becker for getting writers' block, Williams for when you know what you want to say but can't say it well, or O'Connell for when there aren't enough hours in a day, her choices have never failed. So many thanks indeed.

Special thanks also must go to Aruna Palipana and Keith Watling for not only assisting me when I got stuck with technology, but also for showing trust in me by granting responsibility and access to equipment which, had this trust backfired, would have taken them both many hours to rectify. Thanks also to James Hall, who helped me spot so many errors and inconsistencies.

Finally, I would like to thank the other Ph.D. students at Loughborough for the friendly, welcoming and creative atmosphere that they created when I first arrived, and continue to create when I visit. In particular I would like to mention Ria Symonds, who helped me settle when I first started, and Stephanie Thomas, who has been a rock for me as I've been writing up.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis is situated within the literature focusing on undergraduate mathematics education. Much of this literature explores the transition from secondary to tertiary study in mathematics, which is considered by many as particularly difficult for students (Holton, 2001; Smith, 2004; Tall, 1991a). Such difficulties have been at least partially attributed to the changes in teaching and learning style at university (Baker et al., 1973; Clark and Lovric, 2008, 2009; Copes, 1982), changes in the content of mathematics (Alcock and Simpson, 2002; Artigue, 1991; Gueudet, 2008), and changes some authors believe are needed in students' understanding and thinking (Crawford et al., 1994; Tall, 1991b; Tall and Vinner, 1981).

Of the research exploring mathematics learning at this level, much has has focused on material from Real Analysis (Artigue, 1991; Meehan, 2007; Tall, 1991b; Weber, 2008). This is both because the material is typically studied early on in a mathematics degree and so of interest to those studying the transition to university, but also because it is rich in complex formal definitions (Alcock and Simpson, 2002), and yet students often make judgements based on reasoning from their concept images (Tall and Vinner, 1981) and spontaneous conceptions based on the everyday meaning of words (Cornu, 1991).

This thesis contributes to the literature exploring students' understanding in Real Analysis, but it focuses on an area which has had relatively little study in the literature compared with the limiting behaviour and continuity of functions. The mathematical objects under consideration are real-valued sequences, that is, functions

$$f : \mathbb{N} \to \mathbb{R}, \quad f(n) = a_n$$

where (a_n) represents the image of the sequence. This thesis examines how students early on in their degrees understand and work with sequences and the definitions associated with basic sequence properties, such as *strictly increasing*. Such definitions are typically taught before those relating to the limiting behaviour and continuity of functions (e.g. Burn, 1992), and are comparatively simpler in terms of their quantifiers.

The main study reported in the thesis uses an interview-based task where students are asked to generate examples of sequences satisfying combinations of definitions, such as *a strictly increasing sequence which does not tend to infinity*. Using a phenomenographic methodology (Marton and Booth, 1997) in the analysis of the main study, the thesis not only explores students' understanding of sequences and the definitions associated with sequence properties, but also the same students' approaches and reasoning related to the example generation process.

The thesis presents its findings in terms of a set of dimensions of variation, with each dimension focusing on a different aspect of the example generation process: the different ways in which students use definitions when generating examples (Section 6.2); the ways students choose to represent their answers (Section 6.3); the strategies of example construction employed by students (Section 6.4); and the ways in which students justify the correctness of their answers (Section 6.5).

This introduction continues by stating the thesis' research questions, and then outlining the structure of the thesis on a chapter-by-chapter basis, with a final comment on the dependencies between the chapters.

1.1 Research questions

The main purpose of the thesis is to address two research questions:

1. How successful are students at accurately generating examples of sequences satisfying certain combinations of properties?

2. What is the qualitative variation in students' experiences of sequence generation?

The first question is concerned not only with the number of students who provide correct answers in example generation tasks, but also the type of answers they might give, and the ways these answers might vary. It also is concerned with the ways students go about process of example generation, and how successful those different ways may be. When answering this research question in the thesis's conclusion, reference will also be made to what the studies have suggested about students' concept images and concept definitions, and the types of misconceptions and spontaneous conceptions that are associated with the reasoning students have demonstrated.

The second question, phrased within the framework of phenomenography (Marton and Booth, 1997), is concerned with exploring the dimensions of variation which collectively describe students' experience of sequence generation, and the categories of description that comprise each dimension.

1.2 Structure of the thesis

This thesis has nine chapters, the first of which is this introduction chapter. The other chapters fall into four sections: reviewing the literature (Chapters 2 and 3), reports of empirical studies (Chapters 4, 5, and 6), two validation exercises (Chapters 7 and 8), and a thesis discussion and conclusion (Chapter 9).

Literature reviews

Chapter 2 situates the thesis within the mathematics education literature. It begins by exploring issues related to the progression to studying mathematics at the tertiary level, looking at the social, epistemological and cognitive issues, and introducing the theoretical constructs of concept image, concept definition and spontaneous conceptions. Next, it explores research on examples and example generation, defining the notion of an example space, and outlining research which has studied students' example spaces and the process of example generation. After discussing why the module Real Analysis has been well-studied by mathematics education researchers, it highlights that there has been considerably less research on simpler concepts in Analysis such as real sequences.

The chapter concludes that an example generation task is a good tool to use to study how students reason and interact with sequences and the definitions of sequence properties.

Chapter 3 outlines the phenomenographic methodology used in the main study of the thesis. Phenomenography is a research specialism which explores students' experience of a phenomenon, focusing on students' reports of their thinking and interpreting such reports at face-value as possible ways of experiencing the phenomenon. Assumptions of phenomenography are stated and discussed, and the specialism is briefly compared with other interpretive forms of research. Then, various education studies that have taken a phenomenographic approach are described in terms of their methods and presentation of their outcomes, including some studies that explore students' understanding in mathematics and science. Finally, the work of authors that have criticised phenomenographic methods and analyses are commented upon. The chapter concludes that phenomenography is a good research methodology to guide a study of students' experiences in the example generation process, both in terms of methods used, and in the way it presents results in terms of dimensions of possible variation experienced by students.

Empirical research

Two pilot studies are described and briefly analysed in Chapter 4. These pilot studies were conducted and analysed before the inclusion of a phenomenographic methodology; their findings are included in the thesis to give the reader a flavour of the range and type of responses students can give to sequence example generation tasks, and how how such responses helped shape the research questions. The first pilot study was an example generation task given to 101 undergraduate students, asking them to provide examples of sequences subject to certain combinations of constraints. It found that many students did not provide a sequence that satisfied the conditions of the questions, sometimes instead giving a mathematical object which was not a sequence. The second pilot study gave an example generation task to six students as part of a semi-structured interview. Results were constrained by the limited sample size confounded with some students feeling unable to attempt the questions, but the answers given by students did replicate the types of answer found in the first pilot study. Most significantly, the second pilot study suggests that some students are more likely to give answers that are not sequences, some questions are more likely to provoke answers that are not sequences, and when students generate objects which are not sequences, at least some believe the object given does represent a sequence.

The next two chapters outline the main study and present the main study's outcomes. Chapter 5 plans the main study. It takes the research specialism of phenomenography described in Chapter 3 and links it with the findings from the pilot studies in Chapter 4 and the research on example generation from the review of research in Chapter 2 to design an example generation task, and provide a framework in which to analyse the data from such a task. The aim of the main study is to answer the second research question and provide more information on the first research question by conducting example generation interviews and analyse them with techniques from phenomenography (so that dimensions of variation describing the qualitative variation in students' experiences emerge from the data). Chapter 6 first presents a brief discussion on the variation within the definitions and example generation questions from a researchers' perspective, followed by the outcomes of the data analysis: an 'outcome space' consisting of four dimensions of variation: Using Definitions, Representation of Sequences, Sequence Construction Strategies, and Justifications. Each of these dimensions of variation consists of a number of categories of description arranged hierarchically in terms of their relative sophistication, with each category of description separately and collectively addressing the second research question. The chapter concludes by focusing on the categories of description that are associated with the types of incorrect answers seen in the pilot studies.

Validation

The next two chapters of the thesis present two activities designed to explore the validity of the main study's outcomes.

Chapter 7 explores the communicative and pragmatic validity of the dimensions of variation described in Chapter 6; in other words the chapter considers how applicable the dimensions of variation are to new data, and if the dimensions of variation are seen as useful by other researchers. These questions are addressed by reporting on an inter-coder validity exercise in which two colleagues took the dimensions of variation and used them to independently code extracts of new interviews. The chapter is largely based on a discussion between the two researchers and the author based on how consistent and applicable the dimensions of variation were to new data. The chapter concludes that both researchers felt that the structure and content of the dimensions of variation was appropriate for analysing new data, internally consistent, and that the methods provided insight into the data.

Chapter 8 then explores the external validity of the main study by comparing the fifteen students who were interviewed as part of the main study with a wider population of 147 from their year group who were given the task in class seminars. The statistical technique of Rasch Analysis is argued to be ideal for this purpose, and after showing that the dataset satisfied the Rasch Model's assumptions a computer package was used to estimate the data's Rasch Model parameters. The questions are ranked objectively in order of difficulty, and it is noted that in general the more difficult questions combined different types of definitions. Then, the students who took the task during interviews were compared to the wider population. The chapter concludes that the students interviewed in the main study typically answered fewer questions correctly when compared to the entire cohort (including the interviewed students), but that the characteristics of the answers given were similar to those students in the wider population who answered the same number of questions correctly.

Conclusions

Finally, Chapter 9 returns to the two research questions and answers them using the findings from previous chapters. The chapter first considers the second research question by revisiting the main study's dimensions of variation. Then attention is turned to the first research question. It is answered quantitatively based on data from the four studies in the thesis, and then in terms of students' example spaces, concept images, spontaneous conceptions and the way students have approached the process of example generation. Where appropriate the dimensions of variation from the main study are used here also to provide further backing for the conclusions. After this the thesis's methodology is considered, and it is noted that whilst some authors have used the terminology of phenomenography when talking about example generation as a pedagogic tool (for instance Watson and Mason, 2005, p.5, describe students' example spaces in terms of dimensions of possible variation), this thesis has been the first to use the methods of phenomenography to analyse data from example generation tasks used as a research tool. Finally the chapter considers the thesis's pedagogical implications and new research that could further extend the thesis's outcomes within educational research both in Real Analysis and in other areas of mathematics.

1.2.1 Chapter dependencies

It is intended that the thesis is read following the numerical order of its chapters, but the reader may choose to read some chapters before their numerical ordering, as illustrated in Figure 1.1, and outlined below.

The general literature review on relevant areas of mathematics education presented in Chapter 2 should be read first, as it introduces the concepts, constructs and terminology used elsewhere in the thesis.

Chapters 3 and 4 may be read in numerical order if the reader wishes to separate the literature reviews from the empirical studies, although because Chapter 4 reports on pilot-study research that was not conducted from a phenomenographic perspective, it may be read before Chapter 3 presents this methodology. Moreover, reversing the order of these two chapters may give the reader a more solid foundation on which to follow how phenomenography can help address the research questions of the thesis.

The main study's planning and results (in Chapters 5 and 6, respectively) should be read before chapters Chapters 7 and 8 as both these latter chapters present validation exercises based on the main study's data and results. However, the order in which to read these validation exercises is arbitrary.

Finally the discussion and conclusion presented in Chapter 9 draws on content from all the chapters in the thesis.

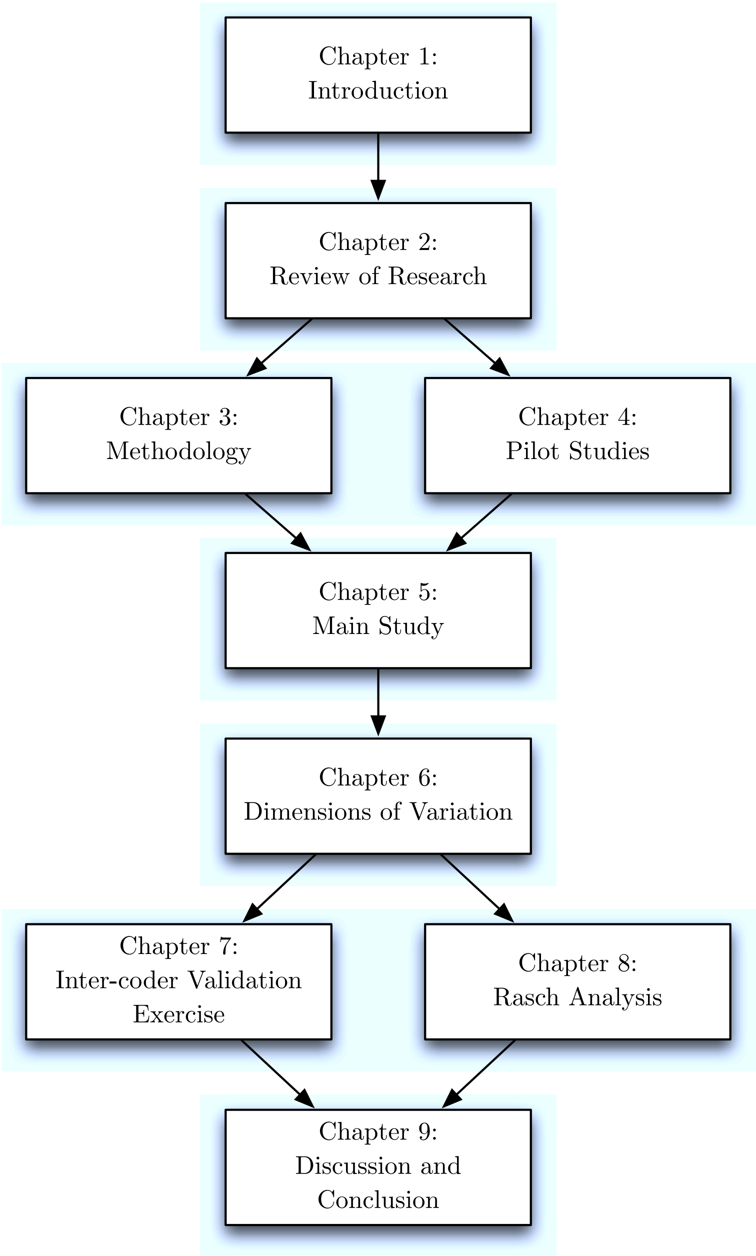


FIGURE 1.1: Chapter dependencies in the thesis. A → B means that it is intended that Chapter A be read before Chapter B.

Chapter 2

Review of Research

This review of research is split into three sections. The first, Section 2.1, gives a brief and general overview of research that deals with the problems students face when they begin the transition to tertiary level mathematics. It identifies general social issues, epistemological issues related to students' beliefs about learning, and cognitive issues related to the transition to advanced mathematical thinking. In subsection 2.1.3 two specific constructs are introduced: Tall and Vinner's (1981) *concept image/concept definition*, and Cornu's (1991) notion of *spontaneous conceptions*. Both of these are used frequently in the rest of the thesis.

The second section, Section 2.2, focuses on research on examples. It considers the fundamental role examples play in mathematics, and looks at how mathematicians and students use examples. It concludes that, for an expert, examples play an important role both when developing new theory and when studying existent theories, but that students often use examples inappropriately. The section also discusses research on example generation, describing how researchers have argued that example generation is not only a useful pedagogic strategy for extending students' examples, but also for exploring students conceptions of a mathematical concept or idea. In subsection 2.2.2 the construct of *example spaces* is introduced (a phrase used by various authors, but particular reference is made to Watson and Mason, 2005). Again, this idea is used frequently in the thesis.

Then, Section 2.3, focuses on research which explores how students learn and understand the undergraduate module Real Analysis. It revisits the themes of formal abstraction and definitions which were introduced in the first section of the chapter, but in the context of students' difficulties learning Analysis. It is identified that there has been relatively little research that explores students' conceptions of some of the simpler ideas in Analysis, and Section 2.4 concludes that it is this gap in the literature that this thesis will address.

2.1 The progression to tertiary level mathematics

This section describes the work of researchers who have identified and explored students' difficulties in the transition from secondary to tertiary level study in mathematics. It places the thesis within the wider context of studies exploring students' difficulties when starting at university, but is deliberately only a brief overview; subsequent sections cover in more detail research more directly applicable to the themes in the thesis.

2.1.1 Social and epistemological issues

The transition to tertiary level study in mathematics is widely documented as difficult for students (e.g. Clark and Lovric, 2008, 2009; Tall, 1991b). Various authors have framed these difficulties in terms of students' approaches to learning more generally. Studies have shown that secondary-level students typically believe mathematics to consist solely of problems that can be solved by applying facts, rules, formulæ and procedures taught by a teacher or presented in a textbook (Garofalo, 1989), and that mathematics problems should take a short time to solve, otherwise there is something wrong with the problem (Frank, 1988). A university student's approach to learning is expected to be more self-directed, their intellectual growth is considered to be non-linear and recursive and it is appropriate to balance different approaches and alternatives (Copes, 1982; Perry, 1988).

Authors have designed constructs which reflect these different approaches to learning. In research outside of mathematics education, Svensson (1977) gave first year undergraduate students of education a passage of text to study, then asked them to summarise the passage and describe what they did when they were studying it. Some students reported that they assembled facts from the passage, without considering links between these facts (an atomistic approach), while other students searched for the whole meaning of the text, considering the author's intention (a holistic approach). In a similar study, Marton and Säljö (1976) asked students to recount how they approached a reading task, finding that there was a distinction within students' descriptions; some reported they did not try to understand the text, only memorising it (a surface approach), whilst others tried to understand the message of the text (a deep approach). As noted by Marton and Säljö (2005), it was often the case that the students who self-reported a surface approach, trying hard to remember the text's content, frequently failed to do so, whilst those that concentrated on the text's meaning tended to remember the content very well. Other authors have added a third, 'strategic' approach to learning (e.g. Ramsden, 1984), but this is not so much a learning style as the targeted application of a knowingly surface approach which chooses which algorithmic approach is most appropriate to solve a particular problem.

Within the context of learning mathematics, rather than approaches to learning in general, Skemp (1976) defined two types of understanding (instrumental and relational). These could be considered as types of understanding that result from the different approaches to learning. Instrumental understanding is described by Skemp as "rules without reason", which has a parallel to the atomistic or surface approach to learning. Relational understanding is "both knowing what to do and why", which is the type of learning likely to result from a holistic or deep approach to learning.

Making a more explicit link between mathematics learning and Marton and Säljö's (1976) construct, Crawford et al. (1994) explored university mathematics students conceptions' of learning and approaches to learning. Students were asked to complete specially designed *Approaches to Learning* and *Conceptions of Mathematics* questionnaires, and their answers were phenomenographically[1] analysed. When analysing students' approaches to learning, Crawford et al. used Marton's deep-surface distinction, and found that 82% of students answered in ways which were classified as surface approaches to learning mathematics. In terms of students' conceptions of mathematics, 77% gave responses that were classified as fragmented (i.e. focused on the parts rather than wholes). The relationship between students identified as having a surface approach to learning and those identified with fragmented conceptions of mathematics was significant, $\chi^2(1, N = 236) = 126$, $p < .001$.

[1] See Chapter 3 for a discussion on phenomenographic methods, and more detailed account of Crawford et al.'s (1994) research.

In a later study, Crawford et al. (1998) gave the same questionnaires to the students later on in their courses. They found that there were the expected correlations between prior and current approaches to learning; for instance prior surface to post surface approach ($r = 0.62$, $p < 0.01$). They also found that students identified with a surface approach to learning were more likely to find their workload inappropriate and teaching quality poor compared to those with a deep approach, which echoes the results of Bessant (1995) who found that students with a deep approach to mathematics were less likely to be anxious about their courses. Students identified by Crawford et al. as having a deep learning style also "achieved at a higher level" in their final exams, although interestingly there was little difference in the groups' views as to the appropriateness of the course's assessment method.

Perhaps as a result of (or anticipating) such studies, there has been a drive to help students have a deeper approach to learning mathematics, and to understand mathematics in a relational fashion. The publication of *Curriculum and Evaluation Standards for School Mathematics* by the National Council of Teachers of Mathematics (1989) which triggered the US 'math wars' as described by Schoenfeld (2004), could be regarded as an attempt to introduce younger students to a relational understanding of mathematics. At the university-level, more recent attempts have included bridging courses (Alcock and Simpson, 2001; Wood, 2001), innovative uses of technology such as electronic voting systems (Draper, 2009), plotting/zooming software (Chae and Tall, 2001; Tall, 2003) and research on the benefits of peer instruction (Crouch and Mazur, 2001).

While theories such as the 'deep-surface' styles of understanding focus on students' attitudes to learning, and others such as the 'relational-instrumental' focus on types of understanding, other researchers have concentrated more on how the content of mathematics changes at university. The next subsection looks at research which considers why some of the more complex mathematical content met in tertiary education is so different to what students have met before, giving one suggestion as to why a procedural 'surface' approach may result in difficulties in the secondary-tertiary transition.

2.1.2 The changing content of mathematics

Authors exploring the changing content of mathematics at university often hypothesise that changes are necessary in the way students think about mathematics itself (c.f. Crawford et al.'s (1994) fragmented/coherent conceptions of mathematics). As Tall describes in an introductory chapter to a book focused on undergraduate mathematics education:

> The move from elementary to advanced mathematical thinking involves a significant transition: that from describing to defining, from convincing to proving in a logical manner based on those definitions. This transition requires a cognitive reconstruction which is seen during the university students' initial struggle with formal abstractions as they tackle the first year of university. (Tall, 1991b, p.20)

Such authors often label this new way of thinking Advanced Mathematical Thinking (AMT). As described by Tall in the above quote, AMT includes skills needed to be successful in tertiary mathematics such as dealing successfully with formal definitions, as well as constructing and understanding formal mathematical proof. In a literature review on the secondary-tertiary transition, Gueudet (2008) reported that many scholars view the transition to formal abstraction and proof to be fundamental conceptual barriers when studying mathematics at university.

Other authors have argued that advanced thinking in mathematics can occur at any age (for instance Harel and Sowder, 2005), which is more in keeping with Skemp's (1976) distinction between relational and instrumental understanding in mathematics, and Crawford et al.'s (1994) fragmented/coherent conceptions of mathematics. In this thesis AMT is considered more in the context used by Tall; it is the types of thinking that are associated with undergraduate study and beyond.

Some authors have argued that because of this change in the content of mathematics at the tertiary level, there should be an effort to change the way students interact with mathematics, especially in modules with a high proportion of formal content, such as mathematical Analysis (Artigue, 1991). In particular, understanding the role that definitions play within tertiary mathematics is particularly difficult for students. For now, a thorough discussion of students' interactions with definitions is postponed until mathematics education research on (real) Analysis is discussed in more detail in Section 2.3. This is because it is helpful to first introduce constructs such as concept image, concept definition, and example spaces before a discussion on definitions in more detail. For now it is noted that many students struggle with their Analysis modules, in part because the formal definitions student meet in Analysis are particularly complex relative to those in other modules (Alcock and Simpson, 2002).

2.1.3 Concept image, concept definition, and spontaneous conceptions

Tall and Vinner (1981) described a construct which distinguishes between reasoning based on students' prior experiences with mathematics and reasoning which is based on definitions within mathematics. The first type of reasoning uses a student's *concept image*.

> [Concept image] describes the total cognitive structure that is associated with the concept, which includes all the mental pictures and associated properties and processes. It is built up over the years through experience of all kinds, changing as the individual meets new stimuli and matures. (Tall and Vinner, 1981, p.152)

A concept image is therefore a mental collection of mathematical and non-mathematical objects and associations, which may be vast. A concept image may not be consistent within itself, it may change over time and, depending on the situation, the student may be aware of different parts of a concept image when faced with different situations (Tall and Vinner, 1981; Vinner, 1991). In this thesis we will see situations where a students's *evoked concept image* (the portion of a concept image which is accessed at a particular time) changes during the course of a short period of time (i.e. during a twenty-minute task).

The second type of reasoning is said to be based on a student's *concept definition*:

> [Concept definition] is a form of words used to specify that concept. It may be learnt by an individual in a rote fashion or more meaningfully learnt and

> related to a greater or lesser degree to the concept as a whole. It may also be a personal reconstruction of a definition. (Tall and Vinner, 1981, p.152)

A student's personal concept definition may be different from the formal mathematical definition, but one would expect that for a more experienced mathematician the two are more similar. In this thesis, 'concept definition' without clarification refers to an individual's personal concept definition; when reference is made to the formal mathematical definition the phrasing 'formal definition' is used (this follows the style used in papers such as Bingolbali and Monaghan, 2008; Tall, 1988).

Ideally, a student's reasoning in mathematics would be based on both their concept image and a concept definition which is similar (or identical) to a formal concept definition. However some students may reason based on their concept image solely (Vinner, 1983). This can be a problem because much of tertiary mathematics involves successfully dealing with formal definitions. Moreover, within a concept image there can often be mathematical and non-mathematical notions and representations which are elaborated from *spontaneous conceptions* (Cornu, 1991).

A spontaneous conception often exists before a mathematical concept is learnt and can be part of a students' evoked concept image. For instance, Schwarzenberger and Tall (1978) note that students will have met the phrases 'tends to' and 'limit' before dealing with formal definitions in tertiary mathematics, and some of the everyday meanings of the words may interfere with their mathematical counterparts (similarly noted by Monaghan, 1991). To continue with the example of a limit, possible spontaneous conceptions may arise from familiarity with a road speed limit which should not be exceeded. However, the limit of a function can be exceeded infinitely many times, such as:

$$\lim_{x \to \infty} \frac{1}{x} \sin(x) = 0$$

Issues of spontaneous conceptions are of course not restricted to the English language. Recently Spyrou and Zagorianakos (2009, 2010) argued that because the Greek word for function (*synartisi*) is often used in everyday language when a relationship is symmetrical or proportional (in the same way that distance a car can travel and the amount of petrol are related), undergraduates often treated the order of variables in an equation as irrelevant; when asked to give examples of functions all but one student wrote a one-to-one function, and many had difficulties when asked to give an explanation of how a function could be many-to-one.

The impact of everyday terms in a logical statement was also discussed by Mason and Pimm (1984), in particular the use of the word 'any' and its relationship with the universal quantifier '$\forall$'. They remarked that

> Mathematicians tend to use 'any' to mean 'every', and occasionally their meaning conflicts with ordinary usage. For example [an assignment] reads 'For any matrix A in W, show that $A^2 = A$.' Six out of 13 submitted answers chose a particular matrix for A and derived the result. When asked about this one student commented 'Well, it said show it for any, so I just picked one.' (Mason and Pimm, 1984, p.281)

Concept image, concept definitions and spontaneous conceptions are constructs which attempt to describe and explain why students struggle with the content of mathematics (Alcock and Simpson, 2009a). For instance, various authors have argued that students do not correctly recognise certain types of tangent because their concept images do not contain such examples. Tall (1986) noted that few students could identify tangents drawn that touched the function more than once, and more recently Biza et al. (2008) noted that few trainee teachers had seen tangents which 'cut' the function in two (e.g. at the inflection point of a cubic). It is unlikely that the students in Tall's or Biza et al.'s study were reasoning with the formal concept definition of a tangent, rather than their concept images which (one can assume) lacked appropriate examples.

As noted by Bingolbali and Monaghan (2008, p.21), concept image and concept definition have been more frequently used in studies looking at tertiary level mathematics learning, and this is possibly because—as discussed in the last subsection—the content of tertiary mathematics is more dependent on understanding formal definitions and the role definitions play within mathematics (a more detailed discussion of definitions within mathematics is presented in Section 2.3.2).

2.2 The use of examples in mathematics

This section shifts the focus from looking at research on the general difficulties students face when studying tertiary mathematics, to exploring research on how experts and students use and reason with examples. First, subsection 2.2.1 outlines what 'an example' means within the context of this thesis. Then, subsection 2.2.2 introduces the terminology of specific, generic, general and particular examples, together with the construct of an example space.

After these preliminary discussions, the attention of the chapter turns to the literature on how experts and students of mathematics use examples. Subsection 2.2.3 explores research which focuses on how mathematicians use, reason with, and understand examples, and then subsection 2.2.4 contrasts the previous subsection by focusing on students' interactions with examples. As well as reasoning with their concept images and not their concept definitions as identified in the last section, some students also base much of their thinking on empirical arguments involving generalisation from specific examples. The section then turns to look at which examples students have access to; in subsection 2.2.5 research is introduced that explores the structure of example spaces, and subsection 2.2.6 the process of example generation. In conclusion, this section identifies that we know little about how students go about the process of example generation, and it is this gap in the literature which this thesis addresses.

2.2.1 The scope of examples considered in this section

In the thesis, when the word *example* is used, it refers to a mathematical object that satisfies certain criteria. Often the criteria will be that of satisfying a formal mathematical definition. This means that 'worked out examples', or 'model solutions' are not considered as 'examples' within this context. Of course, examples of those types have been studied by other authors (for instance Chi et al., 1989; Sweller and Cooper, 1985).

2.2.2 Terminology

This subsection first introduces terminology which describes for what purpose an example is used, whether it be to represent a wider class of objects, or just as a one-off instance of a phenomenon. Then, the construct of an *example space* is introduced.

Specific, Generic, General, and Particular examples

A recurring theme that will be identified in this section is that when individuals interact with examples, some may regard the examples as representatives for a wider class of objects, whilst others will see the examples as specific instances only. Mason and Pimm (1984) illustrated this with the following exposition from Analysis:

> The function
>
> $$x \to |x|$$
>
> is often the only example of a continuous but non-differentiable function presented [to students]. What is happening is that the lecturer, in presenting the example, is seeing it as generic. It indicates a whole class of functions
>
> $$x \to k|x| + C$$
>
> at the very least. The students however are concentrating on the particular example. They see, not a class of functions, but a single function. (Mason and Pimm, 1984, p.285)

It is necessary therefore to formulate a terminology which accounts for how an individual intends to use an example. In the same paper, Mason and Pimm (1984) presented the terminology of 'specific', 'generic', 'general' and 'particular' examples, distinctions which shall also be used in this thesis. In this explanation of their terminology I follow their use of 'the number 6 as an even number' as illustrating each example type:

Specific A one-off situation that may or may not be general

THE even number 6

The existence of such an object is the important point rather than necessarily the representation of a wider collection of objects. In this sense counterexamples to theorems are specific.

Generic Using an example to represent a class of examples with a similar property

AN even number such as 6

Here, the example is used to represent other objects, but there is no intention to represent a complete class of objects.

General Using an example to represent an operation on a wider class

ANY even number like 6

Here the extent of the class to that the example refers to is known, or implied.

Particular Using a general example in a specific situation or argument

$2N$ *is even,* $2N + 2N = 4N$ *so* $4N$ *is also even*

In this example, each $2N$ implicitly refers to the same number, so although N in isolation is a general example, when used in this context $2N$ is a particular example.

(*italics* taken directly from Mason and Pimm, 1984, p.281, 283)

The distinction between Specific and Particular examples is subtle. As an additional illustration, consider the stages of a proof by mathematical induction:

- The aim is to prove the statement $P(n)\ \forall n \in \mathbb{N}$, so in this line n is *general*
- First we prove $P(n_0)$ for a *specific* base case, n_0
- Assume the truth $P(k)$, for a *particular* value of k
- Prove $P(k) \implies P(k+1)$, so although the choice of k in the last step was arbitrary, now it is fixed in calculations, (because it is particular) unlike the general n

When an example is presented, different audiences may perceive the example to be of different types, as illustrated by the quoted passage from Mason and Pimm (1984) on the previous page. In the terminology just introduced, the lecturer is treating the example as generic, while the students are interpreting it as specific, whereas perhaps an audience of academics would also treat the example as generic. This means that exploring how students use and interact with examples can be difficult for a researcher; it will not always be clear how such students are interacting with examples. Such intersubjectivity concerns are more thoroughly discussed in the methods chapter of the main study, Section 5.5.

Example spaces

Various authors have described a construct which attempts to encompass the examples to which an individual has access, either generally or at a specific time. Michener (1978, p.364) used the term *examples space* to describe the set of examples an individual may use when considering a specific mathematical concept and result in mathematics. More recently, authors such as Watson and Mason (2005) have defined an individual's *example space* in a way that is more analogous to Tall and Vinner's (1981) construct of a concept image:

> Think of an example space as a toolshed containing a variety of tools — examples that can be used to illustrate or describe or as raw material. Some tools are familiar and come to hand whenever the shed is opened, whereas others are more specialised and come to hand only when specifically sought. (Watson and Mason, 2005, p.61)

In Section 2.1.3, an individual's concept image was described as the total cognitive structure that is associated with a particular concept. Similarly an example space can be described as the set of examples (and classes of examples) that an individual has access to (with the previously stated proviso that 'worked examples' etc are not included in the context of this thesis). An example space can therefore be considered a subset of a concept image. This may amount to a vast collection of examples which may be in a variety of representations (graphs, formulae, lists, and more vague objects), but at any specific moment only a limited number may be accessed. Goldenberg and Mason (2008, p.187) called this accessible set the *accessible example space*, and Watson and Mason (2005, p.76) call a similar construct a *situated (local) personal (individual) example space*. However, to keep the analogy with an individual's (evoked) concept image I shall follow Zazkis and Leikin's (2007) use of *(evoked) example space* in this thesis.

Within an (evoked) example space, Watson and Mason (2005, p.51) and Bills et al. (2006) argue that there will always be a degree of structure.

> Example spaces are not just lists, but have internal idiosyncratic structure in terms of how the members and classes in the space are interrelated. (Bills et al., 2006, p.133)

This structure is determined implicitly by the individual, rather than imposed by an external agent. In particular, the structure may include the suitability of an example for different purposes, for instance whether typically it is used in a 'specific', 'generic', 'general', or 'particular' sense. Empirical studies, based on asking students to generate examples of concepts, or identify whether objects are examples of certain concepts, suggest that the structure of an individual's example spaces reflects that individual's experience and memory, and that the structure of the evoked example space may change based on the circumstances of the prompt (wording, who is asking), or even that it changes in the same circumstances (see Section 2.2.5).

2.2.3 Experts' use of examples in mathematics

This subsection explores research which looks at how mathematicians use and understand examples. It can be considered that mathematicians look for general relationships within mathematics, describing them in definitions and theorems. Despite this emphasis on the general, many mathematicians consider it important to work with specific and particular examples of a concept, as was remarked by Bills et al. (2006) in a paper looking at research into examples within mathematics education:

> Many would agree that the use of examples is an integral part of the discipline of mathematics and not just an aid for teaching and learning. (Bills et al., 2006, p.126)

Michener (1978) also argued that examples are fundamental in properly understanding mathematics when she attempted to build a conceptual framework of the "ingredients and processes involved in the understanding of mathematics," (Michener, 1978, p.361). She argued that to understand mathematics, one needed to master three major categories of mathematical item: results (theorems and arguments), examples (illustrative material) and concepts (definitions and more informal notions). Michener (p.382) went on to suggest that one can not fully understand results or concepts without good control over the choice and use of examples that are associated with them.

There has understandably been much research which looks at how experienced mathematicians interact with examples, and the rest of this section focuses on such research.

A wide variety of authors suggest that many expert mathematicians (i.e. doctoral level or higher who publish research in their field) use examples when researching new topics (Zazkis and Chernoff, 2008, p.196). Although not strictly a mathematician, Richard Feynman remarked that, when faced with a new equation:

> I can't understand anything in general unless I'm carrying along in my mind a specific example and watching it go. (Feynman et al., 1997, p.244)

For Feynman at least, looking at a situation with a specific example helped him gain a more general understanding of the situation. When asking mathematicians how they use examples, Alcock (2004) found that there were three ways in which experts self-reported using carefully-selected specific examples in the proving process:

> 1. Instantiating examples in order to understand a statement or definition.
> 2. Generating an argument for a universal statement, by (directly) arguing about or manipulating a specific example and translating this to a general case or (indirectly) trying to construct a counterexample and attending to why this is impossible.
> 3. Considering possible counterexamples to general claims in a proof.
>
> (Alcock, 2004, p.2-21)

In a more recent study, Alcock and Inglis (2008) reported on a case study into the use of examples by two doctoral students. They found that when faced with a new topic, one student used examples only when prompted to do so, but the other made substantial use of examples throughout. When interviewing research mathematicians, Weber and Mejia-Ramos (in press) asked if they "used examples to increase their confidence that a proof is correct". Each confirmed they did, and some claimed that they never read a proof without considering examples. Morrow (2004) gave mathematics lecturers a collection of explanations why a theorem was true and asked them to say which was them was the most convincing. A graphical example was chosen by one because:

> [From Professor Jones:] It gives you that "aah" insight that you don't get when you work through the definition. In this one [a definition based answer],

> you do get the "aah" insight at some point, but it's a little more painful in coming, a little slower in coming.
> (Morrow, 2004, p.121)

Some mathematicians have a favourite set of example to use in a variety of situations, which echoes the quote on page 20 about how, within an individual's example space, some examples come to hand more readily than others. Michener (1978, p.366) described this type of example as a *reference example*, standard cases that are used to check out understanding. In the context of discussing the development of example spaces, Watson and Mason (2005, p.75) describe how the mathematician Peter Nyikos has a repertoire of usual counterexamples to try with certain types of questions that are peculiar in some way or another, for example totally ordered sets with two or three elements. Indeed, entire books have been written on the subject of counterexamples (Gelbaum and Olmsted, 2003; Mason and Klymchuk, 2009).

The research cited here suggests that, for a mathematician studying a topic which is unfamiliar, using examples might aid understanding of that topic. However, choosing which examples to use brings with it the concern of ensuring that the example is appropriate to that definition or theorem, especially if the definition or theorem is well established in mathematics. To some extent this concern can be reduced by using reference examples, but such examples may not be sufficiently general to exemplify the entire definition or theorem. Compared with an undergraduate student of mathematics, it is natural that an expert will have far greater experience when selecting, generating, and working with examples that are appropriate to the subject matter (Moore, 1994, p.260). Such experts will be more aware of need to control for undesirable issues that can result when a specific example is treated as general or generic. For instance an example may bring with it further unwanted properties which are not generalisable to the entire class of objects which it represents, or an example may behave in a way which is uncharacteristic of the class as a whole (Mason and Pimm, 1984). This section now leaves expert practice, and explores how students use examples, before looking at literature which explores example spaces more abstractly.

2.2.4 Students' use of examples in mathematics

The previous subsection argued that experts use examples when trying to understand new topics, and so understandably it is considered beneficial by many for students to be presented with examples when learning a new topic (Bills et al., 2006; Meehan, 2007; Selden and Selden, 1998). Indeed, as noted by Chick (2009, p.30), for many educators "there is an underlying assumption that examples facilitate learning," and in terms of pedagogy, examples are important for providing a more concrete grounding when abstract ideas are introduced (Alcock, 2004; Dahlberg and Housman, 1997; Weber et al., 2008). Meehan (2002, p.77) notes that students who can access more (correct) examples of a concept will have better informed concept images, and therefore possess concept images more in line with formal theory.

It is not clear, however, that undergraduate students of mathematics will interact with examples in the same way that experts do. This subsection first introduces the notion of a student's *proof scheme*, and then presents research which suggests that many students find empirical arguments convincing, and that students sometimes try to prove statement by checking with examples rather than by more deductive arguments.

Proof schemes

Harel and Sowder (1998) explored the different mathematical arguments undergraduate students found convincing, which they grouped into general types called students' *proof schemes*.

> A person's proof scheme consists of what constitutes ascertaining and persuading for that person. (Harel and Sowder, 1998, p.244)

Seven proof schemes were presented in the paper, which were grouped into three categories:

- External conviction proof schemes, which includes arguments which are reliant on formal symbolic manipulation (symbolic proof scheme), appeals to an external authority such as a textbook (authoritarian proof scheme), and where conviction is gained from the form of the proof (ritual proof scheme). These proof schemes tend to require no intrinsic justification.

- Empirical proof schemes, which are reliant on physical facts and sensory experiences not backed by deduction (procedural proof scheme), or generalisation from specific examples (inductive proof scheme).
- Analytical proof schemes, which may be based on deductive transformation of verbal or written statements (transformation proof scheme), or deductions based on manipulation of axiomatic statements (axiomatic proof scheme).

This subsection is concerned with students' use of examples, and so most of the focus will be on empirical proof schemes. Weber (2010) noted that there are three types of empirical evidence that may convince a student that the universal statement '$\forall x \in X$, $P(x)$' is true. The first is to check one or several specific examples for x, the second is to arrive at a reason why $P(x)$ is true based on trying specific examples, and the third is to try and generate a generic example with which to check the statement. These types of arguments (which I would agree should be classified as empirical proof schemes) would not be accepted as proofs by the mathematical community, but clearly they play a large part in mathematical reasoning. Harel and Sowder (1998, p.276) noted that arguments similar to these are common amongst undergraduate students, perhaps because "natural, everyday thinking utilises examples so much."

Research on students' tendency towards empirical proof schemes

There has been much research, both before and after Harel and Sowder's (1998) paper, which explores the types of arguments students find convincing, and the arguments students accept as valid proofs. Research with secondary school students typically suggests that they find empirical arguments both to be convincing, and to prove the statement. Chazan (1993) gave two geometry proofs to nine high school students during research interviews. The first was an empirical proof which involved measuring specific examples, and the second was a deductive 'two-column' proof. The empirical measurements were found to be more convincing by three of the students for at least part of the interviews, while two students did not find the deductive proof convincing because of the possibility of counterexamples (a possibility that these two students also identified for the empirical proof). Stylianides and Al-Murani (2010) also explored whether secondary school students thought a proof and a counterexample could coexist, finding evidence of such views when students were surveyed, but less so when they were asked to explain their views during an interview.

This research suggests that when secondary school students view an argument as a proof, they still prefer to use reasoning consistent with empirical proof schemes to gain conviction. Exploring the relationship between what a student found convincing, and what a student viewed as a proof, Fischbein (1982) gave students in grades 10–12 the following statement:

> Dan claims that the expression $E = n^3 - n$ is divisible by [6 for] every n and he gives the following proof:
>
> $n^3 - n = n(n^2 - 1) = n(n + 1)(n - 1) = (n - 1)n(n + 1)$
>
> i.e. $n^3 - n = (n - 1)n(n + 1)$
>
> Thus we have obtained the product of three consecutive numbers ($n - 1$, n, $n + 1$). Among three consecutive numbers there is *always*, at least, one number divisible by two and there is *always* one number divisible by three. Therefore the product is divisible by 2 and by 3. The product of three consecutive numbers will then be divisible by $2.3 = 6$.
>
> (Fischbein, 1982, p.16)

The students were asked if they agreed with Dan's proof (82% did), if they thought the proof was fully correct (69% agreed), and if the proof was indeed general for any n (60% thought it was). Of the 85% of students who thought Dan's proof was correct, over half (57%) agreed that in order to increase their confidence in the theorem, further checks were necessary, such as trying numbers to get "more precise" results.

The above studies indicate that secondary school students typically rely on empirical proof schemes both when they claim to be convinced by a deductive proof, and also when they are more skeptical about them. It is understandable therefore that various authors have identified that students often solely use examples to validate conjectures, so-called 'proof by examples' (Harel and Sowder, 1998; Nardi and Iannone, 2006, p.18). This is in contrast to more experienced mathematicians who tend to reason with examples in conjunction with an analytical proof scheme based on formal definitions (Alcock, 2004).

Studies which consider undergraduate students' proof schemes have found more mixed results. In a series of six interviews with first year undergraduates, Morrow (2004) presented the students with the statement:

> The derivative of every even function is an odd function.
> (Morrow, 2004, p.108)

Students were asked first to reflect on the statement, and then to consider five arguments presented to them: a collection of examples, a graphical argument, a formal deductive proof based on definitions, a proof based on the chain rule, and a flawed algebraic manipulation. As well as indicating whether the arguments were a proof of the statement, they were asked to gauge how convincing they found each argument. Two of the six students did not consider a deductive proof to justify the truth of a statement unless they also used an example as a subsequent check, and one student regarded proof as the same as working with a well-chosen generic example.

Rather than presenting students with arguments and asking if these arguments are proofs, Recio and Godino (2001, p.91) asked 429 first year undergraduates to themselves prove statements from geometry and algebra. A high percentage of the answers were empirical, and so arguably this finding is consistent with the students finding convincing arguments that are consistent with empirical proof schemes. However, Vinner (1997) and Weber (2010) noted that when students are asked to provide proofs there are various reasons why they may provide empirical answers without believing in empirical proof, for instance as not being able to produce a deductive proof and wishing to obtain partial credit for their answer.

Other authors have found empirical proof schemes to be rarer amongst undergraduate students. Weber (2010) gave undergraduate students a collection of mathematical statements and arguments and asked them to (a) rate their understanding of the arguments on a five-point scale, (b) rate how convinced they were on the same scale, and (c) decide if the argument was a proof. 96% of students did not find empirical arguments convincing (i.e. selected 1-4 on the scale rather than 5: "I feel completely convinced"), and 93%

did not view such arguments as proof. Weber (p.328) conjectures a variety of reasons as to why these results are different to that which has previously been shown, one of which was the close proximity of the task to the completion of a transition-to-proof course.

In order to explore students' proof schemes, some of the studies in this subsection's exposition asked students to provide their own proofs, whilst others gave students proofs to read and comment upon. In all but Weber's (2010) study, examples and empirical arguments were seen as convincing (and often preferable over deductive proof) by students. These results are to be taken in the context of other research indicating that students' approaches to proof may change even during the course of one interview (Housman and Porter, 2003), but if students are convinced and prefer working with examples, it is important to study in more detail how they work with examples.

Do students spontaneously choose to work with examples?

If reasoning based on careful use of examples is fundamental to the understanding of mathematics (Bills et al., 2006; Michener, 1978; Watson and Mason, 2005) and students find reasoning based on arguments from empirical proof schemes convincing, it may be the case that teaching based on examples may reinforce or otherwise encourage reliance on (often inappropriate) empirical proof schemes.

As discussed in the last section in respect to expert practice, reasoning based partially on carefully-selected examples is not necessarily undesirable. It is not clear, however, that students do spontaneously choose the same types of examples that experts do, and whether they reason with these examples in a similar way to experts. In particular, it is not clear whether students generate examples of concepts if such examples are not given to them. Dahlberg and Housman (1997) asked eleven undergraduates (one reading computer science, the other ten mathematics) to explore a class of function they had not met before:

> A function is called *fine* if it has a root (zero) at each integer
> (Dahlberg and Housman, 1997, p.287)

Each student, who had taken courses in Analysis, Linear Algebra and Set Theory, was asked to determine if functions presented to them were fine, and then to consider more general conjectures such as:

> No polynomial is a fine function.
> (Dahlberg and Housman, 1997, p.288)

Students were observed to follow one of four strategies: example generation (such as sketching generic examples), reformulation (such as rewording the definition more formally), decomposition and synthesis (breaking down the definition informally), and memorisation (which involved no manipulation other than memory of the definition). Students who spontaneously generated examples when faced with the definition were more successful in understanding that topic in the interview situation. In contrast, Weber (2009) presented a case study of a successful student who claimed in a post-task interview that when faced with new concepts he would "plug in examples to get an idea of what it means", but did not do this when given a task similar to the one used by Dahlberg and Housman; the student instead preferred to work in the abstract.

The research presented in the thesis involves asking students to generate examples of sequences, based on definitions that are provided to them. Dahlberg and Housman's (1997) research would indicate that students who are successful at doing this have a better understanding of the topic in question, but in order to be successful in example generation a students' example space must contain such examples, or their concept image the material to construct (and identify) such examples. Research is needed therefore on the content and structure of students' example spaces.

2.2.5 Attempts to describe the structure of example spaces

The last two subsections concluded that it is of interest to explore students' example spaces, and how they go about reasoning with examples from such spaces. This exploration is useful because both expert mathematicians and students frequently make use of examples, and many students use (or overuse) examples within an empirical proof scheme. If one considers that it is a goal of mathematics education research to describe how students think mathematically, then an exploration of students' example spaces is necessary.

Authors cited in this subsection typically have used *example generation* as a tool to explore students' example spaces. By example generation, I mean the practice of asking individuals to give examples of a mathematical object subject to certain constraints, for instance 'a function $f : \mathbb{R} \rightarrow \mathbb{R}$ which is continuous everywhere.' The next subsection looks at studies which discuss the process of example generation in more detail; in this subsection we focus more on studies exploring the content and structure of example spaces rather than the process of 'evoking' or 'searching' within them.

Studies in which authors have gone about describing the structure of example spaces can roughly be separated into two approaches depending on the unit of analysis. One approach focuses on a specific individual and a specific concept, and attempts to map a portion of that individual's (evoked) example space. Another approach is to focus on the example space as the unit of analysis, attempting to describe common structure in a variety of individuals' example spaces. With respect to Watson and Mason's terminology (2005, p.76), the first type are exploring *personal potential* example spaces, whilst the latter are exploring *collective and situated* example spaces.

The work of this thesis sits amongst the latter approach; one of its aims is to explore students' example spaces of real sequences (this is related to the first research question in section 1.1). In this subsection both approaches to exploring example spaces are discussed in turn.

Exploring an individual's example space

The authors of studies which focus on an individual's (evoked) example space are interested in exploring which examples an individual can access, and which examples are evoked under different conditions. In a paper on using example generation as a research tool, Zazkis and Leikin (2007) introduced an assumption that is common to most studies:

> Our working assumption is that example generating tasks may serve as a research tool in studies that aim to describe and analyze participants' knowledge. (Zazkis and Leikin, 2007, p.19)

In the paper, Zazkis and Leikin (2007) presented a case study of two trainee elementary teachers that identified misconceptions in the teachers' mathematical knowledge by seeing how the teachers dealt with example generation tasks. One such task was to

> Think of a 5-digit number that leaves a remainder of 1, when divided by 2.

They described an account of a teacher who deduced that such a number must be odd (in the next section focusing on example generation such deductions are said to be part of an *analysis* strategy), but despite this the teacher did not feel comfortable that her reasoning was correct when faced with examples of odd numbers; for instance the teacher preferred to check her example (10,003) was correct by dividing her answer by 2.

Associated with research that explores an individual's example space is that which attempts to use example generation to introduce new concepts. Watson and Shipman (2008) encouraged secondary-school students to explore which numbers of the form $(a + \sqrt{b})(c + \sqrt{d})$ would give rational answers when multiplied out. They found that this inquiry-based approach was partially successful, some students had began to discover identities such as $\sqrt{a}\sqrt{b} = \sqrt{ab}$, but many students avoided special cases, instead choosing only square numbers for b and d, and never considering numbers in the more generic form $(a - \sqrt{b})$.

Zazkis and Leikin (2008, p.145) argued that the "examples generated by students mirror their understanding of particular mathematical concepts". In this study, they asked trainee secondary school teachers to generate definitions of a square, then used these definitions for the basis of discussion with the participants. This view—that insights into the structure of an individual's example space will also mirror their conceptions of the concept more generally—provides one link between the evoked example space and the evoked concept image.

Exploring example spaces more generally

The other approach to describing the structure of example spaces is focused at the space itself, rather than at specific individuals. Tsamir et al. (2008) examined the structure of the example space of five- and six year old children by asking them if various geometrical figures were (examples of) triangles or not. By doing this they identified that the children had an implicit notion of 'triangleness'. This research is typical; it does not attempt a general description of an individual's example space (or the structure of this space) in the form of a case study, it instead explores many individuals' evoked example spaces and draws conclusions about the general nature of such example spaces.

When example generation is used as a way to explore topics which are already known to the individuals, researchers typically place constraints on the type of example that is required. Zaslavsky and Peled (1996) asked teachers and student teachers of mathematics to give an example of a binary operation that was commutative but not associative. Their findings were focused on two areas; they first provided evidence of a weak evoked example spaces (the authors used the term *content search-space*, but the two terms are essentially the same), and the second was to highlight the participants' primitive concept images of binary operations, which were typically overgeneralised and the order of operations overlooked. This study used example generation to explore both the participants' evoked example spaces and to draw hypotheses about their concept images.

In a paper discussing general features of example spaces, Goldenberg and Mason (2008) described some of the different ways they may be structured. For instance, it is possible that an example of a mathematical object has associated to it a notion of representativeness. This is similar to the idea presented within Watson and Mason's (2005) 'toolshed' quote (on page 20); that some examples are more frequently used, so to the individual they are somehow more representative of the entire class of examples. For instance, $x = 0$ is perhaps a less representative example of a quadratic equation than $(x-4)(x+3) = 0$. Goldenberg and Mason suggest that such unrepresentative examples typically restrict the possible variation that can be found in the entire class of objects.

Within their discussion, Goldenberg and Mason note that Marton and Säljö's (2005) framework of phenomenography and variation theory are intuitively applicable to research on example spaces. Other authors already discussed here (such as Watson and Shipman, 2008) also make reference to this framework, and this thesis too uses phenomenography as a guiding framework for the main study component. Phenomenography is discussed in more detail in Chapter 3, but it is worth noting that phenomenographical studies are interested in describing the variation of different ways a phenomena can be experienced. Many authors of phenomenographic studies subscribe to the general thesis that in order for an individual to learn, that individual must become (perhaps implicitly) aware of possible variation. Therefore with the focus turned to example spaces, such methodologies assume that: (a) a 'better' example space is one with more variation, and (b) an individual who understands a concept 'better' will have a more varied example space.

2.2.6 The process of example generation

In the last subsection, studies were discussed that described the content and structure of (evoked) example spaces. These studies did not dwell upon the implicit and explicit thought processes that are required during example generation. This subsection explores the literature which deals with this area by outlining and discussing three recent studies. Each study attempts to draw conclusions about the example generation process by observing students tackling and reflecting upon example generation tasks. The tasks themselves vary from those presented in formal mathematical terminology (Antonini, 2006) to graduate students of mathematics, to tasks presented in everyday language (Asghari, 2007) to undergraduates.

Despite the range of tasks and students in the three studies, the authors' findings are similar, and so it is not unreasonable to conclude that the accounts of example generation presented in this subsection are generalisable to example generation more generally. Each study identifies students who use a 'picking and checking / trial and error' approach, others that work with examples to modify them until they are happy they are correct, and some students who deduce properties that the examples must have before thinking of an example itself. In each study students' strategies are reported as idealised versions, based on what the researchers observed and what the students commented upon (c.f. subsection 5.5 which discusses issues of intersubjectivity in research).

Hazzan and Zazkis's account of perceptions of example generation

Hazzan and Zazkis (1999) asked pre-service elementary school teachers three example generation questions. Each question asked for a mathematical object to be generated which satisfied certain criteria:

1. Give an example of a 6-digit number divisible (a) by 9, (b) by 17.

2. Give an example of a function that has a value of -2 at $x = 3$.

3. Give an example of a system of equations with two variables that has $(3, 7)$ as a solution.
   (Hazzan and Zazkis, 1999, p.1)

They found that students tended to construct their examples by making deductions based on the concepts in the question. In the paper, such deductions are described in terms of 'links' between the concepts in the question. Although the 'link' terminology shall not be used further in this thesis, I will describe their results within this framework (the following descriptions are summarised from Hazzan and Zazkis's paper):

Occasional links An individual prefers to work with a concrete object, using a 'picking and checking' strategy.

Procedural links An individual follows an algorithm to produce examples, and when such an algorithm is not given, one is created. Such an algorithm is typically based on the form of question.

Conceptual links The individual begins with a trivial or otherwise simple example (such as 999,999 for task 1, or $\{x = 3, y = 7\}$ for task 3), and then proceeds to modify it by adding some variation. Those who take this approach do not usually present the simple examples as their answer because they feel a need to create more sophisticated answers.

Hazzan and Zazkis's account is interesting because it combines cognitive and affective elements; each type of linking behaviour is presented with reasons (often in the form of quotations from students) why a student embarked on that process. In general, Hazzan and Zazkis inferred that barriers to example generation are often emotional rather than mathematical, indicating that the freedom allowed in an example generation task can result in uncertainty and students wishing to 'quit and avoid making choices when there is no one definite way to proceed.' (Hazzan and Zazkis, 1999, p.11).

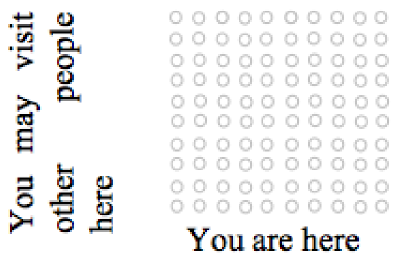
In the research presented in this thesis (see Section 6.4), it could be argued that some of the students in my study made occasional, procedural, and conceptual links (for instance see Haroon's comments in subsection 6.4.2, Phalgun's comments in subsection

6.4.1, and Valter's comments in subsection 6.4.3, respectively). However, unlike Hazzan and Zazkis's account, the results in this thesis present affective dimensions separately from the cognitive strategies.

Asghari's example generation / example checking structure

Compared with Hazzan and Zazkis's emphasis on the affective aspects of strategy choice, Asghari's (2005, 2007) research is concerned with the interplay between how the example is generated and how the example is checked. Asghari gave undergraduate students from various subjects (not just mathematics) a task which was related to an equivalence relation:

> A country has ten cities. A mad dictator of the country has decided that he wants to introduce a strict law about visiting other people. A 'visiting-city' of the city, which you are in, is: A city where you are allowed to visit other people. A visiting law must obey two conditions: (1) when you are in a particular city, you are allowed to visit other people in that city, and (2) for each pair of cities, either their visiting cities are identical or they mustn't have any visiting-cities in common. (Asghari, 2007, p.26)



Observing students tackling the problem, Asghari argued that there are two main ways to go about the generating of a solution, which I paraphrase here:

Conceptual Generating The individual's concept definition is used to construct the example. Provided the concept definition is appropriately aligned to the formal definition, examples generated by this process will be "correct" without the need for further checking. In the context of the mad dictator task, students may make a deduction about the cities (such as the symmetry of dots), and base their example on that.

Figural Generating The concept definition is temporarily set aside, and the individual uses the representation system of the question or answer to guide generation[2]. This requires subsequent comparison (checking) with the concept definition in order to confirm if the example is of the required type. In the context of the mad dictator task, students may begin by filling in some dots, and work out which other ones must also be filled.

This distinction in approaches to generating examples is partly a consequence of the type of example generation task Asghari chose. In the mad dictator task the examples solicited were (a) of a non-formal definition that was not explicitly related to the participants' prior knowledge, and (b) the required representation system of the examples was shading dots on a fixed grid, thus lending itself to figural generating which could loosely be deemed as colouring in some dots and seeing what happens. Despite this, the conceptual/figural framework described here gels well with the terminology of concept image and concept definition.

A shortcoming with Asghari's research is that due to the focus on the checking part of example generation, it is still very unclear how the actual generating takes place in either figural or conceptual generating. This is perhaps a more interesting aspect; as noted by Hazzan and Zazkis (1999) and Selden and Selden (1998), the lack of pre-learned algorithm-like process to solve example generation tasks can be disconcerting to students, and it is often not clear which route a student could or should use to find one.

The final study presented in this subsection is more generalisable to other example generation tasks compared with Asghari's findings and, unlike Hazzan and Zazkis's results, it does not intertwine the choice of strategy with the strategies themselves. Although it does not add substantially to the types of example generation behaviour noted so far, it does widen the scope further than either the occasional/procedural/conceptual or the conceptual/figural distinctions.

Antonini's example generation strategy framework

Antonini's (2006) framework arose by studying what mathematics graduate students did when they were asked to give examples of mathematical objects with varying properties.

[2]In the case of the 'mad dictator' problem the representation system of the answer is the pattern of circles to fill in.

The properties requested were:

1. Give an example of a real function of a real variable, non constant, periodic and not having a minimum period.
2. Give an example of a function $f : [a, b] \cap \mathbb{Q} \to \mathbb{Q}$ (with $a, b \in \mathbb{Q}$) continuous and not bounded.
3. Give an example of a binary operation that is commutative but not associative.[3]
4. Give an example of three natural numbers, relatively prime, whose sum is a number which is not prime to any of them.
   (Antonini, 2006, p.58).

After exploring the processes self-reported by participants during the example generation tasks, Antonini concluded that there were three main approaches to example generation:

Trial and error The example is sought among some recalled objects: for each example the subject only observes whether it has the requested properties or not.

Transformation An object that satisfies part of the requested properties is modified through one or more successive transformations until it is turned into a new object with all the requested characteristics.

Analysis Assuming the constructed object [exists], and possibly assuming that it satisfies other properties added in order to simplify or restrict the search ground, further properties are deduced up to consequences that may evoke either a known object or a procedure to construct the requested one. (Antonini, 2006, p.58–59)

Antonini noted that the trial and error strategy was far more common than either the transformation or analysis, a comment that was also made by Iannone et al. (2009), who asked undergraduates to generate examples of functions satisfying certain criteria and then analysed their data based on Antonini's classification. As was noted at the start of this subsection, 'trial and error'-like strategies were also present in other studies: Hazzan and Zazkis (1999) also commented that 'picking and choosing' was a common feature of the occasional links approach, and it could be argued that the notion of *figural*

[3] Antonini acknowledges that this question is modified from Zaslavsky and Peled (1996).

generation is a type of trial and error, but with the form of an answer guided by the (required) representation.

Transformation-like strategies can also be seen in Hazzan and Zazkis's (1999) notion of conceptual links, where a trivial or prototypical example is modified to become more sophisticated without losing the features which made it valid in the first place. However, Antonini's transformation strategy also encompasses strategies where the object chosen initially does not meet the requirements.

Finally, Asghari's (2005; 2007) conceptual generating strategy is similar to Antonini's analysis strategy. When conceptually generating an example, concept definitions are used to construct the example, and provided the concept definition is appropriate the example will be valid. In an analysis strategy, further properties are deduced or assumed in order to restrict the search to a more accessible portion of the example space.

When presenting the results of a phenomenographic data analysis of the main study data in Chapter 6, Antonini's classification will be used as an underlying basis the dimension of variation that describes the strategies students used to generate examples.

2.3 Real Analysis

In this final section, the attention is turned to a specific undergraduate module: Analysis. Anecdotally, it is well observed that students struggle with the content of an Analysis module, and that the point of understanding formal definitions and proof within Analysis is a milestone for many students, as reflected in two famous mathematician's autobiographies:

> My eyes were first opened by Professor Love, who taught me for a few terms and gave me my first serious conception of analysis. [...] I learnt for the first time as I read [Jordan's *Cours d'analyse*] what mathematics really meant. (Hardy, 1967, p.147)

> The day when the light dawned [...] I finally understood epsilons and limits, it was all clear, it was all beautiful, it was all exciting. [...] I had become a mathematician. (Halmos, 1985, p.48)

In Section 2.1 it was noted that when studying mathematics at university many students have difficulty when moving from "describing to defining, from convincing to proving in a logical manner based on these definitions." (Tall, 1991b). This section explores these issues in the context of dealing with formal definitions in an Analysis module. It will be identified that there is relatively little research on how students work and reason with the simpler definitions found in Analysis, and it is this gap in the literature that will be addressed in this thesis.

2.3.1 Why do students find Analysis so difficult?

It is widely regarded that Analysis is a difficult module to study (Alcock and Simpson, 2001, 2002; Artigue, 1991; Meehan, 2007; Weber, 2008). In a paper exploring research in the teaching and learning of Analysis, Robert and Speer (2001) listed the main mathematical areas they considered to be part of Analysis. These were:

> Functions of one and several variables, considered both locally and globally, limits, continuity, derivatives, sequences, definite and indefinite integrals, and differential equations. (Robert and Speer, 2001, p.283)

For now, this working definition is used, but later in the thesis 'Analysis' will be used in conjunction with the content of Analysis modules studied by students who took part in the example generation tasks. Robert and Speer (2001, p.288) go on to note that contained within several of these topics, in particular continuity, derivatives and sequences, there is a conceptual difference between exploring properties of a particular object (they give the example of the function $f(x) = 7x^3 + 5x^2 - 2x + 4$), and exploring more general properties (say of an arbitrary cubic). They suggest that, for students of mathematics, the disparity between these two ways of dealing with content causes difficulty, especially when it is considered that the focus at the secondary level may well have been more on the particular than the general.

In keeping with this observation, Alcock and Simpson (2002) found that some students extrapolate general features (sometimes in the form of definitions) from single prototype examples, taking the properties of this example as the basis of definitions for a wider class of objects. Such a *prototype procedure* can be successful if the example used is sufficiently general: they present results from an interview where a pair of undergraduate students successfully argue about the convergence of $\sum \frac{(-x)^n}{n}$ for fractional x based on the single prototype $\sum \frac{(-1/2)^n}{n}$.

These two papers suggest that the relationship between general and particular objects can both aid learning and also cause difficulties for students studying Analysis, but such conclusions can also be drawn for many modules met by students early on in tertiary education. Stewart and Thomas (2010) attempted to aid students' understanding in Linear Algebra by dealing with the ideas of the span and a basis for a set of vectors both in terms of individual vectors and in more general terms via concept maps, which I would argue was an effort to help students implicitly compare the behaviour of specific and more general objects. In terms of why Analysis is regarded as difficult, say compared to modules such as Linear Algebra, other authors such as Alcock and Simpson (2002) argue that in Analysis the definitions students meet are more complex. The next subsection focuses on why this may be the case. It looks at the role of definitions in mathematics, eventually concluding that despite the complexity of some definitions within Analysis, there is a need for research which concentrates on how students reason with relatively simple definitions in an Analysis setting.

2.3.2 The role of definitions

Mathematical definitions are important structural concepts within mathematics but often create a serious problem for students (Dubinsky, 1991). Definitions constrain to which objects a theorem can be applied and they outline the scope of technical terms contained within proofs. Without a firm understanding of the status of definitions within mathematics and how definitions should be used, students will struggle in important areas such as proof (see Weber, 2002, for instance).

Recall from Section 2.1.3 that Tall and Vinner's (1981) construct of a concept definition aimed to distinguish between reasoning based on the formal mathematical definition of a concept and reasoning based on alternative ways of understanding the concept. These alternative ways of understanding may originate from spontaneous conceptions based on the everyday use of words, and a student's prior experience with the mathematics in question (Cornu, 1991). These constructs are useful when exploring students' understanding of specific definitions (such as those in this thesis), but less useful for exploring students' epistemological views of definitions. This subsection first examines these views, before returning to definitions that are more closely related to Analysis, and in particular the role of quantification in such definitions.

Students' epistemological views of definitions

To students, what constitutes a (good) definition may change based on experience and instruction (Zaslavsky and Shir, 2005), and yet definitions are often introduced to students with the assumption that the lecturer sees them in a similar way to the student (Vinner, 1976). The strict role that definitions play within mathematics is often not clear to students; when it is appropriate to use definitions, and what is nature of their role and purpose more widely within mathematics? Can they be created on the spot to be used, or are they universal truths that we are merely describing? Are some definitions 'better' than others?

Students entering university are rarely given the chance to 'play around' with definitions, instead definitions are typically regarded by students in a lexical sense (Vinner, 1976), in particular that definitions are not canonical and depend on the situation and context. Vinner gives the example of "a house is a building for human habitation", which depending on the context can be regarded as a definition, statement or fact, noting that within mathematics definitions play a far stricter role, as part of the 'formalistic approach' to mathematics (as described by Vinner, 1976), but students do not appreciate this shift in the way the word definition is used.

Alcock and Simpson's (2002) study, discussed in the previous section, found that students can be reluctant to work with definitions. In a study with undergraduate students studying Analysis, some students needed several prompts to be persuaded to write down a complete definition and use it. The paper concluded with the observation that when asked to work with definitions (say in the context of a lecture) students may instead work with a prototype example, modifying it to be compatible with the definition and basing further reasoning on this example rather than the definition. In the pilot study chapter of the thesis, this phenomenon is observed with respect to the definition of 'increasing' and 'decreasing' sequences where students treated the combination of these definitions to be a sequence which increases for a region then decreases later, rather than the formal interpretation of the definitions: a constant sequence (see Chapter 4, and Edwards and Alcock, 2008).

Logic and quantification in definitions

Aside from studies examining students' epistemological views of definitions, much research has been conducted which explores what students find difficult about the meaning of specific definitions. It is within this body of research where authors have considered why definitions within Analysis are particularly difficult.

Students meet and struggle with definitions in many mathematics courses, whether relatively simple definitions of the style 'all a are b' (where b is a simple statement) (Epp, 2003), to more complex definitions say the continuity of a function within an Analysis course (Tall and Vinner, 1981), or linear independence in linear algebra (Uhlig, 2002). What makes the latter definitions especially difficult is the presence of one or more logical quantifiers (Alcock and Simpson, 2002; Dubinsky et al., 1988; Vinner, 1976).

Undergraduate students of mathematics tend to first meet existential and universal quantifiers ($\exists$ and $\forall$) within proof based courses such as Analysis, Abstract Algebra, or possibly in a bridging course designed to help students move from more procedurally-based courses such as Calculus and Differential Equations (Selden and Selden, 1999).

Dubinsky and Yiparaki (2000) considered how mathematics undergraduates deal with statements involving a single combination of a universal and existential quantifier. Some statements were in everyday language:

> There is a mother for all children

and some were written with mathematical content:

> For every positive number a there exists a positive number b such that $b < a$.

Noting in the paper that in the real life examples the meaning of the statement is more ambiguous (the first statement may arguably be interchanged with 'all children have a mother' in everyday language), they asked the students if the statement was true, and to give an explanation of why. From those explanations they decided if the student was treating the statement in the form '$\forall x \, \exists y$', or in the form '$\exists x \, \forall y$.' The study also asked students questions about the meaning of the quantifiers in isolation, and what would need to be changed in order to negate statements they had identified as true. Whilst students were successful in answering questions about the quantifiers in isolation, when statements combined quantifiers the students did not use the syntax of a statement when analysing it, they either paid no attention to the quantifiers in the everyday statements and continued to fail to do so in the mathematical statements, or attended quantifiers in the everyday statements and failed to do so in the mathematical statements. (*ibid.*, p.264). Few students could identify the changes needed in order to negate statements they claimed to be true.

In an earlier study, Dubinsky et al. (1988) looked at how students interpret and negate statements which contain multiple quantifiers. They found that students struggled with multiple quantifiers in the same definitions, and statements such as

> For every tire in the library, there is a car in the parking lot such that if the tire fits the car, then the car is red. (Dubinsky et al., 1988, p.45)

were extremely difficult for students to deconstruct; students were asked how they might validate such a statement, commonly being side-tracked by the everyday variables (tire, car) opposed to the logic of the statement. Both these studies identify discrepancies between students' concept definitions and the formal mathematical definitions, highlighting a particular issue with respect to quantification.

Aside from the everyday meaning of statements containing multiple quantifiers, other researchers have considered the effect of making objects in statements real life, or imagined. In one part of a study by Sá et al. (1999), undergraduate students from a variety of subjects were asked whether 24 logical statements were true. Eight of the statements were worded such that the truth of the statement was in conflict with the everyday meaning of the conclusion such as the (false) statement:

> (1) All living things need water; Roses need water; therefore, Roses are living things.

Eight statements' truth was not in conflict with the everyday meaning of the conclusion such as the (true) statement:

> (2) All fish can swim; Tuna are a fish; therefore Tuna can swim

The final eight statements involved imaginary content:

> (3) All opprobines run on electricity; Jamtops run on electricity; therefore, Jamtops are opprobines.
> (all from Sá et al., 1999, p.500)

They found a significant 'belief-bias', in that more students incorrectly stated the truth of statements similar to (1) than either those similar to (2) or those similar to (3), with a larger effect size for the (1) and (3). This, and similar studies from the cognitive psychology literature (e.g. Evans et al., 1983) confirm that spontaneous conceptions from the everyday meanings of logical phrases may be present in students' evoked concept images of definitions.

Definitions within Analysis

At the start of this section focusing on Analysis, on page 38, it was noted that a typical Analysis module places emphasis on formal definitions and proof. Whilst this is true for many modules met by students early in undergraduate mathematics, Analysis is unusual in that its definitions typically contain multiple, nested, quantifiers. The module is rich in everyday language (e.g. limit, as discussed below), and authors such as Alcock and Simpson (2002) also note that Analysis conducive to visual representations of concepts and objects, which may result in students reasoning with inferences from imagery rather than formally via definitions. In a subject such as group theory, they note, students may make more use of formal definitions because visual representations are less readily available (Alcock and Simpson, 2002, p.33).

In Analysis students spend some of their time proving statements which they find obvious, despite not being able to follow the proofs themselves (Tall and Vinner, 1981, p.17). The combination of this observation, and the factors noted above result in students finding Analysis difficult and, perhaps due to this, Analysis has been much studied by researchers in Mathematics Education. However, amongst the definitions taught as part of a typical Analysis course, most research has been conducted on those which contain multiple quantifiers, such as the continuity of a function, or the limiting process (Alcock and Simpson (2009b) argue this is because the limiting process is central to much of Analysis). Considerably less has been researched on how students deal with simpler definitions.

The limiting process

Here, mathematics education research on the limiting process in Analysis is discussed to give a flavour of the type of work that exists on how students work with definitions.

Historically mathematicians did not agree completely on notion of converging to a limit, with theorems published which 'admitted exceptions' such as Cauchy's theorem that the limit of a convergent sum of continuous functions is itself convergent (Jahnke, 2008; Sørensen, 2005). Various authors have shown that, for many students, the everyday meaning of limit impedes their understanding of the formal mathematical definition (Cornu, 1991; Monaghan, 1991; Sierpinska, 1987; Tall and Vinner, 1981). Research has shown that students do not see the limit of a function as useful in finding approximations (Çetin, 2009), and that they have difficulty moving from a dynamic conception of a limit to a more formal one (Williams, 1991). Other authors argue that difficulties moving to a more formal conception of a limit are 'at least partially a result of insufficient development of a strong dynamic conception' (Cottrill et al., 1996, p.190).

2.4 Rationale for focusing on sequence properties

As illustrated in Section 2.1, many researchers consider the transition from secondary to tertiary mathematics a particular trouble spot for students (Clark and Lovric, 2008, 2009; Gueudet, 2008; Hong et al., 2009; Wood, 2001). Outside the literature dealing with social aspects of this transition, much of the literature in this area considers the transition to formal thinking within mathematics a particular challenge (Dubinsky, 1991; Harel and Sowder, 2005; Sfard, 1991; Tall, 2004; Tall and Vinner, 1981; Vinner, 1983). Working with formal definitions is seen as a particular obstacle (Alcock and Simpson, 2002; Vinner, 1976), and this has resonated with my own experiences during a mathematics degree. As Alcock and Simpson (2002) noted, analysis is considered to be a difficult course at university by both students and lecturers, and this is in part due to the complexity of formal definitions.

There has been much research on the convergence and limits of functions and sequences (Bérgé, 2006; Cornu, 1991; Monaghan, 1991), and in particular students have have problems dealing with quantifiers in logical statements (Dubinsky et al., 1988), and multiple nested quantifiers (Cottrill et al., 1996). Considerably less attention has been given to properties of sequences which are simpler (in terms of their use of quantifiers) than convergence and limiting behaviour, such as the properties *increasing* and *decreasing*. For instance contrast the definition that a sequence converges to a number, l, with the definition of an increasing sequence.

Converging to a limit The sequence (a_n) converges to the number l if and only if $\forall \epsilon > 0, \exists N \in \mathbb{N}$ s.t. $\forall n > N, |a_n - l| < \epsilon$.

Increasing The sequence (a_n) is increasing if and only if $\forall n \in \mathbb{N}, a_n \leq a_{n+1}$.

In terms of the number and interaction of the quantifiers in the two definitions, the latter is far simpler than the much studied limiting process.

The work of the thesis addresses this gap in the literature by concentrating on students' interactions with these simpler and less studied definitions. The studies in the thesis will show that many mathematics undergraduate students still struggle when understanding, applying and negating these simpler definitions.

2.5 Summary of chapter

The chapter began with a summary of the social, epistemological and cognitive issues related to the progression to study mathematics at tertiary level. Students are given less guidance but more responsibility for understanding the content their work, and the nature of the mathematics has changed from application of techniques to a need to understand how mathematical arguments are dependent on proper use of formal definitions. Rather than reasoning with concept definitions, many students reason primarily with concept images which may contain spontaneous conceptions which are not good representations of the whole class of objects constrained by a definition.

At the same time as they are expected to reason with formal definitions, many students are reliant on empirical proof schemes; they are persuaded by arguments that involve specific examples of mathematical objects behaving in a certain way. Students can therefore run into difficulties in situations where the properties of such examples may not be generalisable to the whole class of objects under consideration.

Research on how students work with examples often uses the technique of example generation to explore students' example spaces. Recently produced work suggests that there are a series of different strategies students typically employ when generating examples, but in general little is known from a students' perspective how to about generating examples.

Within the different modules studied at the tertiary level, Analysis is regarded by many as particularly difficult. It has been suggested that this is partly because definitions within Analysis are particularly complex, often containing nestled combinations of quantifiers. While Analysis has attracted much attention from researchers in mathematics education in topics such as limits, there has been little research however looking at how students interact and reason with simpler definitions within an Analysis context.

This thesis uses an example generation task to explore how students interact and reason with simple definitions, shedding light on students' concept images related to these simple definitions. It simultaneously explores how students go about (and report how they go about) example generation in such a context.

Chapter 3

Methodology

The main study reported in this thesis (see Chapters 5 and 6) has been framed within the research specialism of phenomenography, first described and used by Ference Marton and his research group from the University of Göteborg in Sweden. This methodology chapter is an introduction to and literature review of phenomenography.

This chapter first gives an outline of phenomenography in Section 3.1, noting its unit of analysis (ways of experiencing a concept), and two methodological assumptions (a second-order orientation, and the need to bracket during research). Section 3.2 then considers phenomenography's ontological and epistemological stances, and compares these to other qualitative methodologies in mathematics education, including the related methodology of variation theory. After this, Section 3.3 discusses how authors of phenomenographic and variation theory studies in educational research typically present their results in the form of an outcome space. This section also considers the form that such an outcome space may take. Section 3.4 then outlines the data collection and analysis methods associated with phenomenography, and considers how authors have justified the validity and reliability of their studies. The last section, Section 3.5, outlines some of the criticisms that have been made towards phenomenography and how these relate to the research in this thesis.

In this chapter the word "students" has been used to represent the individuals who are being studied. This is because in my study the individuals who are being studied are undergraduate students of mathematics. However, if direct quotations are taken from a study whose participants were not students and the author(s) use a different term for instance "subjects", I have not edited their language.

3.1 Outline of phenomenography

The research specialism of phenomenography was developed by Marton and his research team in the 1970s (Marton, 1975; Marton and Säljö, 1976; Marton and Svensson, 1979). The word "phenomenography" was first used in Marton (1981); initially the approach was used when investigating qualitative differences observed when students were asked to read, comment upon, and then summarise a text. In an early article, which, according to Entwistle (1997) should not strictly be labelled as phenomenographic, Marton noted:

> The present study is likely to be considered rather unconventional from the methodological point of view. Usually, the variables are, so to speak, given and one attempts to ascertain relationships between them; in our case the categories of description are themselves results (Marton, 1975, p.275).

As Bowden (2005, p.12) notes, "the object of study is not the phenomenon being discussed per se, but rather the relation between the subjects and that phenomenon." Marton was not alone in moving away from the type of research described as *scientific method* (Cohen et al., 2007; Pring, 2000), i.e. an approach to research which first identifies variables, then generates hypotheses and finally tests those hypotheses by controlling for as many variables as possible, attempting to show causality for the remaining variables. In a similar vein to other interpretive approaches to research (phenomenology, ethnomethodology, symbolic interactionism, etc), Marton's focus was on describing students' descriptions of and reflections upon a concept, rather than research describing the concept directly in terms of interaction of variables.

The unit of analysis in phenomenographic studies therefore is students' *ways of experiencing* a concept (Marton and Booth, 1997). A way of experiencing a concept is synonymous with phrases such as "conceptions", "ways of understanding", "ways of comprehending" and "conceptualisations" (Marton and Booth, 1997, p.114). Säljö (1997, p.175) argues "conception" may be an alternative translation from the original Swedish *uppfattning*, and I would also include "descriptions of", "approach towards", "view of", "knowledge of", and "concept image of" as synonyms. Within the phenomenography literature it is usual to use the word *experiences* with the intention of encompassing this wide range of statements, and this thesis will also follow this practice.

Research in phenomenography therefore aims at "description, analysis, and understanding of experiences; that is, research which is directed towards experiential description" (Marton, 1981). The focus is not so much on the source of these ways of experiencing, but on the categories of description of ways of experiencing, and the variation between these categories. In particular, this is what separates phenomenography from other interpretive approaches to research. A particular individual's experience of a concept is not assumed to be constant, but dynamic depending on the situation. The object of focus therefore is not to categorise or label an individual's way of thinking, it is to categorise the possible experiences of an individual. As explained by Marton:

> If we accept the thesis that it is of interest to know about the possible alternative conceptions students may have of the concept or the aspects present in, related to or underlying the subject matter of their study, it is these questions specifically which we must investigate. (Marton, 1981, p.183)

An individual who is aware of more aspects of a phenomenon is considered to be more knowledgeable about that phenomenon, and will have the flexibility be able to handle it in more efficient ways (Marton and Booth, 1997, p.117). Learning is therefore considered to be the process of becoming capable (or aware) of doing something, and "the pattern of variation inherent in the learning situation is fundamental to the development of certain capabilities." (Marton and Tsui, 2004, p.15).

Although there have been refinements between Marton's approach in 1975 and later phenomenographic studies, two methodological assumptions have remained throughout (Hasselgren and Beach, 1997). The first assumption is that phenomenographic research should approached by considering second-order descriptions and questions. Second order descriptions are accounts of peoples ideas about the world from those people themselves, whereas first order descriptions may include statements about the world written by the researchers themselves (Marton and Booth, 1997, p.178). In an educational context, a researcher is not concerned with describing a concept (e.g. what are the salient features of multiplication), but rather students' ways of experiencing the concept (e.g. what is students awareness of multiplication). The second assumption of phenomenography is that of the need to *bracket* during research. Bracketing means that researchers should only focus on the parts of the concept that the student is focusing on at that particular time, suspending judgement on other parts, both during data analysis, but also to a certain extent when interviewing students (Marton and Booth, 1997, p.119).

3.2 Comparison to other research methodologies

The previous section gave an overview of the research specialism of phenomenography, but did not comment on how it relates to other qualitative educational methodologies, although a passing reference was made to the 'big three' qualitative methodologies of phenomenology, ethnomethodology, and symbolic interactionism (Cohen et al., 2007). This section begins by considering phenomenography's ontological and epistemological stances, before relating it to such methodologies.

3.2.1 Ontological stances in phenomenography

Phenomenography is distinct from other research methodologies due in part to its ontological position, or rather its lack of a firm position. Svensson (1997, p.164) argued that phenomenographic research deliberately makes few claims in this area:

> The position taken differs from empiristic and positivistic assumptions about observations as facts, and knowledge as inductively based on facts. It also differs from rationalistic, mentalistic and constructivistic assumptions about knowledge as rational or mental constructions within a more or less closed rational and/or mental system. Thus the view of knowledge is that it is relational, not only empirical or rational, but created through thinking about external reality. ... Knowledge is seen as dependent upon context and perspective. (Svensson, 1997, p.165)

The above quotation includes the phrase "external reality", which is a phrase that could be associated with a positivist ontological stance, i.e. that the world can be described as existing independently from human experience, and so a complete and scientific explanation of reality can be made (Pring, 2000). Similarly, the previous section referred to 'mathematical concepts', which could be interpreted as a positivist statement: that these concepts somehow exist beyond human experience. From a phenomenographic standpoint, the existence or otherwise of an external reality is irrelevant; phenomenographers do not claim to study the reality of the world, but instead individuals' accounts of their awareness of conceptions of the world. Svensson (1997, p.165) argued that within the phenomenography literature different researchers have taken different ontological stances. It is tough to confirm this claim because the majority of authors of research papers in phenomenography do not comment on their ontological stance.

My ontological stance is that I believe mathematicians clearly define objects with axioms and definitions, and attempt to deduce consequences and results based on relationships between these objects. Whatever the status of knowledge contained or represented by these objects, be it via social acceptance (Ernest, 1993) or objectivity via internal certainty (Rowlands et al., 2001), or some Platonic reality argument, these objects still remain clearly defined. Beyond that, in keeping with the phenomenographic stance of the thesis, I do not have a firm opinion on the nature of reality. My attention is focused on students' awareness of these clearly defined objects.

3.2.2 Epistemological stances in phenomenography

In phenomenographic research, an epistemological assumption is that knowledge of a concept is the ability to experience variation of that concept. To know that something is red means that we must have experienced variation between red and other colours in the past. It might be considered that the notion of an individual becoming aware of variation is comparable to an individual constructing knowledge, say in individual constructivism, as described by Piaget (1950) or radical constructivism described by von Glasersfeld (1984). Both these theories do not fit well with phenomenography, however. Piaget's scheme is based on stages of development, which is similar to phenomenography's collection of different ways of experiencing a concept and variation between them, but the latter does not classify learning as development through these stages. Von Glasersfeld's interpretation of constructivism involves the denial of an external reality, whereas phenomenography makes no assumptions in this area (see the last subsection).

Alternatively, it may be considered that phenomenography is compatible with social-cultural theories, which stem from accounts of learning such as Vygotsky's (1978) social constructivism, where learning is seen as a relationship between outer and inner speech. Perhaps from a phenomenographic perspective the relationship between outer and inner speech is the mechanism allowing a student to become aware of more or less variation? As discussed earlier, phenomenographers are not so concerned on the mechanisms of the ways of experiencing, but the variation between possible ways of experiencing.

Marton and Booth clarify their position with regards to various theories of constructivism in the following quote:

> individual constructivism ... regards the outer (acts, behaviour) as being in need of explanation and the inner (mental acts) as explanatory, whereas the reverse is true for social constructivism. ... One should not, and we do not, consider person and world as being separate. ... [Students are] neither bearers of mental structures nor behaviourist actors. ... Thus the dividing line between "the outer" and "the inner" disappears. ... The world is not constructed by the learner, nor is it imposed upon her; it is constituted as an internal relation between them. (Marton and Booth, 1997, p.13)

3.2.3 Comparisons with other qualitative methodologies

If two researchers are interested in exploring how students think, and reject an approach primarily based on the scientific method, they will conduct their research in superficially similar ways (Cohen et al., 2007). Theory will be emergent, grounded in the data (Glaser and Strauss, 1967), and will attempt to yield insight and understanding of an individual's behaviour. Phenomenography thus has much in common with other qualitative research approaches, but it is distinct from the 'big three' of phenomenology, ethnomethodology, and symbolic interactionism. This section concludes by briefly contrasting the various approaches.

Phenomenography has much in common with phenomenology including a second-order approach to research, bracketing out other ways of seeing the world and so taking individuals' accounts at face value (Cohen et al., 2007). From a historical view, phenomenography was not developed from phenomenology (Svensson, 1997), and there are several key differences between the approaches. First, in a phenomenological account, the aim is to describe the essence of a concept as completely as possible, whereas in a phenomenographic account the researcher is interested in the different ways individuals experience the concept and the variation between these ways of experiencing (Marton and Booth, 1997, p.116). Describing the variation is not the same as describing the concept, and it is not thought that the concept is in some way 'the same' as its phenomenographic account (Larsson and Holström, 2007, p.59). Second, in a phenomenological account, a researcher may primarily use their own experiences of a concept (e.g. Mason, in press), rather than accounts from others.

Ethnomethodolgy and symbolic interactionism also take a second-order approach to research, but have less in common with phenomenography. Linguistic ethnomethodologists study the use of language and conversation in everyday life, and situational ethnomethodologists examine how people negotiate the social contexts in which they find themselves (Cohen et al., 2007). Clearly phenomenography has little in common with these research specialisms . Symbolic interactionism relies on a distinction between the 'natural' world which includes human drives and instincts, and an external world which includes symbols, language and objects. It studies the interaction between these worlds, which is an interaction that does not exist in phenomenography (see the quote from Marton and Booth, 1997, at the end of the last subsection).

3.2.4 Variation Theory

Some recent studies which can be considered phenomenographic (Al-Murani, 2006; Runesson, 1999, 2006; Watson and Mason, 2004, 2006) have shifted the perspective so that the objects of research are the categories of description themselves. Such authors effectively first decide which dimension(s) of variation they are studying, then explore students' awareness of this dimension. Such studies typically label themselves within the specialism of *Variation Theory*, and have the goal of describing these dimensions of possible variation.

I see the distinction between a phenomenographic account and a variation theory account as follows. In a phenomenographic account, the object of research is students' awareness of a phenomenon and the qualitatively different ways of experiencing the phenomenon is presented in categories of description, grouped into dimensions of variation. Marton and Säljö (2005, p.336) call this the "referential aspect—i.e. a particular meaning of an object [or concept], anything delimited and attended to by subjects." In a variation theory study, a researcher is interested in exploring a particular dimension of variation, and so looks at individuals' awareness of a phenomenon in this context. Marton and Säljö (2005, p.336) call this the "structural aspect—i.e. the combination of features [of the concept] discerned and focused upon by the subject."

A study that focuses on the referential aspect will ask general questions such as 'how do you go about learning maths?', whereas studies that focus on the structural aspect will typically introduce a mathematical object to the students and then ask them about features of that object associated with the dimension of variation under consideration. It is difficult, if not impossible, to determine where phenomenography ends and variation theory begins, because even a strictly phenomenographic account will frame its results within the framework of variation theory. In some studies, a concept was introduced to the subjects in varying forms, and the subjects were interviewed about their experiences of the variation in the object (Asghari, 2004; Runesson, 2006; Watson and Mason, 2004, 2006). Others have studied the use of variation by teachers when they are teaching (Al-Murani, 2006; Runesson, 1999).

The main study of this thesis uses an example generation task to explore undergraduate mathematics students' awareness of sequences, and also their awareness of example generation. Even though the students were given a task as the focus of the interview, rather than asking referential questions such as "what is a sequence?", I still consider the thesis to be in the tradition of phenomenography rather than variation theory. This is because in the main study, dimensions of variation emerged from categories of description that were grounded in the data. The task was not used to explore a pre-defined aspect of sequences or example generation, rather it was used as a device to focus students' awareness towards these two areas.

3.3 Reporting dimensions of variation

As discussed in the previous section, phenomenographic research aims to describe the different ways individuals may experience a phenomenon, and reports these different ways in terms of categories of description which are then arranged into dimensions of variation. Together, the dimensions of variation form an outcome space of the research (see below). While the specifics of data collection and analysis are covered in Section 3.4, this section presents some examples of the way phenomenographic and variation theory researchers present these dimensions of variation.

Before undertaking a phenomenographic piece of research, the researcher must decide the scope of the study. The conception or object of research must be chosen, and possibly the types of variation to focus on, whether they be referential or structural, or even more precisely defined such as Watson and Mason's (2004) choice of values in equations and coordinate points.

3.3.1 The outcome space

A dimension of variation is a set of related categories of description which are stable between situations, even if individuals may "move" between categories on different occasions (Marton, 1981, p.195). For instance, the dimension of variation Using Definitions (see Section 6.2) outlines students' experiences of formal mathematical definitions during in the main study's task. This dimension of variation includes the categories of description Def-A *Unaware of Definitions*, Def-B *Refers to Definitions*, Def-C *Uses Definitions*, and Def-D *Manipulates Definitions*.

The categories of descriptions in a dimension of variation are usually hierarchical (as is the case for Using Definitions), although an individual's awareness may fluctuate between categories in a short space of time (for some students in the main study, this occurred in the span of a single interview). This is what distinguishes a phenomenographic dimension of variation from a developmental scale construct such as Piaget's (1950) stages of development, where a student is at a certain level, and progresses along a hierarchy of categories as their thinking matures. In a phenomenographic study, awareness associated with more the sophisticated of categories of description, a more varied understanding of a concept, and being more knowledgeable about a concept are synonymous (Marton and Booth, 1997, p.107).

Within a phenomenographic study there may be a single dimension of variation reported. In such cases the outcome space is the same as the dimension of variation. Other studies may have several dimensions; Crawford et al. (1994) reported on students' conceptions of mathematics and how it is learned (discussed in more detail in Section 3.3.2). In such cases, the outcome space is formed by combining the dimensions.

Some researchers have formed dimensions of variation from data sources that were not empirical studies with students, such as the historical emergence of a concept (Renström et al., 1990), or teaching experiments (Neuman, 1997). It is therefore not always the case that a dimension must be remarked upon by the students in the study in order to be included, although phenomenographic researchers attempting to ground their results in their data may consider extensive use of results from outside the study as bringing into question the validity of such research.

3.3.2 Previous research exploring dimensions of variation

This subsection consider how authors have presented dimensions of variation in published phenomenographic studies. It concentrates on studies within mathematics and science education literature to date, at the cost of excluding research in some areas, for instance studies that have been conducted in the field of medicine such as Larsson and Holström's (2007) study on anaesthesiologists' understanding of their work.

Variation in students' approaches to learning

Some of the earliest studies that could be described as phenomenographic took place before the theoretical framework of phenomenography had taken shape. Studies include those looking at students' reading of texts (Marton, 1975; Marton and Säljö, 1976; Svensson, 1977). These and other studies concerned with students' approaches to learning (both from phenomenography and other theoretical frameworks such as Yorke's (2006) social-cultural account) evolved into the well-known deep learning and surface learning distinction that is used commonly used in educational literature (this distinction was discussed in subsection 2.1.1). Marton (1975) gave students a passage of text to read, and then asked them to outline what that thought the meaning of the text was. Besides the partition of subjects into those that displayed deep and surface processing of the text, the paper's results section presented two dimensions of variation[1]: the object of

[1] The paper does not refer to these as dimensions of variation because it predates the formation of phenomenography.

focal attention, and the agent of learning. Within each dimension of variation there were two categories, each of which was hierarchically ordered:

Object of focal attention

- What is signified (what the object is about)
- The sign (the discourse itself, or the recall of it)

Agent of learning

- Active processing (the subject being the agent of learning)
- Passive processing (the subject not being the agent of learning)

(adapted from Marton, 1975, p.276)

In a more recent study, Vermunt (1996) observed within a phenomenographic framework what students actually did to learn, which is slightly different to their approach to learning. The students interviewed were from a variety of different degree programs and were of a wide variety of ages, all registered on degree courses at the Open University. Four learning styles were identified (1) *undirected learning style*, (2) *reproduction directed learning style*, (3) *meaning directed learning style*, and (4) *application directed learning style*. Further description of these categories is outside the scope of this exposition, but it is interesting to note that Vermunt did not comment whether the categories are hierarchically related in terms of their sophistication, although the nomenclature of the learning styles implies this may be the case.

In the field of mathematics education, Crawford et al. (1994) gave 300 beginning university students a questionnaire consisting of five open-ended questions on the students' conceptions of mathematics and their approaches to studying it. Twelve students were selected for an in-depth structured interview in order to clarify their written statements, and after a phenomenographic data analysis (see next section) they reported several dimensions of variation including the one reproduced in this section as Table 3.1. This contains students' approaches to learning mathematics, moving from an aim of reproduction to understanding.

TABLE 3.1: Categories of the responses for students' approaches to learning mathematics (Crawford et al., 1994, p.337)

	Category
A.	Learning by rote memorisation, with an intention to reproduce knowledge and procedures
B.	Learning by doing lots of examples, with an intention to reproduce knowledge and procedures
C.	Learning by doing lots of examples with an intention of gaining a relational understanding of the theory and concepts
D.	Learning by doing lots of difficult problems, with an intention of gaining a relational understanding of the entire theory, and seeing its relationship with existing knowledge
E.	Learning with the intention of gaining a relational understanding of the theory and looking for situations where the theory will apply

Variation relating to students' view of the concept itself

The studies in this section have a slightly different approach to those reported in the last. Rather than taking a phenomenographic approach to study students' concepts of what it means to learn and how they go about learning a concept, the researcher focuses the students' attention on a particular concept or object. Studies which do this vary in how precise the topic under discussion is, ranging from the very general "what do you think mathematics is?" (Crawford et al., 1994, p.335), to a more focused approach on a specific mathematical question (Watson and Mason, 2006). Both the studies cited here are discussed in more detail later in this section.

An example of a study with a specific focus is Renström et al.'s (1990) exploration of students' conceptions of what matter is, (see Figure 3.1 for a reproduction of their outcome space representation). Although the study is not mathematics-related it is included in this chapter for two reasons. The first is that the authors used a combination of methods in the study; not only did they interview twenty secondary school students, but they also used insights gained from the other research and the history of science. The second is that the outcome space they eventually present does not have linear dimensions. This is unusual, and reflects a judgement made by the researchers that there are two distinct ways of experiencing matter which are more or less equally sophisticated.

From mathematics education, another part of Crawford et al.'s (1994) study looked at the beginning undergraduate students' conceptions of mathematics, which can be found

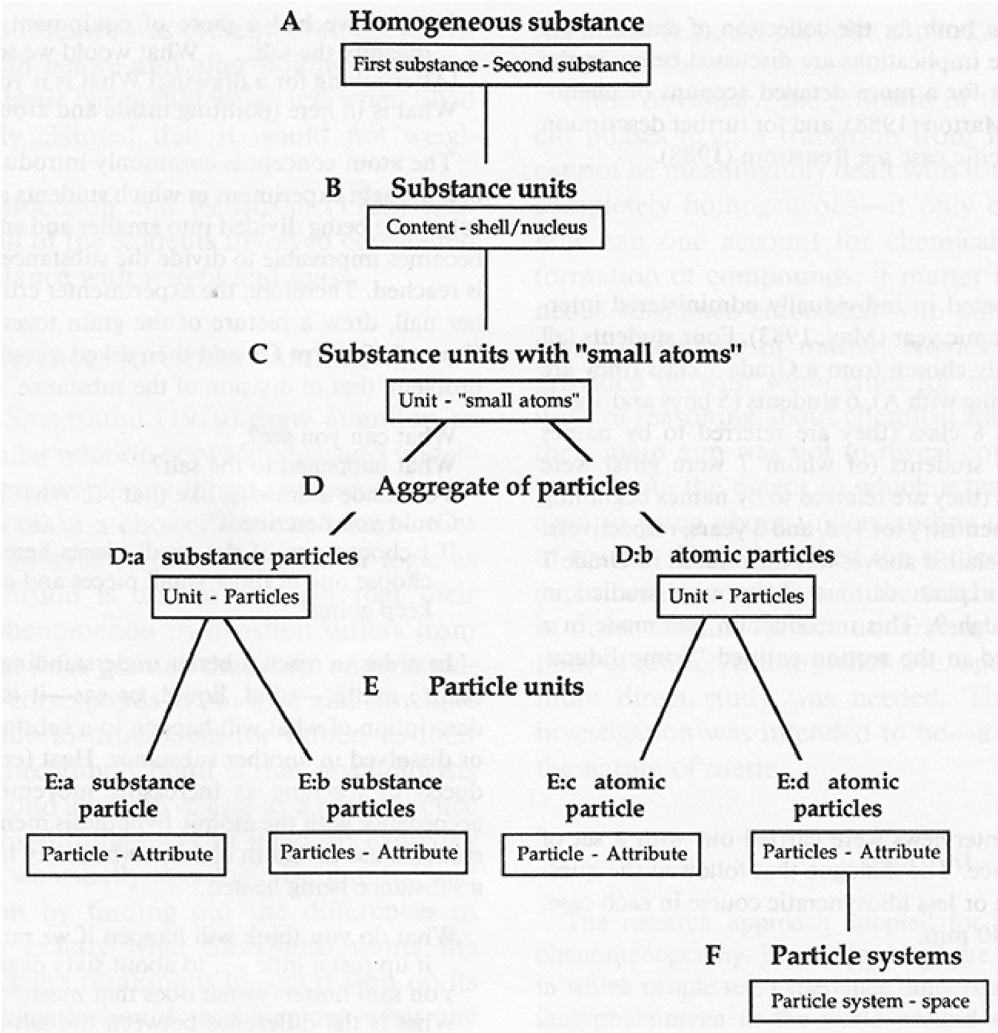


FIGURE 3.1: Students' conceptions of matter. The dimension of variation in the internal structure of each conception is indicted in the boxes. (from Renström et al., 1990, p.558).

in Table 3.2. This dimension was linear and hierarchically structured.

A study which focussed on students' experiences of a much narrower topic than the whole of "maths" was conducted by Watson and Mason (2006). They produced a task where different groups of inservice and preservice teachers explored the set of points which are the same distance apart using taxi-cab geometry, i.e. implicit use of the metric

$$d : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R} \qquad d\left((x_1, y_1), (x_2, y_2)\right) = |x_1 - x_2| + |y_1 - y_2|$$

The task began by giving the teachers a list of points which were 3 units from the point $A = (-2, -1)$, and asking them to both plot the points calculate the distances from

TABLE 3.2: Categories of the responses for students' conceptions of mathematics (from Crawford et al., 1994, p.335)

	Category
A.	Maths is numbers, rules and formulae
B.	Maths is numbers, rules and formulae which can be applied to solve problems
C.	Maths is a complex logical system; a way of thinking
D.	Maths is a complex logical system which can be used to solve complex problems
E.	Maths is a complex logical system which can be used to solve complex problems and provides new insights used for understanding the world

A (it was not stated that they were all the same distance apart). Watson and Mason observed different types of variation which they did not report diagrammatically but rather than in descriptive form. These are paraphrased as:

- Variation in the order the teachers approached the list of points. Some plotted the points to begin with, some calculated each individually then plotted, and some calculated and plotted each point in turn. Regardless of the approach, almost all began to make generalisations early on.
- Variation in the aspects of the task the teachers reflected on. Most made conjectures about which points may be 3 units away, and each successive example jolted them into thinking about other points which may not have fitted in with their sense of pattern. Some operated at a higher level; comparing the points with the Euclidean circle, asking themselves "why this A?" and "why 3?"

In the main study of this thesis, dimensions of variation that are related to students' view of sequences and sequence generation will include students' awareness of:

- The status and use of a formal mathematical definition (see Section 6.2)
- The way a sequence can be represented (see Section 6.3)
- The strategies used to generate sequences (see Section 6.4)
- The way students justify their answers (see Section 6.5)

Variation in teachers' awareness

In the previous sections, and throughout the chapter, the word student has been used in relation to the individuals whose awareness of variation is being studied, and this is because student understanding is the focus of my research. Phenomenographical studies do not always have the student as their focus however. Some researchers interview and observe teachers of mathematics, exploring their awareness of the different types of questions they use within a teaching context, and the outcome space consists of the dimensions of possible variation within the question types. Papers by Al-Murani (2006) and Runesson (1999) have taken this focus; such papers are not discussed in any further detail in the thesis.

3.4 Methods in phenomenography

The methodological assumptions of phenomenography are deliberately unclear. Marton and Booth (1997, p.116) state that it has "the object of research as its only defining attribute, and not methods and theories."

Perhaps as a consequence, until recently there has been little guidance in the phenomenography literature to aid a researcher in data collection and analysis. For example a book with the aim of describing phenomenography as a research specialism, the 224–page "Learning and Awareness" by Marton and Booth (1997), devotes fewer than ten pages to data collection and analysis. More recently, Åkerlind (2005c) has compared different authors' phenomenographic methods, and Bowden and Green (2005) have released a methods-based book based on their interpretation of phenomenography. Much of what follows in this section is drawn from these two accounts, together with literature written for qualitative studies more generally.

3.4.1 Data collection

In phenomenography, data must be gathered to investigate others' ways of experiencing a concept, unlike, say phenomenology, where philosophers might primarily investigate their own experiences (Marton and Booth, 1997, p.116). There are only a certain number of ways of collecting such data, and most authors agree that the most common is to interview students (Entwistle, 1997; Fleming, 1986; Green, 2005; Säljö, 1997). The data in this case are records of the interview, audio and video recordings, field notes and and materials used in the interview. Other approaches to data collection include use of questionnaires (Åkerlind, 2005a), and attempts to use more naturalistic situations, such as Lybeck's (1981) account of students learning about the concept of density in physics by observing students' interactions with each other and the teacher in a classroom. In this main study presented by this thesis (see Chapters 5 and 6), data is gathered from task-based interviews, and so this section will only consider the methods of interview-based studies.

In a book chapter reflecting on phenomenographic research, Bowden and Green (2005) noted that before conducting interviews there were many questions that a phenomenographer needs to answer including: Whom to interview?, How many to interview?, What kinds of questions to ask and what comments to make?, When to begin the analysis? Such questions are answered for the main study's task-based interviews in Section 5.2. What follows in this subsection is a brief discussion of what authors have written about the methodology of task-based interviews.

It was noted in the last section that some variation theory studies have used tasks in an interview situation (Asghari, 2004; Watson and Mason, 2004). There are fewer phenomenographic studies that give students a task during interviews (Marton, 1975, is an example of such a study). The lack of studies may be the result of a belief that a task-based interview is seen as overly constrictive when a researcher wishes to explore a student's experience of a concept. In the main study in this thesis, I have found that using an example generation task is a good middle ground; it functions to steer a student's experiences towards the mathematical object under study, but still allows flexibility in possible responses and the resulting discussion.

Similar to all interview research, in task-based interviews there are certain considerations to take into account, such as choosing an appropriate structure of the interview, and which questions to ask (Bryman, 2004). The questions asked to students during a task-based interview can be separated into two types: those that are asked as part of the task itself and those asked when students are asked to reflect on the task. A researcher may choose to ask the reflective questions during the task itself (at the risk of influencing the student's thinking during the task), or after the task phase of an interview (at the risk of a student forgetting their thoughts during the task).

The questions asked as part of the task itself are usually pre-determined and the same for all students, and so a good pilot study will help construct a suitable set of questions (Bowden, 2005, p.19). The pilot study reported in this thesis in Chapter 4 found which questions were answered well, badly, and not at all, and this was taken into account when planning for the main study (see Chapter 5). Reflective questions are usually asked in the form of a semi-structured series Goldin (1997). In such a series of questions, there is a pre-determined framework of themes to be explored but not a strict list of questions to be followed (as would be the case in a structured interview). Goldin (1997, p.53) argues that the chief reasons for such an approach are the flexibility of being able to pursue different approaches depending on what takes place during the interview and reproducibility in terms of the themes addressed. The form of questioning taken in the main study was that of a conversational interview focused around the student's interaction with the example generation task (Patton, 2002).

Another issue associated with interviewing students in a task-based interview is the effect of the interviewer's instruction and prompts on the student's actions (Koichu and Harel, 2007). Authors do not agree if accurate reports will be produced when a student is requested to 'think-aloud'; Ericsson and Simon (1980) claim that such reports will be representative, while other authors such as Fleming (1986) argue that students may not wish to be truthful about their actual thoughts to an interviewer. When using a think-aloud protocol, an interviewer may decide to prompt the student with variations of the question "what are you thinking about?" after a long period of silence (Koichu and Harel, 2007). Such prompts may be interpreted by the student that their last answer was incorrect, or otherwise disturb their thinking. There is not much that can be done to reduce this, although if there is a long silence an interviewer can time their interruption to not coincide with the student immediately finishing a question.

3.4.2 Data analysis

As described above, the data from an interview-based phenomenographic study consists of the records of the interviews. In a task-based interview such as the ones conducted as part of my main study, these records may also include students' answers to the tasks, and records of the comments students made when completing the task. During data analysis, a researcher takes these snapshots of the possible ways to experience the concept, and attempts synthesise the salient features of students' awareness in terms of categories of description, which then can be formed into dimensions of variation. This subsection outlines how phenomenographers do this by coding the interview transcripts for salient features and then combines similar codes to form categories of description, which in turn are compared in dimensions of variation. Specifics on the data analysis procedure of the thesis's main study can be found in Section 5.4; what follows here is a more general account.

Transcription

The considerations when transcribing interviews are similar to those for other non-phenomenographic studies, as outlined by authors such as Ochs (1979) and Poland (2001). Some considerations, such as transcribing the comments of multiple students are not applicable to my research, but considerations such as not accidentally rephrasing sentences and phrasing structures, and the inability of a transcript to completely reflect the social process of a conversation Fleming (1986), will be kept in mind when the transcription protocol of the thesis's main study is outlined in subsection 5.3.3.

Coding

The coding of transcripts in phenomenographic data analysis are similar to those of grounded theory (Richardson, 1999), in that the categories of description emerge 'bottom up' from the data rather than a 'top down' approach where hypotheses are constructed and deduced (Green, 2005). Transcripts are usually focused on both individually and as a whole set (Åkerlind, 2005c); a phenomenographic researcher is trying to compare the different ways students experience the concept so it is important to keep perspective between different students' accounts (Marton and Booth, 1997).

Such a coding scheme has much in common with Glaser and Strauss's (1967) grounded theory approach to qualitative data analysis, where a constant comparative analysis is made between different individuals or groups of people by first labelling data with codes, then grouping those codes into similarly themed concepts, then grouping similar concepts into categories which are then used to generate a theory explaining the subject of the research. After Glaser and Strauss ended their collaboration they disagreed on the best way to ‘do’ grounded theory data analysis (Glaser, 1992; Strauss and Corbin, 1988), in particular if the codes generated from the data will always interact in the same way. Glaser’s interpretation of the grounded theory does not allow the categories which result from the coding to have a predefined general structure:

> The analyst has no idea that “dimensions,” merely one of many theoretical coding families, is, before emergence, the most relevant. (Glaser, 1992, p.46)

The above quotation is in the context of a severe criticism of Strauss and Corbin’s (1988) approach to grounded theory (rather than phenomenography), but I have little doubt that Glaser would hold similar opinions towards phenomenography’s dimensions of variation. So although in practice there is much in common between a grounded theory approach to data analysis and a phenomenographic one, it is not clear that ‘the method of grounded theory’ can be lifted wholesale and used in a phenomenographic study.

Åkerlind (2005c) has reported on the methods used by phenomenographers to decide which parts of the data should be coded:

> Utterances found to be of interest for the question being investigated... are selected and marked. The meaning of an utterance occasionally lies in the utterance itself, but in general the interpretation must be made in relation to the context from which the utterance was taken. (Marton 1986, as cited in Åkerlind, 2005).

> I have in the back of my mind the question ‘What does this tell me about the way the student understands [the phenomenon under study]?’ In other words, what must [the phenomenon] mean to the student if he or she is saying this or that? (Bowden, 1994, as cited in Åkerlind, 2005).

Codes therefore attempt to pick out parts of a transcript which the researcher finds particularly representative of the extract, or that are otherwise salient to the research. Marton and Booth (1997) note that it is particularly important for a researcher to 'bracket' their own views and attempt to read the extract in the same way as the student intended it. Researchers must not treat the whole interview as representative of a single conception, but be open to the possibility that different conceptions may manifest themselves at different times (Bowden, 2005).

There may appear to be a discrepancy between bracketing data and the general method of constant comparison within grounded theory. I believe there to be no such discrepancy because interpretation of an utterance and comparison with other utterances are isolated in phenomenography. This is because when comparing utterances, the researcher assumes the previous interpretation is a possible way for a student to experience the concept, rather than seeking certainty that the student is actually experiencing the concept in the way believed.

What this means for the main study's data analysis (see Section 5.4 for a complete description) is that the initial open-coding will be focused on interpreting individual utterances. After this when forming categories of description by comparing similarly-coded utterances, alternative explanations for the meaning of utterances are bracketed. Once the categories of description and dimensions of variation are formed, then the original meaning of utterances can be reconsidered.

Forming categories and dimensions of variation

Åkerlind (2005c) quotes a passage from Prosser, explaining how a research assistant began to form categories from the data:

> [The research assistant] was asked to read through the whole set of transcripts. . . several times until she felt she was reasonable familiar with them. She was then to try to construct a set of categories which she felt encompassed her perceptions of what the students were trying to say. She then went back over the transcripts, adjusted the categories, and cycled between the categories and the transcripts until she felt she had a reasonably stable set of categories. (Prosser, 1994, p34, cited in Åkerlind, 2005)

From this quotation it is not at all clear how categories of description were generated from the coded transcripts. Such an account tends to suggest that categories were formed for each transcript and then compared between transcripts. Other accounts of category formation report creating pools of similar meaning that are decontextualised from the transcript, with categories forming from these pools of meaning (Åkerlind, 2005c).

It is important to remember therefore that once categories of description are arrived at, one must still take notice that they are based on different parts of different interviews and so are independent from the interview transcripts and interviewees. Svensson (1997, p.170) notes that "what counts as the 'same' conception may be expressed in many linguistically different ways and what counts as different conceptions may be expressed in a very similar language."

Structural links between categories of description are then examined, and the relative sophistication of related categories is compared forming a hierarchical dimension of variation (Green, 2005; Marton and Booth, 1997). In this thesis the dimensions of variation are reported in Chapter 6, and specific details of the data collection and analysis methods are given in Chapter 5.

3.5 Criticisms of phenomenography

3.5.1 Validation and Reliability

Questions as to the validation and reliability of phenomenographic research originate both from outside and within the specialism. As with all qualitative research it is open to questions related to descriptive and interpretative validity, in other words to what extent has the data collection and analysis process resulted in an accurate refection of the students' experiences. As Maxwell (1992) notes:

> Not all possible accounts of some individual, situational, phenomenon, activity, text, institution, or program are equally useful, credible or legitimate. . . . validity is always relative to, and dependent on, some community of inquirers on whose perspective the account is based. (Maxwell, 1992, p.284)

However, when taking a phenomenographic approach towards data analysis the researcher is looking for possible dimensions of variation, rather than a set of dimensions of variation which most accurately describe the interviews the researcher has conducted. This second-order approach means that descriptive and interpretative validity are therefore alleviated to a certain extent (although they can never be ignored; it is not as if a phenomenographic account with fictional data can be in any way valid), but theoretical validity and generalizability are extenuated as a result of this.

Åkerlind (2005c) notes two types of validity checks that are commonly practiced within phenomenography research. *Communicative validity checks* involve checking whether the resulting outcome space resonates with others. This may include the "research community, the individuals interviewed[2], other members of the population represented by the interview sample, and the intended audience for the findings" (Åkerlind, 2005c, p.330). *Pragmatic validity checks* are where the researcher checks if the results are meaningful and useful for their intended audience. This is perhaps the most useful measure of validity; Entwistle writes that

> for researchers in higher education, however, the test is generally not [phenomenography's] theoretical purity, but its value in producing useful insights into teaching and learning. (Entwistle, 1997, p.128)

In this thesis, Chapter 7 presents an inter-coder validation exercise that explores the communicative and pragmatic validity of the main studies' dimensions of variation.

Ensuring reliability in a phenomenographic account is similar to that of any qualitative research. It is important that the interpretation of data has been consistent and of quality. Methods such as *audit trails* can be used, where the researcher identifies the processes of analysis so that the results are consistent with the data (Cohen et al., 2007, p.149). Other techniques include *coder reliability checks*, where two researchers independently code all or a sample of interview transcripts, and *dialogic reliability checks* where agreement between researchers is reached through discussion and mutual critique of the data (Åkerlind, 2005c, p.331).

[2]It is inappropriate however to seek feedback on a utterance-by-utterance level from interviewees, for there is no claim that an utterance represents a category of description within a dimension of representation and so the "interpretation or categorisation of an individual interview cannot be fully understood without a sense of the group of interviews as a whole"(Åkerlind, 2005c, p.331).

3.5.2 The theoretical position taken

This subsection discusses general criticisms with the theoretical position of phenomenography. A general criticism is that the second-order stance taken does not banish ontological and epistemological issues in general, but merely moves them to a different arena. If the first-order methodological criticisms related to the nature of the concept are avoided by concentrating on students' experiences of the concept, then second-order criticisms remain, namely the likely inconsistency between students' experiences of a concept, their reports of these experiences, and interpretations of the reports of these experiences. This argument is summarised by Ericsson and Simon (1980), and Säljö (1997):

> [verbal reports of thinking] cannot be relied on to produce data stemming directly from the subjects' actual sequences of thought processes. . . the variety of inference and memory processes that might be involved in producing the reports make them extremely difficult to interpret or to use as behavioural data (Ericsson and Simon, 1980, p.221)

> It is doubtful if and in what sense the interview data generated in much of the empirical work within this tradition can be assumed to refer to 'ways of experiencing', the core object of research in phenomenography. (Säljö, 1997, p.173)

What a researcher can do is observe and analyse students' accounts of reflections on a way of experiencing a phenomena. The researcher then attempts to bracket their own ideas and conceptions during a period of analysis, eventually forming dimensions of variation from categories of description generated from these accounts. From my literature search, it is more common for phenomenographic accounts to not claim that the individuals from their study are located "on" a particular dimension of variation, but rather that it is possible to be aware of a concept in a way typified by a category of description. In other words, a way of experiencing a concept is something different to the theoretically constructed outcome space.

> When we talk about "a way of experiencing something" we usually do so in terms of individual awareness. When we talk about "categories of description" [i.e. dimensions of variation] we usually do so in terms of qualitatively different ways a phenomenon may appear to people of one kind or another. (Marton and Booth, 1997, p.128)

This does bring into question the validity and generalisability of results; there is always the danger that the eventual categories of description and dimensions of variation are reflections of the researcher's own ideas. This danger, however, is present with every qualitative approach to social research (Bryman, 2004). In Section 5.5, there is a more focussed discussion as to how these methodological issues relate to the main study's data collection and analysis procedures.

3.6 Summary of chapter

Studies which use the methodology of phenomenography aim to describe the variation in how students experience a concept, and ask questions such as which ways to experience a concept make people able to handle it more and less efficiently. To learn, or to improve one's understanding of a concept is then defined to be an an increase in awareness of the possible variation.

Phenomenography maintains two assumptions. First, that research should be approached by considering second-order descriptions (i.e. students' own accounts of their experiences) and second, that a researcher focuses only on aspects of a concept that a student focuses on, bracketing their own values and judgements. A Phenomenographer then attempts to describe the range of possible ways to experience a concept by grouping these descriptions first into categories of description, then similar categories of description from least to most sophisticated in a dimension of variation. The outcome space, which consists of the dimensions of variation, then outlines the different possible ways students may experience the concept.

Chapter 4

Pilot Studies

This chapter presents findings from two exploratory pilot studies conducted early on in my research. These pilot studies were conducted and analysed before the framework of phenomenography, or the research questions of the thesis had been decided upon, and so the chapter's findings are not presented within a phenomenographic framework. The studies do, however, still provide evidence as to how students interact with the example generation of sequences, and so data from these studies will be used when answering the the first research question in Chapter 9.

The chapter is more autobiographical in tone compared with others in the thesis. It is an account describing the development of my thinking, and also of the evolution of an example generation task, a version of which will be used as the basis of the task-based main interview study. It presents some of the answers given by students taking the task, and reflects on their possible concept images and spontaneous conceptions.

In Section 4.1 the first version of the task is described in conjunction with its implementation as a quantitative pilot study. An analysis of the data from this first pilot study follows, with a brief discussion of the outcomes of this study. Section 4.2 then describes a modification of the task into a smaller scale example generation task conducted in semi-structured interviews. A brief analysis of the data from this second pilot study follows.

4.1 First pilot study

Early on, the opportunity arose for me to give an example generation task to a class of students who were beginning a module on introductory Real Analysis. All students enrolled in this module had taken a class on Sequences and Series either one or two years before, and sat an exam on the material. The initial format of the task was suggested to me by my supervisor, who was the lecturer in the analysis module.

The primary aim of the task (and its subsequent analysis) was for me to obtain a large quantity of data to explore the types of answer students give to an example generation task on an area of mathematics they had seen before. The data analysis would be course-grained, allowing me some insight into students' concept images of real sequences, and their example generation capabilities.

4.1.1 About the task

Participants

The first pilot study of the task was given to 101 undergraduates at Loughborough University. 60 were second year single-honours mathematics students, 38 were third year students on a joint maths degree, one student was a third year single-honours student and two students declined to give their details. All but the third year single honours student had taken a course entitled Sequences and Series in the previous year.

The intended content of Sequences and Series included: Key concepts of real numbers, in particular supremum and infimum; the notion of convergence for sequences; algebra of limits; monotone sequences; subsequences; Cauchy sequences; tests for convergence of series; and absolute and conditional convergence. The lecture in which the task took place was the first of the term, but the Analysis module specification and students' module choice forms reminded students that the Sequences and Series module was a prerequisite.

The task

The task had two stages:

Stage 1.

A comprehensive reference sheet was distributed and the students were given five minutes to re-familiarise themselves with the definitions it contained. Students kept this sheet for the duration of the lecture. The reference sheet can be found in Figure 4.1 and some discussion on the presentation of the definitions in Section 4.1.1.

Stage 2.

Next students were given the example generation task sheet, which asked them to give examples of real sequences that satisfy certain combinations of properties, or to say that the given combination of properties was impossible. This task sheet can be found in Figure 4.2. The students had the rest of the lecture to complete this sheet (around 40 mins), and were reminded after fifteen minutes that they could use the definitions if needed and to move on to a different question if they were stuck.

A comment on the presentation of definitions

There was a notable difference in the way definitions were presented in the Sequences and Series course and in the definitions sheet. In the lecture notes from the course, definitions were presented in a "wordy" way:

Definition 3.10 (divergence to infinity). The sequence s_n is said to diverge to $+\infty$, for which we write $s_n \to +\infty$, if, for every positive real number H, there exists real number n_0 such that for all $n \geq n_0$ we have $s_n \geq H$.

But in the definitions sheet the same definition was presented as:

Definition: $(a_n) \to \infty$ (we say "(a_n) tends to infinity" or "(a_n) diverges to infinity") if and only if $\forall C > 0$, $\exists N \in \mathbb{N}$ s.t. $\forall n > N$, $a_n > C$.

The way definitions were presented in the task definition sheet was more formal in style, but the two definitions are equivalent, and the students had seen mathematical quantifiers and set symbols in other first-year modules such as Mathematical Thinking. Furthermore, the lecturer felt it was desirable for the definitions to be introduced in this style for the purposes of the Analysis module, where similar styles of definitions would be the norm.

Review of Sequences

Definitions

The definitions in the following list should be familiar to you from the course *Sequences and Series*.

Remember that a sequence is a list of real numbers

$$a_1, a_2, a_3, a_4, \ldots$$

where (a_n) denotes the whole sequence.

Definition: A sequence (a_n) is said to be *increasing* if and only if $\forall n \in \mathbb{N}$, $a_{n+1} \geq a_n$.

Definition: A sequence (a_n) is said to be *strictly increasing* if and only if $\forall n \in \mathbb{N}$, $a_{n+1} > a_n$.

Definition: A sequence (a_n) is said to be *decreasing* if and only if $\forall n \in \mathbb{N}$, $a_{n+1} \leq a_n$.

Definition: A sequence (a_n) is said to be *strictly decreasing* if and only if $\forall n \in \mathbb{N}$, $a_{n+1} < a_n$.

Definition: A sequence (a_n) is said to be *monotonic* if and only if it is increasing or decreasing.

Definition: A sequence (a_n) is said to be *bounded above* if and only if $\exists u \in \mathbb{R}$ s.t. $\forall n \in \mathbb{N}$, $a_n \leq u$.

Definition: u is said to be an *upper bound* for the sequence (a_n) if and only if $\forall n \in \mathbb{N}$, $a_n \leq u$.

Definition: A sequence (a_n) is said to be *bounded below* if and only if $\exists l \in \mathbb{R}$ s.t. $\forall n \in \mathbb{N}$, $a_n \geq l$.

Definition: l is said to be an *lower bound* for the sequence (a_n) if and only if $\forall n \in \mathbb{N}$, $a_n \leq l$.

Definition: A sequence (a_n) is said to be *bounded* if and only if it is both bounded above and bounded below.

Definition: $(a_n) \to a$ (we say "(a_n) tends to a" or "(a_n) converges to a") if and only if $\forall \epsilon > 0$, $\exists N \in \mathbb{N}$ s.t. $\forall n > N$, $|a_n - a| < \epsilon$.

Definition: A sequence (a_n) *diverges* if and only if it does not converge to any finite limit.

Definition: $(a_n) \to \infty$ (we say "(a_n) tends to infinity" or "(a_n) diverges to infinity") if and only if $\forall C > 0$, $\exists N \in \mathbb{N}$ s.t. $\forall n > N$, $a_n > C$.

FIGURE 4.1: The definition sheet given to students in the first pilot study.

Information about this task

This task **does not form part of any assessment** and will not affect any degree classification awarded. The purpose of this task is to help us better understand the class' knowledge of sequences.

Sequences Review

Give an example of each of the following, **or state that this is impossible**. You may write your sequence in any way you choose (e.g. using a formula or a list of numbers).

Q1. A strictly increasing sequence.

Q2. An increasing sequence that is not strictly increasing.

Q3. A sequence that is both increasing and decreasing.

Q4. A sequence that is bounded below, but not above.

Q5. A sequence that has neither an upper bound nor a lower bound.

Q6. A decreasing sequence that is bounded below.

Q7. A monotonic sequence that is not bounded below.

Q8. A bounded, monotonic sequence.

Q9. A monotonic sequence that has neither a lower bound nor an upper bound.

Q10. A non-monotonic sequence that has neither a lower bound nor an upper bound.

Q11. A sequence that converges to 100.

Q12. A sequence that converges to two different limits.

Q13. A convergent sequence which is not monotonic.

Q14. A sequence that tends to minus infinity.

Q15. A sequence that tends to minus infinity and is not monotonic.

Q16. A strictly increasing sequence that does not tend to infinity.

Q17. A sequence that tends to infinity and is not increasing.

Q18. A sequence that tends to infinity, of which infinitely many terms are 0.

Q19. A divergent, bounded sequence.

Q20. A divergent sequence that is not bounded.

FIGURE 4.2: Questions on the first pilot study's task sheet. The actual sheet had gaps between questions.

Possible answers to the example generation questions

Below are possible answers to each of the questions, with a brief explanation to the types of sequences which satisfy the properties.

Q1. A strictly increasing sequence.
This condition is satisfied if each subsequent term is higher than the last, for instance $(a_n) = n$

Q2. An increasing sequence that is not strictly increasing.
Here, each subsequent term must be greater than or equal to the last, for instance $(a_n) = 1, 1, 2, 3, 4, \ldots$

Q3. A sequence that is both increasing and decreasing.
The only sequences satisfying this question are constant sequences, for instance $(a_n) = 1$

Q4. A sequence that is bounded below, but not above.
Here, sequences such as $(a_n) = n$ *are fine*

Q5. A sequence that has neither an upper bound nor a lower bound.
A typical answer to this question might be $(a_n) = n(-1)^n$

Q6. A decreasing sequence that is bounded below.
The sequence must converge from above, for instance $(a_n) = 1/n$

Q7. A monotonic sequence that is not bounded below.
Here, we can have a sequence that is decreasing and not bounded below such as $(a_n) = -n$

Q8. A bounded, monotonic sequence.
The sequence must either be constant or converge, for instance $(a_n) = 1/n$

Q9. A monotonic sequence that has neither a lower bound nor an upper bound.
The combination of properties requested is impossible

Q10. A non-monotonic sequence that has neither a lower bound nor an upper bound.
Any correct answer to Question 5 will work here

Q11. A sequence that converges to 100.

The constant sequence $(a_n) = 100$ *is fine*

Q12. A sequence that converges to two different limits.

Impossible

Q13. A convergent sequence which is not monotonic.

This can be satisfied by a sequence which oscillates around a number with decreasing magnitude such as $(a_n) = \frac{(-1)^n}{n}$, *or by a sequence such as* $(a_n) = 1, 0, 1, 1, \ldots$

Q14. A sequence that tends to minus infinity.

Here, a sequence such as $(a_n) = -n$ *is fine*

Q15. A sequence that tends to minus infinity and is not monotonic.

A sequence such as $(a_n) = 1, 0, 1, 0, -1, -2, \ldots$ *satisfies this question*

Q16. A strictly increasing sequence that does not tend to infinity.

The sequence must converge, for instance $(a_n) = -1/n$

Q17. A sequence that tends to infinity and is not increasing.

Here, there must be at least one pair of terms that decrease, $(a_n) = 1, 0, 1, 2, 3, \ldots$ *works*

Q18. A sequence that tends to infinity, of which infinitely many terms are 0.

This is impossible

Q19. A divergent, bounded sequence.

Here the sequence must not converge, so $(a_n) = (-1)^n$ *is fine*

Q20. A divergent sequence that is not bounded.

Any answer to Question 5 also satisfies this question

4.1.2 Data analysis

In keeping with the exploratory nature of the pilot study I began by reading through the task sheets returned by the students. There was a large volume of data; 101 students had answered 20 questions and of the 2,020 possible responses only 30% of questions were left blank (for a more detailed breakdown per question see Figure 4.3).

Difficulty of questions

The first analysis I made was an attempt to gauge which questions were more difficult than the others. If I were attempting to determine relative difficulty of the questions now I would have used a Rasch Analysis (see Chapter 8), but with this dataset I just considered the proportion of correct answers.

I went through each sheet marking the tasks as if they were a class assignment, marking them correct (**C**), incorrect (**I**) or not attempted/blank (**B**). Figure 4.3 gives a breakdown of this coding for each question on the task. There was a degree of interpretation here; minor slips of notation were overlooked, but the scripts were generally marked harshly (for a more detailed discussion of the marking of scripts see Section 4.1.4).

The difficulty of a question was calculated by considering the number of students that answered a question incorrectly. In such a calculation, it is unclear if blank answers should be counted as incorrect; if a student leaves a question blank perhaps they attempted it and couldn't give an answer, or perhaps they did not have sufficient time to complete the question. There are two metrics that take these perspectives into account, d_1 and d_2.

The first metric, $d_1(i)$, gives the proportion of students who did not correctly answer Question i, which assumes that a blank answer represents a student not being able to answer a question:

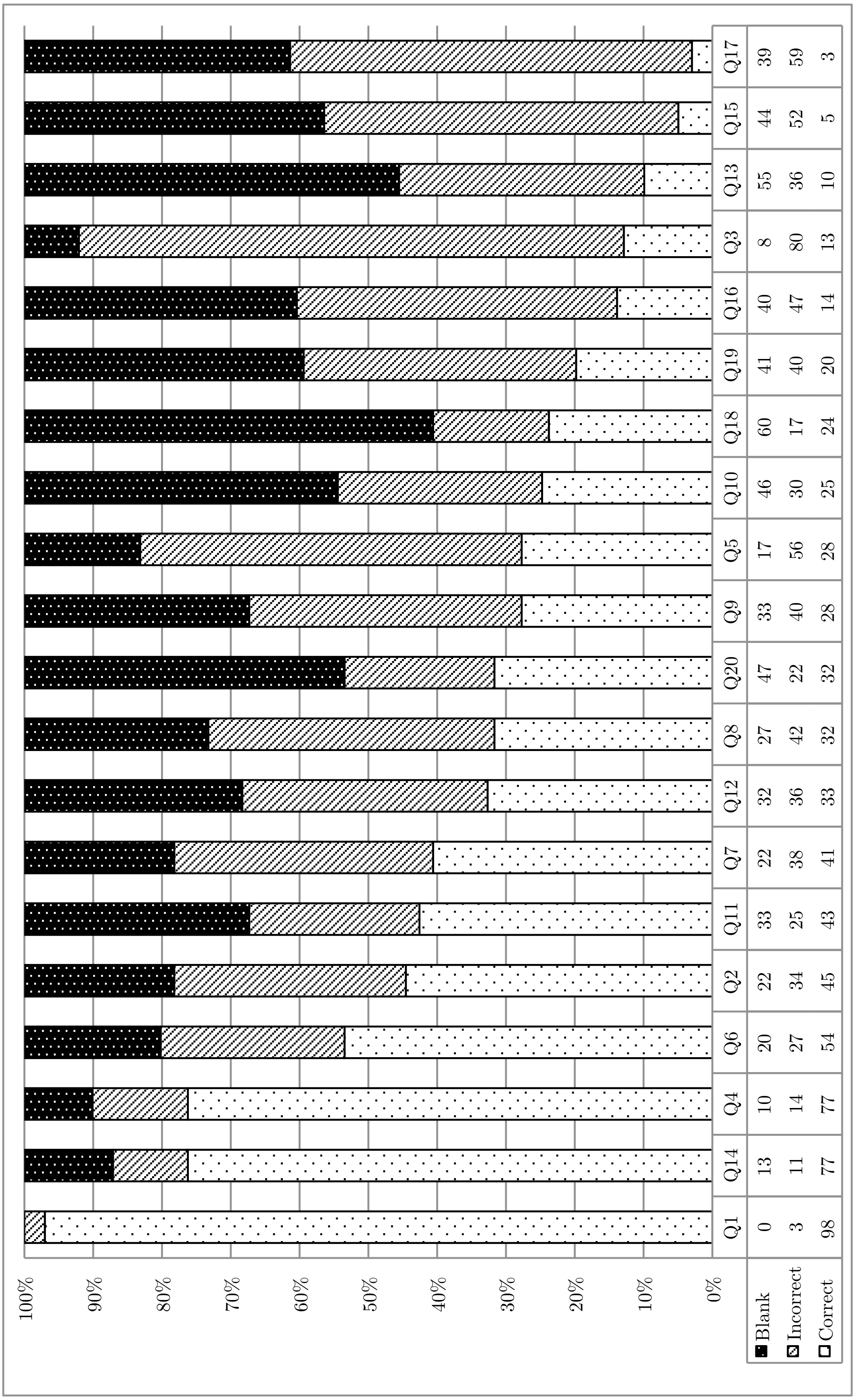
$$d_1(i) = \frac{\#\mathbf{I}_i + \#\mathbf{B}_i}{101}$$

where I_i are those students incorrectly answering Question i and B_i are those students leaving Question i blank.

The second metric, $d_2(i)$, gives the proportion of students answering Question i incorrectly from those who attempted the question. This metric assumes that blank answers are due to a student not having time to attempt the question:

$$d_2(i) = \frac{\#\mathbf{I}_i}{\#\mathbf{I}_i + \#\mathbf{C}_i}$$

where I_i are those students incorrectly answering Question i and C_i are those students correctly answering Question i.



	Q1	Q14	Q4	Q6	Q2	Q11	Q7	Q12	Q8	Q20	Q9	Q5	Q10	Q18	Q19	Q16	Q3	Q13	Q15	Q17
Blank	0	13	10	20	22	33	22	32	27	47	33	17	46	60	41	40	8	55	44	39
Incorrect	3	11	14	27	34	25	38	36	42	22	40	56	30	17	40	47	80	36	52	59
Correct	98	77	77	54	45	43	41	33	32	32	28	28	25	24	20	14	13	10	5	3

Figure 4.3: Breakdown of answers in the initial coding of correct, incorrect and blank answers.

TABLE 4.1: Difficulty metrics for each question. Higher numbers correspond to more difficult questions.

Qn	Difficulty $\frac{\#\mathbf{I}+\#\mathbf{B}}{101}$	Difficulty $\frac{\#\mathbf{I}}{\#\mathbf{I}+\#\mathbf{C}}$
1	0.03	0.03
2	0.55	0.43
3	0.87	0.86
4	0.24	0.15
5	0.72	0.67
6	0.47	0.33
7	0.59	0.48
8	0.68	0.57
9	0.72	0.59
10	0.75	0.55
11	0.57	0.37
12	0.67	0.52
13	0.90	0.78
14	0.24	0.13
15	0.95	0.91
16	0.86	0.77
17	0.97	0.95
18	0.76	0.41
19	0.80	0.67
20	0.68	0.41

Each metric gives the difficulty of a question as a value $0 \leq d \leq 1$, with more difficult questions having a difficulty value closer to 1. The metrics were calculated for each question and are presented in Table 4.1. Unsurprisingly there is a significant positive correlation between these two metrics, $r = .942$, $p < .01$, which indicates that the choice of how to treat a blank answer is arbitrary. For the remainder of the chapter I shall use the first metric (answers that were not correct as a proportion of all answers).

Based on this metric, the most difficult questions (with first metric difficulty in brackets) were, in order:

Q17. A sequence that tends to infinity and is not increasing ($d_1(17) = 0.97$)

Q15. A sequence that tends to minus infinity and is not monotonic ($d_1(15) = 0.95$)

Q13. A convergent sequence which is not monotonic ($d_1(13) = 0.90$)

Q3. A sequence that is both increasing and decreasing ($d_1(3) = 0.87$)

Q16. A strictly increasing sequence that does not tend to infinity ($d_1(16) = 0.86$)

It can immediately be seen that four of these questions (17, 15, 13, and 16) involve combining definitions which focus on different aspects of a sequence. Some properties, such as *increasing*, *decreasing*, and *monotonic* compare terms pairwise across the entire sequence. Other properties, such as *tending to infinity*, focus on the long term behaviour of the sequence. It is not obvious why these particular combinations may cause more difficulties than other combinations which mix local and longer term properties, such as Q7. A monotonic sequence that is not bounded below. The issue of combining local and longer term sequence properties successfully is discussed in more detail in the main study, Section 6.1.

Analysis of individual questions

I then went through each question looking at the students' responses in an attempt to explore the incorrect responses given to the most difficult questions. I made a log of the different answers students gave, and wrote brief reflective essays on each question on the task.

During this process of reflection, I decided that within the incorrect answers there were two qualitatively different types of answer. There were answers which were sequences that did not possess all the properties required by the question, and those answers which weren't sequences. These non-sequence responses often had properties which were similar to the properties specified, such as an answer which gave an unbounded set when an unbounded sequence was requested.

I then extended my initial categorisation of answers (**C**/**I**/**B**) into four categories by partitioning the incorrect answers into the two types I had observed. The second coding scheme is given in Table 4.2. This coding scheme is used for the remainder of this thesis.

As with the previous categorisation, there was a degree of interpretation as to whether an answer is actually intended to be a function $f : \mathbb{N} \to \mathbb{R}$ or not. The answers at this stage which were most difficult to code were examples written as functions $f : X \to \mathbb{R}$ where the set X was not stated, or otherwise ambiguous. Section 4.1.4 discusses the methodological issues of this ambiguity.

TABLE 4.2: Categorisation scheme for responses to the first pilot study.

Code	Description
C	*Correct* Answers which were correct.
B	*Blank* No answer was given.
IS	*Incorrect sequence* An incorrect answer given in the form $f : \mathbb{N} \to \mathbb{R}$, or one that could reasonably be interpreted as such.
INS	*Incorrect non-sequence* Other incorrect answers

Analysis focusing on IS and INS sequences

Only questions which had a high proportion of incorrect responses were recoded to the (**C**/**B**/**IS**/**INS**) categorisation. As explained in Section 4.1.2, my focus at this stage was towards the type of answers students gave to the more difficult questions, and so I recoded the five questions with the highest proportion of incorrect answers relative to the metric d_1, Questions 17, 15, 13, 3 and 16.

I also looked in more detail at a further question that, based on my reflective log on each question, had many interesting **INS** responses.

Q9. A monotonic sequence that has neither a lower nor an upper bound ($d_1(9) = 0.72$);

Within these six questions, three had a large number of incorrect responses because the majority of students attempting the question stated that the combination of properties requested were impossible. These were questions 15, 16, and 17. I shall first discuss these three questions, and then consider the ones remaining.

Analysis of Questions 15, 16, and 17

Questions 15, 16, and 17 were:

Q15. A sequence that tends to minus infinity and is not monotonic.

Q16. A strictly increasing sequence that does not tend to infinity.

Q17. A sequence that tends to infinity and is not increasing.

TABLE 4.3: Breakdown of answer types for Questions 15, 16 and 17.

	C	**B**	**IS**	**INS**	"Impossible"
Question 15	5	44	6	14	32
Question 16	14	40	9	12	26
Question 17	3	39	14	5	40

A breakdown of the answer types for each of these questions is given in Table 4.3. These three questions could be considered to be targeting standard misconceptions within sequences, such as the idea that a strictly increasing sequence must tend infinity. A task such as this one does not provide any evidence why many students thought Questions 15 and 16 were impossible, and also why many students did not attempt these questions, but there are similar features between the three questions. I briefly comment on these features here, leaving further discussion to later chapters).

- The general comments made at the end of Section 4.1.2 regarding different definitions focusing on different aspects of the sequence also resonates here. Properties such as *increasing* and *monotonic* involve comparing sequences term-by-term, whereas a property such as *tending to infinity* relies on considering the long-term behaviour of a sequence. If a student tries to consider the term-by-term properties as long-term then perhaps the properties become impossible. For instance, a long-term version of *increasing* may mean 'going up' to some extent. With such a meaning, the combination of properties requested by Questions 15, 16, and 17 will appear to be impossible.

- All three questions rely on negating a property, and research suggests that students find this particularly difficult (Antonini, 2001; Evans and Handley, 1999). If the students in question are not formally negating the definitions, it may well be the case that they are instead taking the opposite of an everyday meaning of a property. For instance, the opposite of the everyday meaning of increasing may be considered to be decreasing, and it is correct that there is no decreasing sequence which tends to infinity.

Analysis of Questions 3, 5, and 9

The remaining three questions had varying proportions of **IS** and **INS** answers (see Table 4.4 for a breakdown), and I now discuss the responses to each of these questions in more detail.

Questions 3, 5, and 9 were:

Q3. A sequence that is both increasing and decreasing.

Q5. A sequence that has neither an upper bound nor a lower bound.

Q9. A monotonic sequence that has neither a lower bound nor an upper bound.

Table 4.4: Breakdown of answer types for Questions 3, 5, and 9.

	C	**B**	**IS**	**INS**	"Impossible"
Question 3	13	8	52	7	21
Question 5	28	17	24	31	1
Question 9	28	33	21	19	as **C**

In Questions 3, 5, and 9 a large proportion of the responses given were not correct. Many student gave sequences that did not have the required properties and, on the whole, these **IS** responses were not wholly unreasonable. Many were sequences that did not have all of the required properties but did have some of these. For instance two students gave sequence

$$(a_n) = (-2)^n$$

as an answer to Question 9. This sequence has neither a lower nor an upper bound, though it is not monotonic. Answers of this type are perhaps to be expected as students may attempt to give an appropriate answer but be unable to find one that satisfies all the requirements simultaneously.

Less expected was the large number of responses that were not infinite sequences of real numbers, though typically these **INS** responses did not appear to be arbitrary. For instance, as a response to Question 5, eight students gave the answer

$$(-\infty, \infty)$$

and five students gave the double-sided sequence

$$-\infty, \ldots, -1, 0, 1, \ldots, \infty$$

These answers, together with sixteen given in standard set notation ($\mathbb{R}$, $\mathbb{Z}$, $n^3\ \forall n \in \mathbb{R}$, etc.) are indeed unbounded, and a generous interpretation might allow that answers such as may be thought of as monotonic in some sense also. It seems that in a typical **INS** response, the definition of a sequence is violated (specifically changing its domain) in order to fit the other properties, rather than violating the other properties while keeping the primary one of being a sequence. This result echoes that of Dahlberg and Housman (1997), who reported that that some students modify or reinterpret the meaning of a concept if they are unable find examples to satisfy it.

With hindsight, this large number of **INS** responses is perhaps not so surprising. Students will typically see a formal definition of a sequence early in their Sequences and Series course, but this definition may not be reiterated or used explicitly in subsequent reasoning. So, although it is alarming that students may reach the end of such a course without having internalised the idea that 'sequence' means 'infinite real one-sided sequence' in this context, it is understandable if they are unable to maintain sufficient control to exclude examples of other, related objects (series, sets etc.).

4.1.3 Ethical considerations in the first pilot study

When designing the first pilot study ethical issues were first addressed by completion of the Loughborough University's ethical clearance checklist. No facet of the pilot study required clearance from the ethical advisory committee. Possible issues that were identified as a result of a discussion with my supervisor were that students are used to completing short tasks in lectures as part of their assessed work, and so it may appear to students as if this task was also assessed. Consequently it was heavily stressed that the task was not assessed work, and that the results would only be used by us to gain an understanding of the types of responses they gave to this type of task.

A further ethical consideration with the first pilot study is the more general one of using the students' lecture time wholly or partly for research purposes. While research that gathers data during teaching time can face issues of ethics, in the first pilot study the task was designed to correspond with the learning objectives of the course, it formed a recap of material which had been previously studied, and it introduced them to the definition-focused approach of their future lectures. Furthermore, students' responses to the task informed the lecturer of their prior understanding, allowing more time to be spent recapping last year's material.

Other ethical considerations included the amount of time given to students and the format of the task sheet. I felt that both these parts of the design were appropriate for the audience in question, and the lecturer (who was involved in the design of the definition and task sheets) treated the task as she would any other in-lecture activity, with the expectation that the students would engage with it and with the intention of using it as the basis for later discussion. Students were therefore not told that they could decline to complete the task, although any student that did not do so had no attention brought to them.

4.1.4 Methodological issues in the first pilot study

Before summarising this study and its implications for the main study, I briefly acknowledge and discuss some difficulties in gathering and interpreting this type of data.

Timing of task

Almost all (99%) of the students had completed the module Sequences and Series three months before the task, and the lecture in which the task was given was the first one after the Summer vacation. So although the relevant definitions were provided, it may be that a larger proportion of correct results would have been seen had the task been given immediately following Sequences and Series module. This concern is addressed to a certain extent in both the second pilot study and the main study. The participants in these latter studies had recently been taught the material. This (first) pilot study perhaps provides some indication, however, of the retention of the material over time.

The appropriateness of the definition sheet

There is the issue of how many definitions to provide on the definition sheet, in what form the definitions should be, how much detail to give on the sheet, and even if it is appropriate to have a definitions sheet at all. For the purposes of this pilot study it seemed only fair to give each student a collection of all the relevant definitions, as students who participated in the pilot study were not given a chance to prepare. Furthermore, the high proportion of incorrect answers with the definition sheet available is a striking result, and one would expect that students would perform worse on the task without the information available. There is also the question of whether the number of **INS** would be lower if the definition sheet emphasised the definition of a sequence more explicitly. I believe that the explanation given on the definition sheet (see Figure 4.1) was sufficient, however.

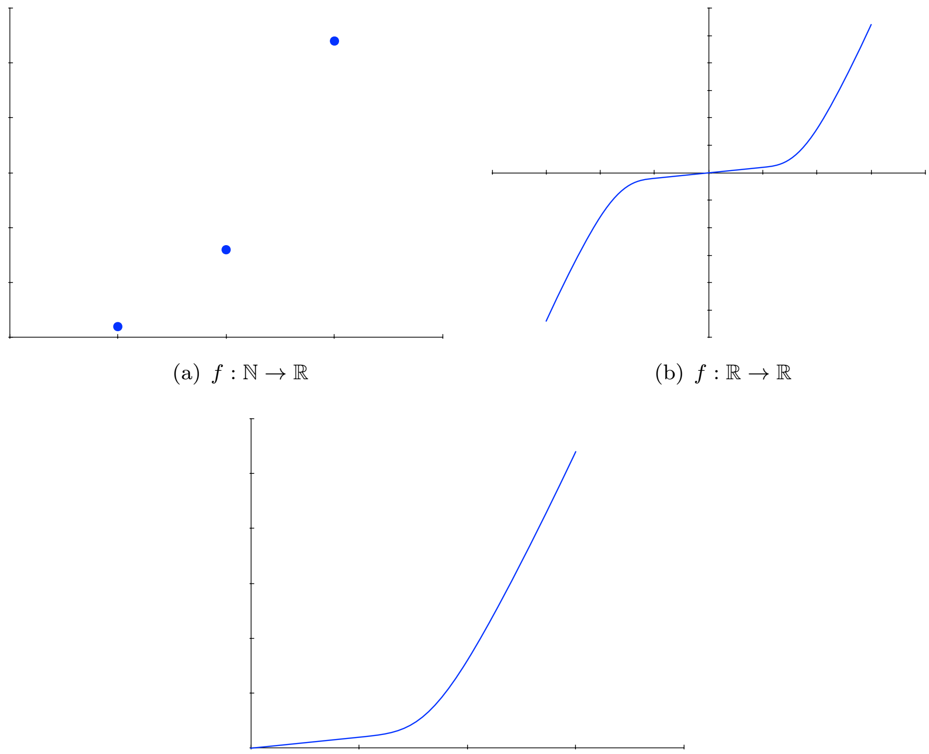
The coding of answers

As noted in Section 4.1.2, there is a problem regarding inferring a students' meaning from the examples given in tasks such as this. In Question 5 examples such as $(a_n) = n^3$ can be found in both category **IS** and **INS**, depending on how explicit an answer was given, but it is impossible to know which domain was intended. For instance, consider the following answers given by three students to Question 5: A sequence which is not bounded above or below.

1. $(a_n) = n^3$, $n \in \mathbb{N}$
2. $(a_n) = n^3$, $n \in \mathbb{R}$
3. $(a_n) = n^3$
4. $(a_n) = n^3$ with a sketch of a function (see Figure 4.4 for three possible sketches)

The first answer will have been coded as **IS**; the domain is specified as the natural numbers and so the object written is a sequence. The second answer's domain is the real line, so the object is certainly not a sequence, and will have been coded **INS**. The third answer does not specify the domain, and so the benefit of doubt is given as it can be reasonably interpreted as a sequence, although it is perfectly possible that the second answer too was intended to be the real-valued function, but for the purposes of the categorisation the benefit of the doubt was given towards **IS** categorisation where there was ambiguity.

My conviction of the categorisation of the fourth answer depended on the accompanying graph. Three examples of graphs of functions are given in Figure 4.4. If a student gave a "dotty" graph, such as the one presented in Figure 4.4(a), then the domain is clearly $\mathbb{N}$ and so the answer would be categorised **IS**. At the other extreme, if a student included a graph of the continuous function over the whole of $\mathbb{R}$, such as in Figure 4.4(c) then the answer is certainly **INS**. The answer most difficult to categorise would be with a graph of the positive region only, but still a continuous function rather than a "dotty" sequence. This type of answer would have still been categorised **INS**, because technically the function drawn has domain $\mathbb{R}^+$, but with less conviction on my part because anecdotal evidence suggests that students often make this type of sketch even when they are otherwise dealing with the function as a sequence.



(a) $f : \mathbb{N} \to \mathbb{R}$

(b) $f : \mathbb{R} \to \mathbb{R}$

(c) $f : \mathbb{R}^+ \to \mathbb{R}$

FIGURE 4.4: Possible sketches to accompany the answer $a_n = n^3$

There was generally less ambiguity with sequences given as a list of numbers, for instance double-sided sequences such as

$$-\infty, \ldots, -1, 0, 1, \ldots, \infty$$

were considered to be a function $f : \mathbb{Z} \to \mathbb{R}$, and so was also marked **INS**.

It is not obvious how a researcher should address this problem without an interview component in the study—as noted in the previous paragraph, had the definition and task sheets been more specific in reminding the students that a sequence should have $\mathbb{N}$ as its domain, some of the interesting responses may have been lost.

The research questions addressed by the pilot study

Finally there is a more theoretically-based methodological issue which was highlighted by an anonymous reviewer to a PME short oral paper I submitted (Edwards and Alcock, 2008). Did the pilot study aim to (a) use example generation as a lens to explore students' concept images of sequences, or (b) explore how example generation can provide information about students' concept images more generally, using sequences as as the object in focus? This was a very valid criticism, and due to the exploratory nature of the analysis the answer probably is "a little of both." In my main study I have reflected on which parts of data collection and analysis have supported which theoretical questions.

4.1.5 Conclusions from the first pilot study

In this section I have presented the results of an example generation task given to second and third year university students. Data and conclusions from this study have been published in conference proceedings (Edwards and Alcock, 2008), but not in the form presented by this thesis.

Although the results should be treated with caution for the reasons stated in Section 4.1.4, it is clear that many students have problems dealing with even the simple sequence properties of monotonicity and boundedness. A large proportion of students gave responses that indicated spontaneous conceptions at odds with the formal definitions; only

13% could combine the definitions increasing and decreasing to give a sequence satisfying both, despite 98% of students correctly giving an example of a strictly increasing sequence. Often attempts to generate a sequence that satisfied certain properties apparently led to a failure to control for the requirement that the answer be a sequence.

This pilot study succeeded in its aim of gathering a relatively large amount of data in order to provide some initial information about students' knowledge and understanding of sequence properties. Together with the second case study, this exploratory analysis of example generation data highlighted issues to explore further in the design and implementation of my main study. The key issues are summarised in Section 4.2.5, after the second pilot study in Section 4.2.

4.2 Second pilot study

The second pilot study was a small interview study whose aim was to explore in more detail the types of interesting **IS** and **INS** responses I observed in the first pilot study. The data analysis was focused on examining situations where such responses were given.

This section begins with a description of the choice of questions used in the second pilot study task and an outline of the format each semi-structured interview took. Finally there is a description of the analysis process and some general observations on the type of answers seen. In the next section I draw together findings from both pilot studies in anticipation for the main study.

4.2.1 About the interviews

Participants

The second pilot study took place in February 2008, with a group of students who were taking a course on Sequences and Series. This module is the first time students at Loughborough University are taught sequences and was the compulsory module taken by the first pilot study's participants.

I canvassed for volunteers to attend an hour-long interview where they were told they would be asked to complete a short task, receive feedback on their mathematics and also be given a sigma-branded USB pen drive as compensation for their time. Seven students volunteered and six attended interviews.

Reducing the number of questions

Each interview was scheduled to last for an hour, with a discussion section at the beginning and a more focused discussion section at the end. This meant the number of questions had to be reduced from the first pilot task, as I did not wish for the students to feel rushed and there was a danger of this had I kept all twenty questions. In order to decide which questions to include in this second version of the task several factors were taken into account.

- Although it was expected that the most interesting responses would emerge from more difficult questions which combined several sequence properties, I did not want to make the task appear to be an unsettling list of 'trick questions'. I therefore included some of the more straightforward questions from the first pilot. An unexpected consequence of including such questions was that they served as an informal benchmark for a student's comfort-level with sequences.
- The first pilot study had indicated that **IS** and **INS** responses were commonly given to some questions, and so questions were included with the expectation of yielding these types of responses. Questions 3 and 5 were included directly from the first pilot study, but Question 9 was modified to reduce the number of properties under consideration (see next bullet).
- Questions were limited to combining at most two properties. In the pilot study those questions with three clauses were poorly attempted and poorly answered.
- I felt it better to group similar questions together; this allowed me to explore in more depth a person's concept image of one property, such as boundedness. It also allowed me to target consistencies and inconsistencies within a group of questions in the post-task discussion period.

The task was structured into three themes: *increasing and decreasing*, *boundedness*, and *dealing with infinity*, with three or four questions in each theme. Figure 4.5 lists the questions on the second pilot study's tasksheet, together with my rationale for including them. The themes were not stated on the sheet.

Each theme began with an "easy" question, and progressed to more difficult ones as indicated by (a) the type of responses I noted in the first pilot study, (b) the metrics in Table 4.1, and (c) my refections on what answering the question entailed. Two questions were included that were not trialled on the pilot study: Question 4 which asked for a sequence which was neither increasing nor decreasing, and Question 5, which asked for a sequence with no upper bound.

Planning the style of the interviews

I structured the interview into four parts. First was an introductory period. Students were welcomed, and the purpose of the interview was explained to them (i.e. that I was interested in how they were finding the transition to university and that the task was designed to see how they were finding the Sequences and Series module). They were also told that I would not be reporting back to their lecturer any information that could personally identify the students. This stage of the interview was very brief.

During the second stage of the interview I asked each student how they found the transition to university. This part of the interview was designed not only so I could find out about their background and how they were finding university (data that I do not comment upon here), but also to establish a rapport with the students and put them at ease. This stage of the interview lasted between five and ten minutes.

In the third stage of the interview, the definition sheet was presented to the students, and they were given the task sheet. I explained that I did not want to influence the way they think during the task, so I would remain silent throughout, but that I was happy to go through their answers afterwards. They were also encouraged to 'think aloud' where possible. Each student was given as long as they needed to complete the task sheet.

Finally, each student was given feedback on their performance, in the form of a conversation focused around their answers. This part of the interview was semi-structured (Ginsburg, 1981) around the following themes:

- How did you find the task?

Page 1 — Increasing and Decreasing

Q1. A strictly increasing sequence
Q1 on pilot study, difficulty 0.03

Q2. An increasing sequence that is not strictly increasing
Q2 on pilot study, difficulty 0.55

Q3. A sequence that is both increasing and decreasing
Q3 on pilot study, difficulty 0.87

Q4. A sequence that is neither increasing nor decreasing
A new question, related to the previous question

Page 2 — Boundedness

5. A sequence that has no upper bound
A new question, intended to be an easy introduction to page 2

6. A sequence that has neither an upper bound nor a lower bound
Q5 on pilot study, difficulty 0.72

7. A bounded, monotonic sequence
Q8 on pilot study, difficulty 0.68

Page 3 — Dealing with Infinity

8. A sequence that tends to infinity
Similar to Q14 on pilot study, difficulty 0.24

9. A sequence that tends to infinity that is not increasing
Q17 on pilot study, difficulty 0.97

10. A sequence that tends to infinity that is not bounded below
A new question based on Q9 from the pilot study, but with fewer properties

11. A strictly increasing sequence that does not tend to infinity
Q16 on pilot study, difficulty 0.86

FIGURE 4.5: Questions asked in the second pilot study task, with comments in italics. Difficulties are taken from the first metric given in Table 4.1 (higher numbers correspond to more difficult questions).

- Were any questions easier/harder than others?
- Questions focused on their answers, especially any **INS** ones
- Provide assistance to any questions they were stuck / incorrect on

4.2.2 Data analysis

The audio data from each interviews was transcribed and the task sheets examined for instances of **IS** and **INS** responses. Relevant extracts of transcripts were then examined for comments on the answers given. Unlike in the first pilot study, it was practical to compare responses both across questions and participants. A more general analysis focusing on every response to the task in conjunction with the the entire transcribed data was not completed due to time constraints.

Outcomes from the second pilot study

The students who were interviewed in the second pilot study struggled with the task. 36% of questions were left blank (compared to 30% in the first pilot), and only 29% of questions were answered correctly. The answers the students gave can be found in Table 4.5. In this table each response has been coded in relation to the coding system which was given in Table 4.2. Some student gave multiple answers, or went back later giving another answer, and the Table 4.5 reflects this by including both answers where appropriate.

As noted earlier, the aim of the interviews was to explore further the thought processes of students who produced **IS** and **INS** responses. As with the first pilot study, many of the answers given to the task were incorrect. Bearing in mind some students gave multiple answers, I saw fewer **IS** responses (6) than **INS** ones (17).

Within the **INS** responses it was notable that there were both students who tended to give **INS** responses, such as Student 2, and also questions which tended to be answered with **INS** responses, such as Question 6. The remainder of this sections looks at Student 2 and Question 6 in more detail.

	Student 1		Student 2		Student 3		Student 4		Student 5		Student 6	
Q1	$2, 4, 8, 16$ $a_n = 2^n$	C C	$a, a+1, a+2\dots$	INS	$1, 2, 3, 4\dots$	C	$1, 2, 3$	C		B	$1\ 2\ 3\ 4\ 5\dots$	C
Q2	$1, 2, 2, 4, 8$ $a_n = n(n-1)$	C IS	$a, a, a+1, a+1$	INS	$1, 1, 2, 4, 3$	IS	x^2, x^3, x^4 $-4^2, -4^3, -4^4$	INS IS	could be inc or decreasing	INS	$1\ 2\ 2\ 3\ 4\dots$	C
Q3	Impossible	I		B		B		B	Impossible	I	—	B
Q4	$(-1)^n = s_n$	C	*Graph drawn*	B	$1, 1, 1, 1$	IS		B	$-1, 1, -1\dots$	C		B
Q5	$2^n = a_n$	C	$s_n \to \infty$ $[0, \infty]$	INS INS	$[1, \infty)$	INS	$1, 2, 4, 16\dots, \infty$	C	$1, \dots, \infty$	C		B
Q6	$(-2)^n$	C	$-\infty, \infty$	INS	$(-\infty, \infty)$ $-2, -1, -0, 1, 2$	INS INS		B		B	$(-\infty, \infty)$	INS
Q7	$\frac{1}{n} = a_n$	C	$[1, 2\dots, 5]$	INS	—	B		B	*?*	B		B
Q8	$2^n = a_n$	C	$s_n = n$	C		B	$1, 2, 3, 4\dots,$	C		B	$1, 2, 3, 4, 5, \dots$	C
Q9	Impossible	I		B	—	B	Impossible	I	Impossible	I		B
Q10	$(-2)^n = a_n$	IS	$(0, \infty)$	INS	*Collection of Objects*	INS		B		B		B
Q11	$\frac{1}{n} + \frac{1}{n-1} = a_n$	IS	$s_n \to 1$	INS	$1, 1\frac{1}{4}, 1\frac{1}{\frac{1}{2}}, 1\frac{3}{4}\dots$	C*	$1, 2, 3, 4$	INS*		B	$1, 2, 3, 4, 5.$	INS

TABLE 4.5: Responses to each task question, with appropriate classification from the taxonomy given in Table 4.2. All but *italic* statements are reproduced verbatim. * Starred classifications are judgement calls based on the particular circumstances of the answer.

A student who gave many INS solutions

Although there was a very limited sample of people who were interviewed it was clear that certain students who sat the task had a poor control over the way in which sequences can be represented formally. Student 2 generated examples which were of many different mathematical types, including

Sequences with undefined variables, such as "$a, a+1, a+2$"

In the answer to Question 1 and 2, the example generated was correct, but given in a general form rather than a specific example. These answers were classified as **INS**, but they are clearly correct to a certain extent.

Intervals of the real line, such as "$[0, \infty]$"

Examples for Questions 5, 6, and 10 were each given as a connected interval of the real line. The examples chosen are to some extent correct in regards to the requested properties, for instance as an example of a sequence with neither an upper nor a lower bound, the interval $(-\infty, \infty)$ was given.

A finite list of numbers, such as "$[1, 2 \ldots, 5]$"

The answer to Question 7 was given as a finite list of numbers, and so is bounded and monotonic, but not a sequence.

A further property the sequence may have, such as "$s_n \to 1$"

In two questions the student drew a conclusion about the properties the sequence must or may have, but did not give an example. As soon as these statements were written the student immediately moved on to the next question.

After the student finished the task, I discussed the different types of answer with the student. For example:

> **Interviewer:** You've sort of written them in three different ways throughout the paper.
>
> **Student 2:** Yeah.
>
> **Interviewer:** And can I ask why? If that's OK?
>
> **Student 2:** It just depended on what I was thinking really. To me when you get to infinity you can't, well I suppose I could do dot-dot-dot. I don't

know. I don't know why I did it like that. It just depends on how I think of it in my head at the time.

Interviewer: So you would say that the ones that are written, say with the brackets, you somehow thought of those differently perhaps to those two?

Student 2: No, I reckon I could probably write that as a bracket one. And I could write all these as brackets, I could probably write them all in each of the other ways.

Here 'bracket ones' referred to the examples the student gave in terms of intervals of the real line. For this student, a sequence could be represented as an interval of the real line, or both other representations such as including a list of numbers. Perhaps this student was confusing the values of a sequence and the range of possible values of the sequence, in other words the set $\{a_n : n \in \mathbb{N}\} \subset \mathbb{R}$ and the interval $(\inf\{a_n : n \in \mathbb{N}\}, \sup\{a_n : n \in \mathbb{N}\}) \subset \mathbb{R}$.

A question in which many students gave INS responses

As well as students who predominately gave **INS** responses, there were also questions which seemed to evoke such responses from students who in other questions gave **C** or **IS** responses. Two such questions were:

Q6: A sequence that has neither an upper bound nor a lower bound
Three of the six students gave the interval $(-\infty, \infty)$ for this question. Of course, this is a mathematical object which does not have an upper nor a lower bound. Looking at the students lecture notes, it seems the topic of boundedness was covered in some depth in conjunction with sets early on in the course, and this may have some bearing on the answers seen in the task.

Q11: A strictly increasing sequence that does not tend to infinity
Two of the students gave a finite sequence, which again is understandable as if such objects were sequences also then this example would satisfy the requested properties.

4.2.3 Ethical considerations in the second pilot study

When designing the second pilot study ethical issues were first considered by completion of the Loughborough University's ethical clearance checklist (the second pilot study did not required clearance from the ethical advisory committee). I also attended a short training session focused on ethics in research held by my department.

I recognise that being asked to 'do maths' whilst being audio-recorded and observed by a stranger might be intimidating for some students, and so the early stages of the interview were designed to have little to do with the task and were more focused on chatting to the student to put them more at ease. The semi-structured nature of the final stage of the interview was also designed to adapt for the way the student had progressed with the task. As I shall mention in more detail in the next subsection, the students struggled with the task, and in some cases I felt it more appropriate to guide the students through the questions than to let them become unsettled, feeling that they could not answer any of the questions on the task.

4.2.4 Methodological issues in the second pilot study

As reported earlier, the students who were interviewed in the second pilot study struggled with the task: 36% of questions were left blank and only 29% of questions were answered correctly. This is perhaps less surprising when one considers that the second pilot study focused on questions which yielded a large proportion of **IS** and **INS** responses in the first pilot, but even the easier 'warm-up' questions (1, 5 and 8) were answered by two-thirds of students at most. Answers left blank in an interview-based study are, in general, less of a concern than in a class-based task because they (a) allow students do some thinking about the question which might be verbally self reported at the time, and (b) there was also the opportunity for me to discuss questions left blank with the student during the discussion period. These were not possibilities in the first quantitative task.

Most students appeared to be comfortable making clear that they were struggling during the task. The following excerpt illustrates this, taken from when Student 5 was attempting the third set of questions:

> *Q8. A sequence that tends to infinity*
>
> **Student 5:** I don't really know how to, I can't read things off. I don't know how to write like you say a sequence that tends to infinity.
>
> *Q9. A sequence that tends to infinity that is not increasing*
>
> **Student 5:** I don't know what to say.
>
> *Q10. A sequence that tends to infinity that is not bounded below*
>
> **Student 5:** I don't know what would you want for a sequence that tends to infinity, not bounded below. I don't know what sort of thing would be bounded below, like how you'd represent that.

Later in the interview, we were discussing the difference between this type of task and coursework more generally:

> **Student 5:** Yes. It's harder. Probably because for the courseworks we can refer to the examples that have been in lectures, or tutorial questions and pick up similarities. I guess you follow a pattern that they have taken and get an answer, rather than here it's probably more important questions I should know the answer to, and if I was given this as coursework I probably wouldn't do as well.

This student was particularly articulate and confident in voicing her difficulties with the task, but other students became more unsettled as the task went on. This had the result that in some interviews my plan of non-involvement during the task was not followed for ethical reasons; one student became agitated that they could not attempt many questions and so in this interview I took on a more of a tutorial role than that of an interviewer. This type of occurrence suggests that, for some students, an example generation task may be a very poor tool to explore concept images of sequence properties; they either had relatively empty concept images or were unable to evoke their concept image. For a student such as Student 5, who in the quote above described what Lithner (2008) calls *imitative reasoning*, perhaps a near-empty evoked concept image is a valid conclusion.

In most interviews, the schedule of questions to be asked during the semi-structured discussion period (see Subsection 4.2.1) was not strictly followed; instead the conversation was drawn to "spontaneous questions in the natural flow of an interaction," described by

Patton (2002) as an *informal conversational interview.* Consequently in the main study, described in the next chapter, the discussion period was planned to be an interaction of this type.

4.2.5 Conclusions from the second pilot study

The second pilot study has provided some more insight into the thought processes of students, or at least the thought process reported by students as they 'thought aloud' during the task. In one case, an **INS** representation of a sequence as an interval of the real line was treated as a valid sequence, and a student believed they could represent any sequence both in this form, and as a list of numbers. It was speculated that some students are more prone to providing **INS** examples, and some questions more likely to provoke **INS** examples, although this may be a result of students having been recently taught about boundedness in the context of intervals rather than sequences.

In conjunction with the first pilot study, the second study has demonstrated that example generation tasks can be successfully used to explore students' concept images of sequence properties, but only when questions are of appropriate difficulty for the students answering them.

4.3 Summary of chapter

This chapter has presented a reflective account of two pilot studies which were conducted a year before the main study. The pilot studies were not designed or analysed from a phenomenological perspective, and were presented as illustrative of the types of behaviour, concept images and spontaneous conceptions that I observed from students prior to the main study.

The first pilot study was quantitative, and reported on an example generation task given to 101 undergraduate students. The task asked the students to provide examples of sequences subject to certain constraints, and the students also had access to a comprehensive definition sheet throughout the task. It was found that students often not only did not manage to provide a sequence which satisfied the constraints, they sometimes provided mathematical objects which were not sequences. These two types of incorrect answers were classified as incorrect sequences (**IS**) and incorrect non-sequences (**INS**). Typically, an **INS** response would, to some extent, reflect the constraints requested (such as an unbounded set given when an unbounded sequence was requested).

The second pilot study aimed to shed more light on the thought process which leads to **INS** answers. It selected some of the questions from the first pilot study and asked students to complete them as part of a semi-structured interview. It was found that some students provided **INS** answers to several questions, and that certain questions prompted many students to give **INS** answers. In the discussion period of the interview, after the task had been completed, some students indicated that they believed they could switch between a valid representations of a sequence, and an **INS** answer without corrupting the sequence.

A summary of the number of participants, number of questions, with the mean student scores and standard deviations are given in Table 4.6.

TABLE 4.6: A summary of average student scores in the pilot studies.

Study	No. students	No. questions	Mean score	St. dev
Pilot study 1	101	20	6.95 (34.7%)	3.98 (19.9%)
Pilot study 2	6	11	2.83 (25.8%)	2.23 (20.2%)

In each pilot study there has been evidence that students' spontaneous conceptions have remained in focus during the task, most commonly when dealing with the properties of *increasing* and *decreasing*, where 87% of students could not give an example of a sequence which was both increasing and decreasing, and 2/5 of such students instead giving a sequence which is oscillatory to some extent. These properties are considerably simpler (in terms of their definitions) than other properties studied in the literature.

Chapter 5

Main Study: Planning and Data Analysis

This chapter describes the planning and phenomenographic data analysis procedure of an interview-based example generation task given to students. After briefly discussing the aims of the study, Section 5.2 discusses the how the task is related to the pilot studies, and describes the participants, the task given to students in the interview, and the interview itself. Issues related to the transcription of the data are then explored in Section 5.3. Next the phenomenographic data analysis of the study is outlined in Section 5.4, which includes descriptions including open codes, headings, categories of description and dimensions of variation (Chapter 6 presents the outcomes of this data analysis). Lastly, intersubjective and ethical considerations are highlighted in Sections 5.5 and 5.6, respectively.

5.1 Aims of the main study

The aims of the main study are related to the thesis' two research questions, which are restated here for convenience:

1. How successful are students at accurately generating examples of sequences satisfying certain combinations of properties?
2. What is the qualitative variation in students' experiences of sequence generation?

The primary aim of the main study is to answer the second research question. This chapter describes the data collection and analysis procedures of the main study, the outcomes of which are reported in the next chapter. In common to all the studies in the thesis, the first research question is also addressed by the main study, which provides more evidence for the conclusions eventually drawn in Chapter 9.

5.1.1 Relation to phenomenography

Chapter 3 described in detail the research specialism of phenomenography. The second research question of this thesis is a phenomenographic one and, as discussed above, the main study was conducted to address this research question. Therefore throughout the following discussion of data collection and analysis, the description and terminology used is oriented to those used in phenomenography ('bracketing', 'category of description', 'dimension of variation' etc).

5.2 Planning for the main study

5.2.1 Participants

The participants in the main study were first year undergraduate mathematics students from the University of Warwick. A different cohort from the pilot study was used because at the time the second study was scheduled to take place (autumn 2008), the appropriate cohort of students at Loughborough University would be second year undergraduate mathematicians, some of whom took part in the pilot studies.

The University of Warwick had around 350 single-honours students and more than 200 on a joint degree including mathematics in 2008, and so there was theoretically a large number of students who could volunteer to complete the task under interview conditions. All students were contacted via e-mail and asked to participate in the study. It was hoped 20 students would volunteer, but only 15 responded and attended their respective interviews. By means of payment for their time each was offered feedback on their performance and a 'sigma'-branded USB memory stick.

These undergraduates had previously met formal definitions of sequence properties in lectures, i.e. definitions with quantifiers, rather than as 'wordy' definitions. It was hoped that this would eliminate some of the problems commented upon (in the pilot studies) involving students' unfamiliarity with the definitions.

5.2.2 Design of the task sheet

The version of the task sheet used in the second (interview-based) pilot study was answered badly by students, with five of the six students answering fewer than four of the eleven questions correctly (see Section 4.2). In the subsection reflecting on the pilot study's methodology (4.2.4), it was noted that students not being able to answer questions correctly is not necessarily a bad thing in an interview-based study, for the think-aloud protocols and discussion period can provide insight into students' thought processes. However, as reflected upon in the subsection related to ethical considerations (4.2.3), it was unnerving for students to be asked to complete a task they could not approach, and so serious consideration was given to using a simpler task in the main study.

On reflection, I felt it appropriate to use the same task in the main study because the participants in the main study had been introduced to formal definitions of sequence properties in the weeks leading to the interviews, and had been dealing with theorems and proofs using such formal definitions. The definition sheet was slightly modified to correspond to the notation used at The University of Warwick (in particular, an entire sequence is written "(a_n)" to distinguish it from the specific term of the sequence "a_n"). The definition sheet used in the main task can be found as Figure 5.1, and the main study's task sheet can be found as Figure 5.2.

5.2.3 Description of interviews

Each student had agreed to meet for an hour at a mutually convenient time, although not all sessions lasted a full hour. They were first asked to read the participant information sheet (see Appendix A.1 for a copy), and then asked if they would consent to audio and video recording during the interview. All participants agreed to be audio recorded, one participant declined the video recording and the video camera failed for one interview.

Once the student had signed an informed consent form (see Appendix A.2), I began by introducing myself and discussing my research. Then for around 15-10 minutes I chatted

Review of Sequences

Definitions

Remember that a sequence is a list of real numbers

$$(a_1, a_2, a_3, a_4, \dots)$$

where $(a_n)_{n=1}^{\infty}$ denotes the whole sequence.

Definition: A sequence $(a_n)_{n=1}^{\infty}$ is *increasing* if and only if $\forall n \in \mathbb{N}$, $a_{n+1} \geq a_n$.

Definition: A sequence $(a_n)_{n=1}^{\infty}$ is *strictly increasing* if and only if $\forall n \in \mathbb{N}$, $a_{n+1} > a_n$.

Definition: A sequence $(a_n)_{n=1}^{\infty}$ is *decreasing* if and only if $\forall n \in \mathbb{N}$, $a_{n+1} \leq a_n$.

Definition: A sequence $(a_n)_{n=1}^{\infty}$ is *strictly decreasing* if and only if $\forall n \in \mathbb{N}$, $a_{n+1} < a_n$.

Definition: A sequence $(a_n)_{n=1}^{\infty}$ is *monotonic* if and only if it is increasing or decreasing.

Definition: A sequence $(a_n)_{n=1}^{\infty}$ is *bounded above* if and only if $\exists U \in \mathbb{R}$ s.t. $\forall n \in \mathbb{N}$, $a_n \leq U$.

Definition: U is an *upper bound* for the sequence $(a_n)_{n=1}^{\infty}$ if and only if $\forall n \in \mathbb{N}$, $a_n \leq U$.

Definition: A sequence $(a_n)_{n=1}^{\infty}$ is *bounded below* if and only if $\exists L \in \mathbb{R}$ s.t. $\forall n \in \mathbb{N}$, $a_n \geq L$.

Definition: L is an *lower bound* for the sequence $(a_n)_{n=1}^{\infty}$ if and only if $\forall n \in \mathbb{N}$, $a_n \leq L$.

Definition: A sequence $(a_n)_{n=1}^{\infty}$ is *bounded* if and only if it is both bounded above and below.

Definition: A sequence $(a_n)_{n=1}^{\infty}$ *diverges* if and only if it does not converge to any finite limit.

Definition: $(a_n)_{n=1}^{\infty}$ tends to infinity if and only if $\forall C > 0$, $\exists N \in \mathbb{N}$ s.t. $n > N \Rightarrow a_n > C$.

FIGURE 5.1: The definition sheet given to students in the main study, interview component.

with the student about their experiences in secondary and tertiary education, and how they were finding their courses. Although this data would not form part of the data for my analysis because it was not relevant to example generation, this portion of the interview was always included as it functioned as an ice breaker, allowing a rapport to be established between myself and the student.

The student was then given the list of definitions and asked if they had seen them before. Every student indicated that they had seen the definitions in that form before, although some students indicated that they were uncomfortable with them. In these cases the student was reassured that although they were going to be given a task related to the

Please give an example of each of the following, **or state that this is impossible**.

You can write your sequence in any way you choose:
As a list of numbers, as a formula, etc.

You do no need to prove your answers.

[Questions on page 1]

1. A strictly increasing sequence
2. An increasing sequence that is not strictly increasing
3. A sequence that is both increasing and decreasing
4. A sequence that is neither increasing nor decreasing

[Questions on page 2]

5. A sequence that has no upper bound
6. A sequence that has neither an upper bound nor a lower bound
7. A bounded, monotonic sequence

[Questions on page 3]

8. A sequence that tends to infinity
9. A sequence that tends to infinity that is not increasing
10. A sequence that tends to infinity that is not bounded below
11. A strictly increasing sequence that does not tend to infinity

Figure 5.2: Questions on the the main study task sheet, interview component. The actual sheet was on three pages and had three-inch gaps between questions.

definitions they would not have to formally use them, and it was reiterated they were not being assessed in any way.

Students were then presented with the task sheet and asked to work through the questions. They were reminded that they could skip any question and tackle them in any order they liked, and that they were encouraged to think aloud, even if this felt alien to them. Finally it was restated that the task was not a test, but even so they would be offered no assistance or encouragement while they were completing it. It was explained that I was happy to go through their answers with them afterwards if they wished me to. There was no formal time limit on the sheet; each student had as long as they needed to complete the task and were asked to indicate when they were finished.

The final part of the interview was an unstructured discussion period, where I discussed with the student the answers they had given. This discussion period could loosely be categorised as an *informal conversational interview*, as defined by Patton:

> The informal conversational interview relies entirely on the spontaneous generation of questions in the natural flow of an interaction. (Patton, 2002, p.342)

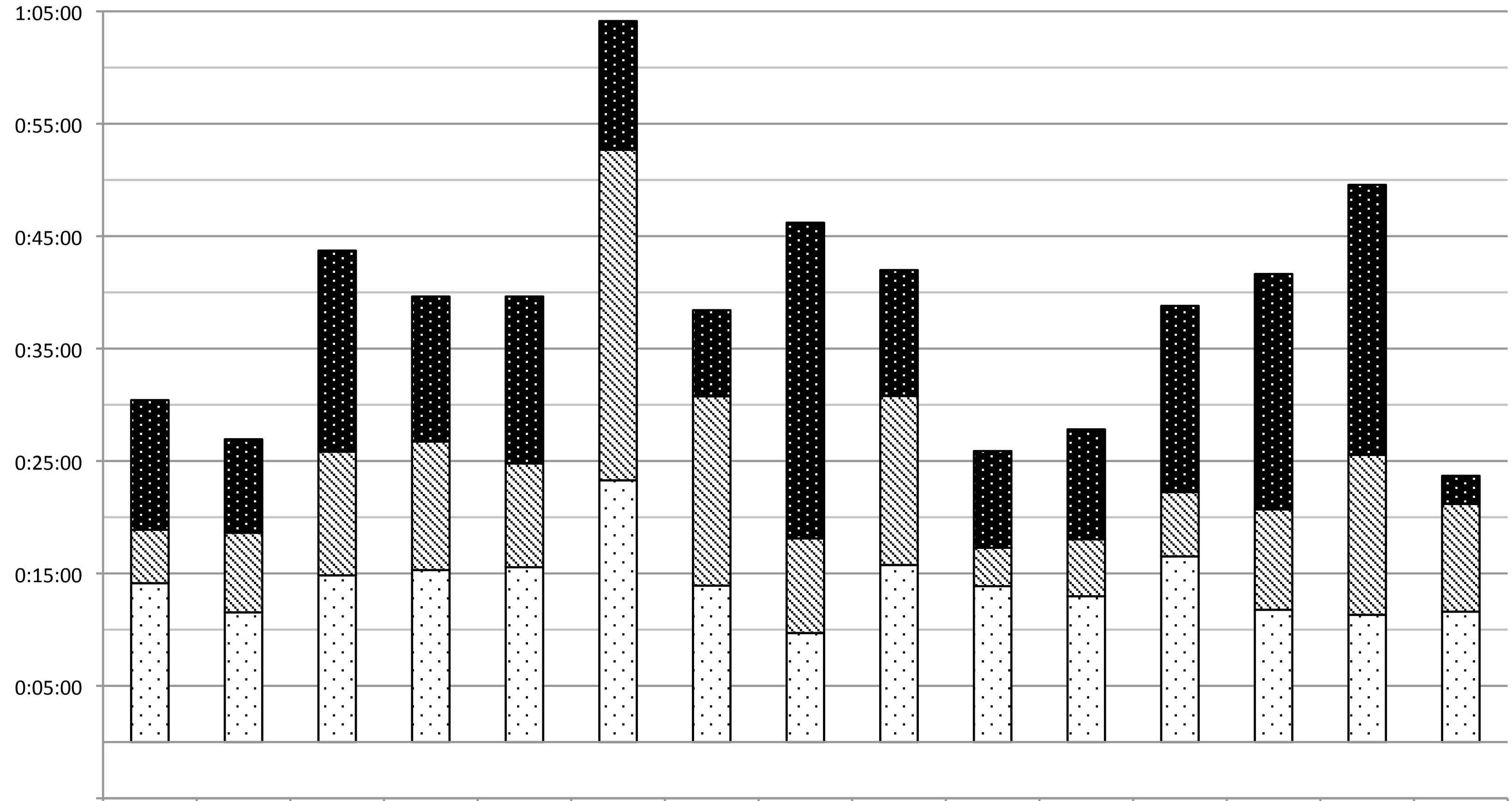
I say loosely categorised because the matters which were discussed in the interview were focused on the students' responses to the example generation task, so the phrase 'natural flow of an interaction' is taken more in the context of a mathematics tutorial/supervision. Because the student and interviewer had the common ground of the example generation task, many of the problems with this type of interview (such as the style being less systematic and comprehensive) are reduced because the conversation, although not formally structured, was focused on the questions on the task sheet and the answers the students gave. In each interview I initially focussed on questions the student had left blank or answered 'impossible', and then let the conversation progress as naturally as possible from there. This is common in phenomenographic studies, even when there are a structured series of questions (Åkerlind, 2005b).

In summary, each interview's hour-slot was scheduled into four periods:

- Reading the information sheet and completing consent form
- An initial chat about first year mathematics and an introduction to the task
- Completing the task
- Discussion of answers and general advice

Figure 5.3 outlines the time taken for each student in the chat, task, and discussion periods. The length of time spent reading the information sheet and completing the informed consent form is not included in Figure 5.3 because I did not begin the audio recording until the student consented that I may do so, and so there is no data to indicate the length of this period. The length of the discussion period for Oksana[1] was brief because she arrived to the interview late, and had to catch a bus home.

[1] a pseudonym



	Anna	Ben	Valter	David	Edha	Phalgum	Guan	Haaroon	Ian	Joe	Ken	Laura	Mike	Nicola	Oksana
Discussion	0:11:32	0:08:19	0:17:53	0:12:54	0:14:51	0:11:27	0:07:40	0:28:05	0:11:12	0:08:36	0:09:47	0:16:34	0:20:56	0:24:01	0:02:30
Task	0:04:45	0:07:05	0:11:00	0:11:26	0:09:14	0:29:24	0:16:50	0:08:25	0:15:02	0:03:25	0:05:04	0:05:43	0:08:56	0:14:14	0:09:35
Chat	0:14:08	0:11:32	0:14:50	0:15:18	0:15:33	0:23:17	0:13:55	0:09:42	0:15:45	0:13:52	0:12:58	0:16:31	0:11:46	0:11:19	0:11:36

FIGURE 5.3: Duration of interviews, split into initial chat, task completion and discussion of answers. All times are presented in the form h:mm:ss. The data is measured from the audio recordings.

Relation to phenomenography

In each interview there were two sources of verbal data. The first was the comments the student made as they completed the task. The choice of which comments were appropriate was left to the student, although each student was prompted to talk as much as they felt comfortable doing. Comments ranged from stating their interpretation of what they were doing:

> I'm just writing the definitions.

to deeper reflections on their thought process:

> I'm not too sure about this one either, because if it's tending to infinity but then it's not bounded below, but it has to be bounded below because you have to start at a certain point.

The second source of data was the discussion period after the task phase, and as indicated by Figure 5.3, the duration of this part of the interview was typically longer. Here the conversation was lead by the interviewer, but focussed on the answers given by the student in the task and what the student had talked about during the task. Following Bowden's (2005) suggestions for phenomenographic interviewing, questions typically asked for more information, or encouraged reflection upon ideas raised by the student. At all times I attempted to bracket my own judgement. Despite this, the social contract between myself (someone who is experienced in mathematics) and the student (someone who is less experienced) meant that this type of questioning was at times perceived by a student as hints that they are mistaken (Koichu and Harel, 2007), as is illustrated in the following extract from the main study interviews:

> **Interviewer:** That one, you sort of paused.
> **Student:** I was thinking monotonic, if I wanted it to be increasing or decreasing.
> **Interviewer:** And which did you decide?
> **Student:** I decided decreasing, and changed it to increasing. Nah, I decided increasing and changed it to decreasing.

> **Interviewer:** Ok, why was that?
>
> **Student:** I don't know, I might change it back [laughs].

There is little that can be done to remedy this situation. At the start of the interview I had remarked to each student that I would avoid telling them if they were correct or incorrect until the end of the interview. In hindsight, it would have also been prudent to stress that I would be asking the same questions whether their answers were correct or incorrect.

5.3 Transcription and marking of answers

5.3.1 Pseudonyms assigned to students

During the data analysis each student was assigned a pseudonym, as listed in Table 5.1. This is because each student was informed at the start of their interview that "the recordings will be treated in strictest confidence and any reference to them shall hide the identity of the student in question" (Participant Information Sheet, Appendix A.1). In Table 5.1, each student's gender and ethnicity has been maintained in an effort to keep as much (unidentifiable) information about each participant present as possible whilst still acting in accordance with the Participant Information Sheet.

The reader may feel that this conflicts with the phenomenographic principle of bracketing data. Perhaps it might be easier for a reader to bracket the quoted incidents and analysis presented in the next chapter had the students' identities been entirely erased? While to a certain extent this might be true, I believe two other factors outweigh this. The first is still related to bracketing, but to my bracketing rather than the reader's. When analysing and reporting the data, I have had access to more information than the reader, and so the interpretative analysis presented in the next chapter will be subject to my value-laden judgements and interpretations. Although I have made an effort to be as objective as possible in this regard, the reader is in a better position to evaluate my bracketing if provided with as much information as possible. The second factor is that I believe an alternative allocation of names or markers would remove some of the humanity from the students in question. Had they been labelled "A", "B" etc, I believe they would have seemed less human, and I also felt it would be inappropriate to allocate

TABLE 5.1: Pseudonyms for interviewed students

Interview order.	Pseudonym
1.	Anna
2.	Ben
3.	Valter
4.	David
5.	Edha
6.	Phalgun
7.	Guan
8.	Haaroon
9.	Ian
10.	Joe
11.	Ken
12.	Laura
13.	Mike
14.	Nicola
15.	Oksana

every student a Western (or Eastern) name, a male (or female) gender, regardless of their actual background. I feel the actual presentation of names hides the identity of the students, while reminding the reader that they are unique individuals.

5.3.2 Categorisation of answers

The answers each student had given during the task phase of the interview are reproduced in Table 5.2 (for Questions 1-5) and Table 5.3 (for Questions 6-11). In these tables, intended to give the reader an overview of the different interviews, only the 'final' answer is reproduced, rather than answers given during working, or alternative answers. The answers in Table 5.2 and 5.3 are re-created as to be as faithful as possible to the originals (so, for instance commas and ellipses within a student's list notation are never corrected).

5.3.3 Transcription of interviews

Transcription protocol

In the first stage of analysis the audio from each interview was transcribed. Apart from decisions of how to typeset mathematical statements, no consideration was made of the

	Q1.		Q2.		Q3.		Q4.		Q5.	
Anna	$(a_n) = n$	C	Graph Drawn	B	$a_n = \sin n$ $0 \leq n \leq \pi$	INS	$a_n = 2$	IS	$(a_n) = n$	C
Ben	$1, 2, 3, 4, \ldots$	C	$1, 1, 2, 3, 3, \ldots$	C	$1, 1, 1, 1 \ldots$	C	$n(-1)^n$	C	$1, 2, 3, 4, \ldots$	C
Valter	$(a_n) = n$	C	$(a_n) = 1$	C	$(a_n) = 1$	C	$(a_n) = (-1)^n$	C	$(a_n) = n$	C
David	$a_n = n$	C	$a_n = 1$	C	$a_n = \sin(n)$	IS	$a_n = k$ $k \in \mathbb{R}$	IS	$a_n = n$ $n \in \mathbb{N}$	C
Edha	$\frac{n+1}{n}$ $n > 0$	IS	$1, 2, 3, 1, 3, 4 \ldots$	IS	$a_n = (-1)^n(n+1)$	IS	$1, 1, 1, 1, 1, 1, \ldots.$	IS	$a_n = n$ $n > 0$	C
Phalgun	$(a_n) = n$	C	$(a_n) = \{1, 1, 2, 2, 3, 3, 4, 4, \ldots\}$	C	$(a_n) = 1$	C	$(a_n) = \sin n$	C	$(a_n) = n$	C
Guan	$a_n = 1, 2, 3, 4, 5, \ldots$	C	$a_n = 2, 2, 2, 2, 2, \ldots,$	C	$a_n = 1, 1, 1, 1, \ldots,$	C	$a_n = 1, 1, 1, 1, \ldots$	IS	$a_n = 1, 2, 3, 4, 5, 6, \ldots$	C
Haaroon	$(a_n) = n$	C	$(a_n) = \log n$	IS	$(a_n) = \sin n$	IS	$(a_n) = 1$	IS	$(a_n) = \log n$	C
Ian	$a_n = n$	C	$1, 2, 1, 3, 1, 4, 1 \ldots$	IS	I		$a_n = 1$	IS	$a_n = n$	C
Joe	n	C	$[a_{2n}] = 0$ $[a_{2n-1}] = n$	IS	$(-1)^n n$	IS	$\{a_n\} = 3$	IS	n	C
Ken	(n)	C	(3)	C	(3)	C	$(-1)^n$	C	(n)	C
Laura	$1, 2, 3, 4 \ldots$	C	$1, 1, 2, 2, 3, 3 \ldots$	C	$3, 3, 3 \ldots$	C	I		$a_n = n^2$	C
Mike	$a_n = n + 1$	C	$a_n = 7$	C	$a_n = 1$	C	$a_n = (-1)^n$	C	$a_n = n!$	C
Nicola	$(a_n) = n$	C	$a_n = 1, 1, 2, 2, 3, 3, \ldots$	C	$(a_n) = 1$	C	$(a_n) = (-1)^n$	C	$(a_n) = n$	C
Oksana	$1, 2, 3, 4 \ldots$	C	$1, 1, 2, 2 \ldots$	C	$2, 2, 2 \ldots$	C	$2, 3, 1, 4, 0, 5, -1 \ldots$	C	$1, 2, 3 \ldots$	C

TABLE 5.2: Responses to Questions 1–5, with appropriate classification from the taxonomy given in Table 4.2.

	6	7	8	9	10	11
Anna	$(a_n) = (-1)^n n$ C	$a_n = 2$ C	$a_n = n$ C	I	I C	$(a_n) = \frac{1}{n}$ C
Ben	$1, -1, 2, -2, 3, -3 \ldots$ C	$1, 1, 1, 1, 1,$ C	$1, 2, 3, 4, \ldots$ C	— [Interpreted as 'I']	— [Interpreted as 'I'] C	$\frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \frac{15}{16}$ C
Valter	$(a_n) = (-2)^n$ C	$(a_n) = \frac{1}{n}$ C	$(a_n) = n^2$ C	Graph Drawn B	Graph Drawn B	$(a_n) = x^{\frac{1}{n}}$ $0 < x < 1$ C
David	$a_n = (-1)^n n$ C	$a_n = -\frac{1}{n}$ C	$a_n = ke^n$ $k \in \mathbb{R}^+$ C	I	I C	$a_n = \sqrt{n}$ IS
Edha	$a_n = (-1)^n(n+1)$ C	$a_n = n + \frac{3}{2}$ IS	$a_n = \frac{n+1}{n}$ IS	I	$a_n = n$ For all integers INS	I
Phalgun	$(a_n) = (-n)^n$ C	$(a_n) = 2$ $(a_n) = \frac{1}{n}$ C	$(a_n) = 2^n$ $(a_n) = e^n$ C	B	B	B
Guan	$a_n = (-1)^n n$ C	$0, 0, 0, 0, 0 \ldots,$ C	$a_n = n = 1, 2,$ C	$a_n = 1, 0, 2, 0,$ $3, 0, 4, 0 \ldots$ IS	I C	I
Haaroon	B	$(a_n) = \frac{1}{n}$ C	$(a_n) = \log n$ C	I	B	B
Ian	$a_n = (-2)^n$ C	$a_n = (-1)^n$ $-1 \leq a_n \leq 1$ IS	$a_n = n^2$ C	I	$a_n = n$ $n \in \mathbb{Z}$ INS	$a_n = 1 - \frac{1}{n}$ $n \to \infty$ C
Joe	$(-1)^n n$ C	$1 - \frac{1}{n}$ C	n C	$-n$ IS	B	$1 - \frac{1}{n}$ C
Ken	$((-1)^n n)$ C	(3) C	(n) C	I	I C	$(10 - \frac{1}{n})$ C
Laura	$1, 2, -2, 3, -3 \ldots$ C	$a_n = \frac{1}{n}$ C	$a_n = n$ C	I	I C	I
Mike	$a_n = (-2)^n n$ C	$a_n = \frac{1}{n}$ C	$a_n = n^{32}$ C	$10, 9, 30, 29,$ C	$a_n = -n^4$ IS	$a_n = (\frac{n}{n+1})$ C
Nicola	$a_n = n(-1)^n$ C	Graph Drawn B	$(a_n) = 5^n$ C	I	B	Graph Drawn B
Oksana	$0, 1, -1, 2, -2, 3 \ldots$ C	$\frac{1}{n}$ C	n C	$231564897 \ldots$ C	B	$-\frac{1}{n}$ C

TABLE 5.3: Responses to Questions 6–11, with appropriate classification from the taxonomy given in Table 4.2.

interviews' meaning. This was for two reasons, first to enable a constant comparison during analysis and second to function as a safeguard for transcription quality.

After examining descriptions of interview protocols from various authors such as Poland (2001) and Ochs (1979) I decided to write my own protocol for the interview data. This was because the protocols in these accounts were designed for conversation-based interviews with multiple interviewees, whereas in the main study students were interviewed individually, and during the task phase for large pieces of the interview there was a single speaker. The protocol used was as follows:

- Each transcript begins with a time in square brackets, such as [03:23], which corresponds to when the conversation began relative to the timer on the audio file. At various points during the interview, times may be added into the transcript to aid navigation of the recordings, for instance after a long pause.
- Any comments that needed to be made during transcription, such as something being inaudible the comment would be made [in square brackets]. Square brackets were also used for elements of the conversation which are impossible to transcribe, such as [laugh] or [sigh].
- Mathematical content would be typeset in standard latex form based on the context. Where the context was ambiguous a comment would be added.
- A new line is started after one of three conditions:
    1. A new speaker begins
    2. After a long pause (dependent on context, but typically 5+ seconds)
    3. When it was clear the student moved onto a new topic or changed question
- Lines that start **Interviewer:** were said by me
- Lines that start **Student:** were said by the student
- A line that ends with / is interrupted by the next line (it does not indicate a pause)
- Shorter pauses, and discourse markers (such as "erm" and "like") were not transcribed, instead they were marked with appropriate punctuation.
- An abrupt pause in the middle of a statement (but not an interruption), was marked by —.

For the 13 students where a video recording existed to accompany the interview audio transcript, the video was synchronised with the transcript so that ambiguous passages could be interpreted with more evidence. To a certain extent, this helps avoid the problems associated with written transcripts of data as commented upon in the methodology, subsection 3.4.2.

5.4 Coding of data

When each interview had been transcribed, the transcripts were coded inductively from the "bottom-up", in a way similar to a grounded theory analysis (Glaser, 1992; Glaser and Strauss, 1967; Strauss and Corbin, 1998). The aim here was to "identify concepts" and "discover their properties and dimensions" within the transcripts of each interview (Strauss and Corbin, 1998, p.101). Such data analysis techniques are common when taking a phenomenographical approach (Åkerlind, 2005c; Richardson, 1999), as I discussed in my methodology chapter. This section will go in to more detail about the different stages of my coding process, which are first summarised here:

1. Open coding of data
   Here salient passages of transcript are labelled with a representative code in an attempt to objectively capture their meaning.
   Described in Subsection 5.4.1, open codes are written in SMALL CAPS.

2. Collection of codes under headings
   Codes with the same meaning are merged, and those with similar meanings are grouped together under a descriptive heading.
   Described in Subsection 5.4.2, headings are written in *lower case italics*.

3. Formation of categories of description
   Headings that describe related phenomena are grouped, and the associated transcript extracts compared.
   Described in Subsection 5.4.3, categories of description are written in *Title Case Italics*.

4. Formation of dimensions of variation
   Related categories of description are ordered in terms of their sophistication.

> Described in Subsection 5.4.3 also, dimensions of variation are not given a special typeface. From the analysis of the main study data four dimensions of variation were generated: Using definitions, Representation of sequences, Sequence construction strategies, and Justifications (see next chapter).

At each stage, the aim remains to capture every salient detail and aspect from the data. What is salient at a particular point is up to a matter of interpretation, but in this interpretation of a phenomenographic data analysis which is strongly based on the writings of Åkerlind (2005c), Green (2005) and Marton and Booth (1997), it should be possible to take the dimensions of variation formed in the final stage and use their categories of description to code the transcripts, and still capture the salient features of the interview.

The above account is quite abstract and should be thought of an an overview only. In the descriptions of each stage below, examples from the transcripts, codes, headings, categories and dimensions from the main study data and analysis will provide some firmer ground.

Use of technology in data analysis

Throughout the coding process the qualitative management software Atlas.ti (Muhr, 2010) was used. The purpose of Atlas.ti within qualitative data coding is discussed by Muhr (1991) and more recently by Lewins and Silver (2007), and so the intricacies of the software package will not be discussed in detail here. In general, the software is equivalent to having unlimited margin space to write notes and attach codes, but with the modification, searching, and retrieval efficiency of a desktop computer.

5.4.1 Open coding

During open coding, the interview transcripts are disassembled in order to explore and identify units of analysis to code for meanings, feelings, actions, events and so on (Cohen et al., 2007). More specifically the transcripts were read line-by-line, labelling incidents (i.e. what happens at a particular time) with short codes of between 1 and 8 words. Where it was more appropriate to label passages of text, say at the paragraph level, whole paragraphs were coded (Atlas.ti is able to attach a code to a quotation of any length). Some incidents from the interview could not be adequately summarised by a short code. For these incidents a short name was still chosen for the code, but comment was also attached to the code outlining the incident in more detail. After the first coding there were 130 codes. A sample list of codes is given in Table 5.4.

Table 5.4: A selection of codes from the initial open coding. All codes beginning with the letter 'C' or 'D' are listed.

C and N increase $\Longrightarrow$ sequence must increase	can reach infinity	combining definitions
common sense	Comparing to functions	constant $\Longrightarrow$ increasing
constant is increasing & decreasing	constant is monotonic	constant sequence
constructing sequence	correct combination of global and local properties	correctly use definition
decreasing $\Longrightarrow$ can't tend to infinity	definition of infinity	degrees of difficulty
different domains	difficulty of questions	distance from limit
domain as $\mathbb{R}$	domain as $\mathbb{Z}$	domain can not be real
draws graph	dynamic sequence	

The open coding process attempted to label passages as objectively as possible, with the aim to code what happened or what was said in the corresponding interview segment, rather than my interpretation of the passage. Despite this there will always be a degree of interpretation as to the meaning of events (such issues with the intersubjectivity of interview study analysis is described in more depth in Section 5.5). For instance, the code common sense was used as a label for extracts such as when Valter remarked:

Valter: Just appealing to common sense

but also to his statement:

Valter: Intuitively I'd say

and also to Phalgun's comment:

Phalgun: Both increasing and decreasing, a trivial one, $(a_n) = 1$

These last two extracts are arguably less explicitly about common sense, certainly in the sense that the student did not say the phrase itself, but in my initial open coding they were labelled as such. This is because, for instance in the case of the final quote, when a mathematician says something is 'trivial' they are using a type of mathematical common sense; the student was appealing to our collective common sense as mathematicians. Atlas.ti allows extracts to be coded with multiple codes, and so the latter quote was also coded TRIVIAL and CONSTANT SEQUENCE. Latter stages of data analysis would reflect upon how appropriate the codes were to the passages coded and remove duplications; here the aim was to capture the most salient features with appropriate codes.

Where I wished to make an interpretive observation, such as on Phalgun's use of the word 'trivial' and its links to the notion of common sense, a theoretical memo was created. These memos could be attached to a passage of text of any length (and so are equivalent to writing a note on a post-it note and sticking it on the page). Memos were written which focused both on line-by-line incidents but also over longer periods of analysis (such as a general comment about a student over the whole interview). When re-reading memos, it became clear that some identified themes across wider durations of interviews and possibly across different interviews, and accordingly the relevant passage(s) of the interview were coded to reflect this.

The transcripts were coded in the same order as the students were interviewed, and codes were re-used where appropriate (as was seen in the case of COMMON SENSE, above). During the open coding other transcripts and previous parts of the same transcript were 'bracketed', i.e. the statements were taken at face value in the sense the student intended (see the methodology in Chapter 3 for a more detailed account of this phenomenographic term).

After all interviews were coded there was a total of 130 codes, a brief selection of which was given in Table 5.4. Some of these codes dealt with what the student remarked, for instance the use of code COMMON SENSE was described in the last section. Similarly the code DEGREES OF DIFFICULTY referred to instances where a student remarked that one question was more difficult than another. Other codes deal with what the student did (DRAWS GRAPH), a student's interactions with their work (DYNAMIC GRAPH), or the logical implications of what a student said, such as a statement where a student makes a claim that a decreasing sequence can't tend to infinity (DECREASING $\implies$ CAN'T TEND TO INFINITY). Finally, some codes labelled phenomena first observed in the pilot studies, such as codes related to **INS** answers (DOMAIN AS $\mathbb{Z}$).

At this stage some codes with essentially the same meanings were combined, provided the codes' associated transcript extracts indicated this was appropriate. For instance the codes USES TECHNIQUE SEEN IN CLASS and RECALLS CLASSROOM PROOF were both merged with the PRIOR WORK code.

5.4.2 Collection of codes under headings

Once all codes had been created, and merged where appropriate, the codes were then grouped under headings. A heading is a more abstract higher order concept, which describes a similar group of codes. The headings were created so that when it came to forming categories of description, not only would there be fewer units to work with, but also so that the "problems, issues, concerns and matters" that are important to the students being studied could be highlighted (Strauss and Corbin, 1998, p.114). Continuing to look at the code COMMON SENSE, we can see that it was placed under a heading which collected codes which described how students justified their answers:

> Heading: *justifying answers*
>
> COMMON SENSE (see above)
>
> DYNAMIC SEQUENCE (phrases such as "the sequence goes to")
>
> IMPOSSIBLE (where the only justification was the word impossible)
>
> JUSTIFIES ANSWER (catch-all code)
>
> LIST EASIER (justifying the form of the answer)
>
> NOT CONFIDENT WITH ANSWERS (no justification, only doubt)
>
> NOTATION DEPENDENT UPON COMPLEXITY OF SEQUENCE (justifying the form of the answer)
>
> UNSEEN SEQUENCE (unsure of answer because it was new to them)
>
> UNSURE OF ANSWER (otherwise unsure of the answer)

Some of the codes included within the *justifying answers* heading included codes related to the lack of justification, such as UNSEEN SEQUENCE (where a student remarked that they had not seen a sequence of a certain type before), and UNSURE OF ANSWER. This is an illustration that although an open coding cannot code for the absence of a phenomenon, these two codes suggest what students may do rather than justify their answers.

Occasionally when it became clear that the phenomenon under consideration had a relative grounding in the literature, codes were later introduced under a heading. In practice the only instance of this is when Antonini's (2006) classification of example generation was used to categorise some of the variation within the heading *construction strategies*. Antonini's (2006) strategy is outlined in the literature review, and the corresponding dimension in the data is described in more detail in Section 6.4.

5.4.3 Categories of description and dimensions of variation

In the next stage of analysis, all the incidents associated to codes under a specific heading were listed. Where two headings covered similar content, initially the incidents were considered together. In conjunction with a student's answer sheet and the audio/video recoding (but still bracketing other incidents in the interview), the incidents under a heading were compared, and ordered roughly in terms of their sophistication. This was usually possible because the set of incidents came from the same or similar headings, so it was rare to have two extracts that could not be compared at this stage. The aim was not to create a definitive ranking, but to determine and contrast the different type of phenomena coded under a particular heading(s).

Once a rough ordering of sophistication of extracts had been made, the salient features of each were described. These descriptions became 'categories of description' that collectively form the hierarchal structure of a dimension of variation. Extracts that exemplified categories of description particularly well were specially marked, and it is a selection of these incidents that make up the quotations presented in the next chapter. Once the related categories of description had been drafted, they were again compared with the codes and quotations from the relevant heading, and the categories modified if they did not capture the data as fully as possible.

Finally, when the extent of a dimension of variation had been defined as completely as possible, I reread through the transcripts looking for extracts that might have been missed by earlier stages of coding, but which still exemplified the dimensions of variation. After this process, for a given category of description within a dimension of variation, there were (1) incidents which originated from the initial open-coding analysis, and (2) incidents that were later included after the dimension of variation had been defined. When reporting the dimensions in the next section, I do not distinguish between incidents with different origins; quotations should be read as illustrating the dimension of variation, rather than originating from a code→heading→category path.

5.4.4 Relation to grounded theory

The type of process that forms categories of description and dimensions of variation is similar to 'axial coding' in grounded theory, where a researcher attempts to make links between headings and codes to integrate codes around the axes of central categories (Cohen et al., 2007). The procedure used in phenomenography is different however, as it has the aim of producing dimensions of possible variation from the phenomena delimited by the headings and its codes.

Unlike the progressions from transcripts to codes, and then to headings, it is not the case that categories of description are formed from headings, and afterwards grouped into dimensions of variation. Instead, similar headings are considered together and the possible variation across and within these headings is explored, which consists of going back to the transcripts and comparing the coded extracts. Variation of experience unearthed by these comparisons was captured by a series of categories of description. These categories of description are, in some sense, a dimension of possible variation. However the structure of a definition is hierarchical, ranging from less sophisticated incidents to more sophisticated ones. Moreover a dimension of variation should have the contrast, generalisation, separation, and fusion properties as described in Section 3.3.1 of the methodology. In other words a dimension of variation should allow comparisons between their component categories of description, suggest generalisation to other concepts and objects, be considered in isolation from the other dimensions, and combined with other dimensions without losing the meaning from one dimensions.

5.5 Interpretation and Intersubjectivity

Any study that is interested in interpreting the ways students explore their example spaces or the ways students go about example generation will be subject to certain issues with regards to the interpretation of data and the intersubjectivity of researchers and interviewees. Such issues are also present within much of mathematics education, especially in areas where qualitative approaches are taken (Ginsburg, 1981; Lester, 2005; Oliver et al., 2005), and Section 3.5 of this thesis discussed some of these in relation to phenomenography. No research will ever completely determine an individual's thought process, because to do this an individual must be conscious of the processes and be able to articulate these accurately to a researcher, which is clearly impossible even when the researcher is the individual.

In this section, 'issues of interpretation' are problems that may emerge when considering multiple meanings of events from the interviews. During the data collection and analysis described in this chapter and reported on in the next chapter, I have tried to bracket my thoughts and align my way of thinking with the student's, in the hope that the resulting categories of description have emerged from the data, rather than be biased towards only discerning aspects of which I was already aware. I have tried to both anticipate and articulate instances where an event or extract can be interpreted in multiple ways. The validation exercise discussed in Chapter 7 also addresses some of these concerns in more detail.

'Issues of intersubjectivity' are closely related to issues of interpretation, but are more philosophical in nature. Suppose we ask a participant to generate an example of a certain type of mathematical object. In order for the researcher to evaluate whether a suggested example is of the required type a definition is needed, which may be an informal one (such as a square in the context of elementary mathematics), or a formal mathematical one (an equivalence relation in the context of undergraduate mathematics). In the terminology introduced in Section 2.1.3, the researcher must consider whether the generated example is consistent with the researcher's personal concept definition of the object type. The researcher must assume further that their personal concept definition is compatible with the formal mathematical definition. This is before we even consider whether a formal mathematical definition, or knowledge in general, can exist beyond subjective cognitive construction (for an extensive debate in this area see Kieren, 2000; Lerman, 1996, 2000;

Steffe and Thompson, 2000). So although when interpreting interview data researchers can attempt to address issues of interpretation, issues of intersubjectivity are inherent.

Another issue of intersubjectivity arises when exploring a student's example space. When viewed from a student's perspective, an example generation task requires use of their personal concept definition (which they may or may not be explicitly aware of) when generating and checking whether an example is of the required type. In some cases, the student's concept definition will be identical to the researcher's and/or the formal concept definition, but in general this shall not be the case; a student may reason with elements of their concept image in conjunction with a concept definition (Vinner, 1991).

This means that a student's example space may contain objects which, although incompatible with interviewer's example space, are examples as far as the student is concerned (by the trivial argument that they are present in the student's evoked example space). Establishing whether such examples should be valid is now significantly more difficult. The approach taken in this thesis is to highlight and discuss instances where the validity of an example is debatable from an intersubjective perspective, drawing a conclusion from such discussion, but leaving the reader to determine whether they agree with such a conclusion (for instance, see Figure 4.4 and the associated section considering whether a graph of n^3 can be considered a real-valued sequence). Section 3.5 described how Phenomenography avoids some of the concerns about intersubjectivity by interpreting students' accounts of their thinking as representing 'ways of thinking that are possible', rather than 'what the student is exactly thinking'.

5.6 Ethical considerations

As with both pilot studies, ethical considerations for the main study included completion of Loughborough University's ethical clearance checklist. One question on this checklist prompted some consideration:

> [Does the study] involve procedures which are likely to cause physical, psychological, social or emotional distress to participants?

When completing the checklist for the pilot studies this question was answered 'no' because I did not think the task (and interview) situation would cause any physical, psychological, social or emotional distress. However, recall that in the second pilot study some students felt uncomfortable because they could not answer any question on the task (see Section 4.2.4). After reflecting on this question, and after discussion with my supervisor, I decided that for the main study the answer to this question would remain 'no', because being unable to answer a question, then working on it with guidance from a tutor is part of being an undergraduate student. However, I decided on a procedure to follow should the situation arise again. Initially a student would be encouraged to persevere (by saying that the questions are difficult and many people have struggled with them), but ultimately I would repeat what was done in the pilot study, changing the tone to that of a tutorial rather than a research interview if necessary.

The remainder of the questions on the ethical clearance checklist were more clearcut, and indicated that clearance was not required from the ethical advisory committee. After confirming that no further ethical approval was needed from Warwick University for either this study or the larger-scale validation study (see Chapter 8), I lodged a signed copy of the ethical clearance checklist with my head of department.

Another ethical decision was how to advertise the task to the students. Although students were compensated for their time with payment (a memory stick), students were also told that the interview experience would also be a good opportunity to receive feedback on their mathematical thinking. I do believe this to be the case, although it may have resulted in the students who volunteered for the study to be those that were struggling in comparison to their peers (this is discussed in more detail in Section 8.4.4, when the interviewed students are compared to the wider cohort of first year mathematics students at Warwick University).

The remainder of this section discusses informal reflections on ethical considerations on a phenomenographic task-based interview.

Unlike the interviews from the second pilot study, the interviews in the main study were undertaken from a phenomenographic perspective. Such a perspective involves the interviewer suspending judgement and 'bracketing' as much as possible so to reflect with the student on their understanding of a topic. In an interview, typically this was manifested by the interviewer acting as though incorrect mathematics was correct (relative to formal theory). There are clearly ethical dimensions to such a practice, and these were resolved as much as possible by explaining the situation to the student at the start of the interview. I would typically explain that “as you answer the questions, and when we discuss them afterwards, I don’t want to say if you are right or wrong. This is because I’m interested in you describing how you are thinking and I don’t want to influence the way you are thinking. I am happy to go through your answers properly at the end of the interview.” Although the wording of this statement changed from student to student, each interview introduction included the sentiment of this statement.

Furthermore, phenomenographic studies typically focus on ‘why’ questions, and quite often a student will feel unable to articulate their thinking to a level they themselves feel satisfied with (Åkerlind, 2005b, p.115). This may mean that it is effortful and potentially tiring for students to reflect deeply on their thought processes, and why their answers are correct. I therefore attempted to make the interview as pleasant an experience as possible for the student, for instance each interview began with a ‘chat’ period designed to allow a rapport to be established between myself and the student (see Section 5.2.3).

5.7 Summary of chapter

This chapter has described the planning and data analysis of the main study. Designed to address the second research question of the thesis, and to provide more evidence on the first research question, the study is framed within the methodology of phenomenography, as presented in Chapter 3.

In the study, fifteen students from Warwick University were individually interviewed. During these interviews, each student was given a modified version of the example generation task used in the pilot studies in Chapter 4. Each student discussed their thoughts as they completed the task via a ‘think-aloud’ protocol, and later during an informal discussion period with the interviewer. Throughout the each interview, the interviewer attempted to ‘bracket’ their judgement on the content of the mathematics, asking questions that requested more information and those which encouraged reflection on ideas raised by the student.

The data analysis procedure began by transcribing the entire set of interviews. The transcripts were then phenomenographically analysed by first open coding the data, then collecting of codes under headings, and finally forming of categories of description and dimensions of variation. The next chapter reports on the dimensions of variation which emerged from this data analysis.

Chapter 6

Main Study: Dimensions of Variation

This chapter addresses the second research question:

> What is the qualitative variation in students' experiences of sequence generation?

As described in the previous chapter, the thesis aims to consider qualitative variation of experience from the research specialism of phenomenography. Four dimensions of variation are presented in this chapter: Using Definitions, Representation of Sequences, Sequence Construction Strategies, and Justifications. Each of these dimensions emerged from the phenomenographic data analysis, as described in Chapter 5. This chapter begins with a brief discussion of the variation present in the definitions and questions from the perspective of a researcher in Section 6.1. Then, in Sections 6.2–6.5, the chapter outlines each dimension of variation by discussing the categories of variation present, and presenting and commenting upon illustrative incidents from the data.

Each section focusing on a dimension of variation begins with a general discussion of the emergence of the dimension, and its relation to the second research question. A summary table of the categories of description is then presented (recall that a dimension of variation has a hierarchical structure, beginning with categories including few features of the concept to categories which describe richer or deeper ways of seeing the concept). Then, each category of description is broken down into types of incident, with illustrations of the category from the data. These incidents are only considered from within the context of the relevant dimension of variation, postponing comparison across dimensions of variation until the end of this chapter. Finally, each section concludes with a summary of the dimension.

Within a particular report of an incident, quotes present the data in the order in which events occurred. If the quote is taken from when the student was answering a question during the task phase of the interview, the quoted passage begins by stating the question under consideration, then what the student said, then finally the student's answer:

> *Q1. A strictly increasing sequence.*
>
> **Student:** I think this would work.
>
> [Answer given: $a_n = n$]

If the quote is taken from discussion period after the task had been completed, the student's answer is presented immediately after the question, and then the discussion between interviewer and student is given.

> *Q1. A strictly increasing sequence.*
>
> [Answer given: $a_n = n$]
>
> **Interviewer:** Why did you write that?
>
> **Student:** I thought it would work.

6.1 Variation in the questions

Although the purpose of a phenomenographic study such as this is to explore students' awareness of sequence generation, it is important to also consider the variation present in the questions students were asked. Phenomenographers usually subscribe to Marton and Booth's (1997) belief that learning is the same as becoming aware of possible variation in the object of study, and so before considering the dimensions of variation in students' awareness it is first worthwhile therefore to first consider the variation that is present in questions and the definitions under consideration.

Such an analysis relies on my thoughts as a researcher, rather from empirical data, and so such an account is more phenomenological than phenomenographical. Recently, some authors have researched the variation in examples from a similar perspective (e.g. Mason, in press; Watson and Chick, in press). Note that here it is variation in the definitions and example generation questions that is considered, rather than variation in the examples themselves. Also, the variation discussed here is the result of my reflections before and during the data analysis; I make no claim that the variation of sequence property types discussed in this section is the only way in which variation of sequence properties can be considered.

The definitions presented to students on the definition sheet are superficially similar (this definition sheet was reproduced as Figure 5.1). Many definitions from this sheet make some stipulation on the terms of the sequence based on the ordering of their subscripts, and many contain either or both of the quantifiers $\forall$ and $\exists$. It is understandable therefore that when students are considering a definition together with a sequence, they misinterpret which features of the sequence the definition constrains.

The features that the sequence properties constrain fall into three types. Some, such as increasing and strictly increasing, give a rule that dictate a condition on subsequent terms that must hold for each term of the sequence. Some definitions do not dictate a sequence's term by term behaviour, but provide condition for all terms, such as those involving bounding a sequence. Other definitions do not stipulate the behaviour of a finite number of terms, but give conditions for the long term behaviour of the sequence, such as the definition of tending to infinity. In summary, I consider there to be roughly three different 'types' of definition on the sheet:

> *Term-by-term (pairwise) definitions* [T-T]
> These define the behaviour of the sequence as it moves from term to term. Longer term behaviour can sometimes be inferred.
> Such as: increasing, decreasing, strictly increasing, strictly decreasing, monotonic
>
> *Sequence-wide properties* [S-W]
> These are definitions that specify a rule (which may involve a universal quantifier) for all terms in the sequence, treating each term in isolation.
> Such as: bounded above, bounded below, upper bound, lower bound, bounded
>
> *Long-term properties* [L-T]
> These give a rule (which may involve a universal quantifier) which must be

satisfied from (or at) some point in the sequence.

Such as: tending to infinity, diverges

In terms of these definition types, each question asks students to use a single definition, or to combine the same or different types. Looking at each question on the task sheet:

1. A strictly increasing sequence [T-T]
2. An increasing sequence that is not strictly increasing [T-T/T-T]
3. A sequence that is both increasing and decreasing [T-T/T-T]
4. A sequence that is neither increasing nor decreasing [T-T/T-T]
5. A sequence that has no upper bound [S-W]
6. A sequence that has neither an upper bound nor a lower bound [S-W/S-W]
7. A bounded, monotonic sequence [T-T/S-W]
8. A sequence that tends to infinity [L-T]
9. A sequence that tends to infinity that is not increasing [T-T/L-T]
10. A sequence that tends to infinity that is not bounded below [S-W/L-T]
11. A strictly increasing sequence that does not tend to infinity [T-T/L-T]

Chapter 8, which reports on an analysis of the task when given to a large group of students, will note that questions combining different types of properties were more difficult than those which combined a single type, which in turn were more difficult than those which asked students to consider a single definition (in Section 8.4.2). In this chapter, in the following section which discusses the dimensions of variation Using Definitions, we will see that some issues students had may have resulted from treating one type of definitions as if it were another.

6.2 Using definitions (Def)

The focus of this dimension of variation is students' awareness of (that is, their use of and comments on) the definitions of sequence properties in the task. The focus is not extended to the way students deal with the definition of a sequence (this awareness is covered in the Representation of Sequences dimension of variation).

Recall that students were presented with a comprehensive definition sheet immediately before the task phase of the interview. It is unsurprising therefore that many students made use of these definitions in some way or another during the task. However, from the perspective of the second research question, students were aware of different aspects of the definitions, and how they can be applied when generating examples of sequences.

At one extreme we have the category of description Def-A *Unaware of Definitions*, where students rely almost wholly on non-mathematical spontaneous conceptions, taking the 'everyday' meaning of words in the place of a sequence property, much as (Cornu, 1991) observed was the case when some students dealt with the concept of a limit. At the other extreme, in category of description Def-D *Manipulates Definitions*, students use and manipulate aspects from the definitions both when constructing their sequences and when justifying why their sequence is correct. Between these categories of description, there are instances of students referring to definitions, but with no evidence that they are using them in any way (Def-B *Refers to Definitions*), and those students who use definitions, but do not demonstrate that they are treating them as objects (Def-C *Uses Definitions*).

An outline of this dimension of variation is given in Table 6.1.

Table 6.1: Dimension of Variation: Using definitions.

	Description
Def-A.	*Unaware of Definitions*
Def-B.	*Refers to Definitions*
Def-C.	*Uses Definitions*
Def-D.	*Manipulates Definitions*

6.2.1 Def-A. Unaware of definitions

This category of description originates from statements and actions which were coded to suggest that students were not aware of how the definitions were related to the example generation task (and so from a phenomenographic perspective possibly also less aware of the role of definitions in mathematics more generally). Incidents which were classified Def-A were often the result of a failed attempt by student to interact with a definition after they had been prompted to do so by the interviewer.

Other incidents of being unaware of definitions involve students who used the everyday meaning of the properties in place of the formal definitions. Some of these everyday meanings had little in common with the formal definition. A summary of the types of incident that are in this category include students who:

- Refers to definitions only when prompted
- Used spontaneous conceptions related to the everyday meaning of a property in place of a formal definition
- Ignored / not referred to the definitions[1]

Incident Def-A1: Refers to definitions only when prompted

Consider the following incident that took place during the discussion period with Ian. During the task, Ian had not referred to or commented upon the definitions:

> *Q1. A strictly increasing sequence*
> [Answer given: $a_n = n \quad \text{where} \quad n \in \mathbb{N}$]
>
> *Q2. An increasing sequence that is not strictly increasing*
> [Answer given: $1, 2, 1, 3, 1, 4, 1$]
>
> **Interviewer:** Do you want to just have a quick read of the definitions of increasing and strictly increasing, and have another quick look at Questions 1 and 2. Do you follow what the definitions are saying here?

[1] There is no illustrative incident corresponding to this event, it is a general observation from the data.

> **Ian:** Yeah, it's just that the small thing between two successive terms can be equal with increasing, but with strictly increasing they are always just/
>
> **Interviewer:** So looking again to your answers to [Question] 1 and 2, are you still comfortable with them or not?
>
> **Ian:** No, Question 2's wrong, because, well, yeah I'm happy with increasing, because every successive term's going to be greater than the previous term, and [29 second pause]. So I suppose you need something like, some example of a term being the same so it's not always increasing
> [Changes answer to $1, 1, 2, 2, 3, 3$].

Ian had demonstrated that although he had not referred to the definitions during the task (and had, perhaps as a consequence, given incorrect answers), when prompted to recall definitions he was capable of applying them successfully. Such incidents imply that even if an incident has been categorised as Def-A, it does not necessarily follow that the student is unable to recall and apply definitions in all situations.

Incident Def-A2: Using everyday meanings of properties in place of formal definitions

A well-documented type of phenomenon that is categorised within Def-A is when the the everyday meaning of words are used as the concept definition for a property. Cornu (1991, p.154) noted that when students deal with formal definitions of mathematical concepts their prior knowledge, for instance the everyday meaning of the words, can mix with the mathematical definition resulting in a potentially conflicting concept image. This prior knowledge which may or may not conflict with the formal definition is referred to as a spontaneous conception. Incidents in Def-A2 make no reference to the formal definition at all; students' use of the property is wholly composed of their spontaneous conceptions (in other words no reference is made to the formal definitions at all).

Looking back at the results from the pilot studies, it was clear that for certain students, the formal properties of increasing and decreasing were synonymous with "going up" and "going down". Some students in the main study also used the terms in a similar way, such as when Edha was answering Question 3 in the task period:

Q3. A sequence that is both increasing and decreasing

Edha: Erm, both increasing and decreasing, sine or cos curve?

Interviewer: What made you think of the sine or cos curve?

Edha: It just keeps going up and down, but I don't think that's right. I'm not too sure.
[Answer given: $a_n = (-1)^n(n+1)$]

David had similar thoughts:

David: Both increasing and decreasing. For that one then, the first one I immediately think of it, lets go for $\sin(n)$.
[Answer given: $a_n = \sin(n)$]

Such everyday meanings may be mathematical, but not consistent with formal mathematics, such as David's answer to Question 6:

Q6. A sequence that has neither an upper bound nor a lower bound

David: Ok. That's quite a tough one actually. Because I'm thinking the first term of the sequence is normally going to be the lower bound.

The use of everyday meanings may also have repercussions when negating properties, for instance when Valter commented:

Valter: If a sequence is not increasing then it must be either oscillating or decreasing.

Interviewer: What do you mean by oscillate?

Valter: Erm, it's neither increasing nor decreasing.

This everyday negation of increasing stated here is not contradictory with the negation of the formal definition of increasing (i.e. that there is at least one term which is lower than its predecessor), but it relies on defining another property—oscillating—in a circular argument.

6.2.2 Def-B. Refers to definitions

It was rare for there to be no reference to mathematical definitions, but this did not mean that students used the formal definitions correctly, or remained consistent in their use of definitions. Most students made reference to the contrasting definitions when moving from Question 1 to Question 2, as these questions deal with very similar properties. This is different to actually using the definition, whether it be during example generation or validation of answers. Other students referred to definitions only when they got stuck with the task, checking for instance that their notion of tends to infinity was the same as the one on the sheet.

In general, for an incident to be classified Def-B, either an improper (but mathematically-based) definition was used, or the definition was not used in a consistent way. The types of event which are in this category include:

- Modifying definitions with everyday spontaneous conceptions
- Modifying definitions with spontaneous conceptions from mathematics
- Referring to definitions without using them when answering the questions
- May look at definition for inspiration but does not use later

Incident Def-B1: Modifying definitions with everyday spontaneous conceptions

Incidents of this nature are similar to those described in incident Def-A2, but whereas in Def-A the entire definition was taken to be the same as an everyday meaning, incidents in Def-B typically take the formal definition as a base, but then include spontaneous conceptions which add to, or otherwise modify the definition.

In the last section, we considered David's comments behind his answer to Question 3 as an example of Def-A. If we now consider this in the wider context of his actions during the first four questions we can see that his usage of increasing and strictly increasing are in keeping with the formal definitions, but once he considers increasing and decreasing together the everyday meaning of the words begin to modify his use of the formal definition of increasing:

Q1. A strictly increasing sequence

David: Ok. So the first one is pretty simple. Just straightforward progression of the natural numbers.
[Answer given: $a_n = n$]

Q2. An increasing sequence that is not strictly increasing

David: Second one. Erm, Ok so I'm trying to think of a sequence that have some terms that are equal to each other maybe. Or a whole sequence that is all equal to each other. To show it's increasing, but not strictly[...] I'm not sure if that qualifies, that's the thing. Obviously all the terms are equal and then there are no increasing terms. Erm, OK. I'll leave that how it is.
[Answer given: $a_n = 1$]

Q3. A sequence which is both increasing and decreasing

David: Both increasing and decreasing. For that one then, the first one I immediately think of it, lets go for $\sin(n)$.
[Answer given: $a_n = \sin(n)$]

Q4. A sequence that is neither increasing nor decreasing

David: Ok. Haha, yeah that one's going to be my straight-line one. Just a constant at k, just for a bit of rigour. Yeah, so that one [points at Question 2] I'm going to have to rethink, but there we go.
[Answer given: $a_n = k \quad k \in \mathbb{R}$]

At this point, it seems that provided David considers increasing and decreasing in isolation from each other he is happy with the view that a constant sequence is increasing (and so presumably, also considered in isolation, decreasing). When it comes to combining the two properties, his viewpoint is shifted to fit in with the spontaneous conception that 'increasing and decreasing' means 'going up and down,' and so the sequence $a_n = \sin(n)$ is increasing and decreasing, and the constant sequence $a_n = k$, where $k \in \mathbb{R}$ is neither increasing nor decreasing.

David did not comment on the inconsistency of his answers at this point, but did so later on, after first attempting the other questions on the task:

Q2. An increasing sequence that is not strictly increasing

David: Ok, that's, this one here increasing but not strictly increasing. I want some level terms really.

Interviewer: You don't have to give a formula if you don't want to, if you can make it obvious what the pattern is from the/

David: I'm going to go for — I wonder if you can do this
[Changes answer to $a_n = n$ for $n < 5$, $a_n = 5$ for $n \geq 5$]

David had not explicitly recalled a formal definition during the task, but from his initial and subsequent attempt at Question 2 it appears that he understood the distinction between an increasing sequence and a strictly increasing sequence. However, once he decided that a constant sequence is neither increasing nor decreasing (c.f. his answer to Question 4) he concludes that a constant sequence is therefore not suitable for Question 2 either. He then decided to modify his (correct) answer to Question 2 to a different (correct) answer which conflicts less with his answer to Question 4. This was raised during the discussion period:

Interviewer: And so for this one here [Question 2]- you were quite happy about that one [$a_n = 1$] and that one [$a_n = n$ for $n < 5$, $a_n = 5$ for $n \geq 5$] satisfying it, but then that one you were a bit/

David: Yeah, because obviously for an increasing one that's not strictly increasing you can have terms that are level. But all the terms being level just doesn't seem to qualify for an increasing sequence, just by the meaning of the word.

Here we have seen an incident where David initially treated the sequence properties in a way which was correct relative to the formal definitions, but then his spontaneous conceptions (e.g. that a sequence which is both increasing and decreasing goes up and down) has modified his initial understanding of the definitions.

Incident Def-B2: Modifying definitions with spontaneous conceptions from mathematics

The spontaneous conception outlined in the last incident has an everyday use, whereas some students demonstrated that spontaneous conceptions could have a more mathematical basis. In the following quote from Ben, he takes the definition of an upper bound, and modifies the constraint that an upper bound be a real number, allowing an upper bound to have the value infinity:

> *Q6. A sequence that has neither an upper bound nor a lower bound*
>
> **Ben:** Ok so here we want one that has a U at infinity and minus infinity to get both of those. So you could have $-1, 2, -2, 3, -3$.
> [Answer given: $1, -1, 2, -2, 3, -3\ldots$]

Incident Def-B3: Referring to definitions without using them when answering the questions

This incident explores comments that relate to students who made reference to the existence of definitions, but did not necessarily use an appropriate definition. Typically such students also did not apply these definitions appropriately. One common instance was using definitions as a justification for the impossibility of the combination of properties requested by Question 9. Here are a collection of quotes to illustrate this, taken from both the task and discussion periods:

> *Q9. A sequence that tends to infinity that is not increasing*
>
> **Guan:** If it tends to infinity it has to be increasing.
>
> **Ian:** I suppose, by the definition of infinity, because there's always going to have to be terms that exceed a limit so you just always seem to be increasing, I don't know, all the terms that I think of are going to infinity in other ways, always increasing.
>
> **Joe:** Because otherwise, if it tends to infinity and it's positive it has to increase.
>
> **Ken:** A sequence that tends to infinity that is not increasing. I think that's impossible. By definition.

Some of these quotes only refer to definitions implicitly (such as the one from Guan, which depending on context could be classified Def-A also), but the reasoning these students report make logical deductions based on what it means to tend to infinity (and so these quotes are considered as belonging to the students' concept definitions of tending to infinity).

Incident Def-B4: Mixing sequence-wide, long-term, and term-by-term properties

Recall from Section (6.1) that from this researchers' perspective, there is variation in the way a definition constrains the terms of a sequence; whether they constrain term-by-term (T-T), sequence-wide (S-W), or in long-term behaviour (L-T). Treating one of these definition types as another is another way in which a student may refer to a definition whilst still not use it appropriately:

For instance, consider the following comment from Phalgun, who is treating tending to infinity and bounded below as if they were term-by-term properties:

> *Q10. A sequence that tends to infinity that is not bounded below*
>
> **Phalgun:** Something going up and down at the same time

It is easily argued that Phalgun has the spontaneous conceptions 'tending to infinity is going up' and 'not bounded below is going down', but from the perspective of definition types Phalgun may be treating tending to infinity and being bounded below as T-T property.

The following quote can be interpreted as Ken treating increasing as a S-W property:

> *Q4. A sequence that is neither increasing nor decreasing*
> [Answer given: $(-1)^n$]
>
> **Ken:** It's not getting larger, it's just going from 1 to -1

Here, one interpretation of Ken's statement "it's not getting larger" is that it is an implicit reference to the definition of increasing, only the definition is being treated as if it were a sequence wide property, rather than a term-by-term one.

6.2.3 Def-C. Uses definitions

In this category students not only make reference to mathematical definitions when constructing or validating answers, they also provide evidence that they are using the example in a way that is consistent with how definitions are used in formal mathematics. Particular features of a definition may be focussed on (such as the $\forall$ quantifier), or the meaning of the definition may be attended to. This category also includes instances where a student uses a definition, but inappropriately (see incident Def-C3).

The type of incidents categorised as Def-C were:

- Using definitions to confirm answers are correct
- Using definitions when constructing answers
- Mixing sequence-wide, long-term, and term-by-term properties

Incident Def-C1: Using definitions to confirm answers are correct

Incidents categorised Def-C were almost exclusively related to the sequence properties of increasing and decreasing, typically when students justified that their answers to Question 1 and 2 were correct.

> *Q1. A strictly increasing sequence*
>
> [Answer given: $a_n = n$]
>
> **Haaroon:** So it's strictly increasing so it's always going to go above the next term, so it just needs to go up.

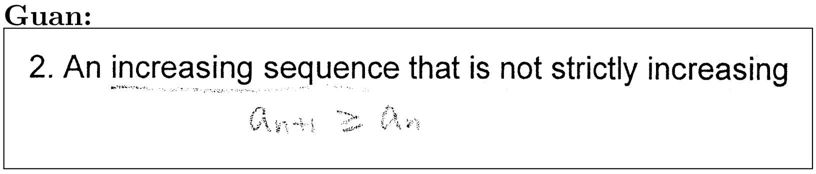
> *Q2. An increasing sequence that is not strictly increasing*
>
> **Nicola:** I was thinking for this one you need one that's going to, at some point plateau, stay at the same point, 'Cus it's not strictly increasing.
> [Answer given: $a_n = 1, 1, 2, 2, 3, 3,$]

These incidents may seem superficially similar to the incidents described in Def-B3, where definitions were referred to when justifying that answers were impossible. However, the types of incident shown in Def-B3 did not make reference, or use, the content of the definitions, they were instead using the language of formal mathematics but not in a way that was consistent with the meaning of the formal properties (such as Guan who said that 'if it tends to infinity it has to be increasing'). Here, in the incidents classified as Def-C, the both the definitions themselves and the framework of their use are consistent with formal mathematics; implicit reference is made to the '$\forall n \in \mathbb{N}$, $a_{n+1} \geq a_n$' part of the definition of increasing.

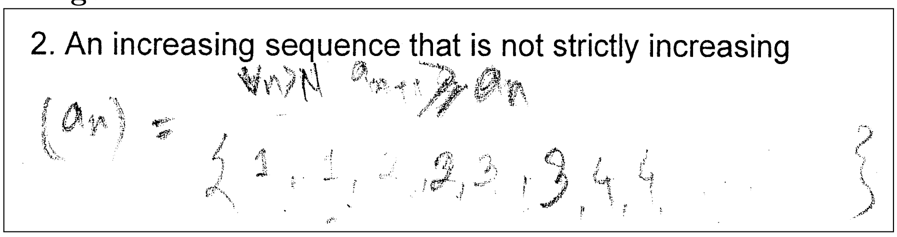
Incident Def-C2: Using definitions when constructing answers

Some students chose to write the appropriate definitions down near the questions, indicating that they were trying to keep the definitions in focus when answering the questions:

Guan:



Phalgun:



Others looked at the definition sheet to check definitions, or because they had forgotten the meaning of a particular property:

[Nicola looks at definition sheet]

Interviewer: Which one did you just look at?

Nicola: The monotonic. Increasing or decreasing.

In these incidents students were making explicit reference to the content of a definition when completing the task.

Incident Def-C3: Mixing the content of sequence-wide, long-term, and term-by-term properties

The type of incident presented here is similar to those in Def-B4, where a definition of a certain type (T-T. S-W, L-T) was treated as if it were another. In Def-B4, it was the idea of what the definition did, perhaps making reference to the definition but not using the content of the definition at all. Here we present an incident where Valter certainly uses the content of the definition, but the way he works with the definition suggests that he is treating 'not increasing' as a T-T definition (where in fact it is a S-W property).

I first present the incident, summarise it, and then explain why it was categorised Def-C.

> *Q9. A sequence that tends to infinity that is not increasing.*
> [Answer given: impossible]
>
> **Valter:** It might be easier to use the definition, tending to infinity in graphical way. You have C here which is an arbitrary value bigger than zero, and we know that for a sequence to tend to infinity every term of the sequence after a certain N must be greater than this C. For all n. For all C. So if we take another one higher then we can find a corresponding N_2. So we keep on choosing Cs and have to look for corresponding Ns, and one can see that the terms of the sequence have to increase at some point.

In summary, Valter first considers the definition of tending to infinity, in particular noting that it contains a C, which is fixed, but arbitrary. For the particular value, C_1, he notes that there is an N_1 where for all $n > N_1$, $a_n > C_1$. He then chooses a larger value $C_2 > C_1$, and defines a N_2, which without loss of generality we can assume $N_2 > N_1$. Then he claims the sequence of (N_i) must increase. This summary is illustrated in Figure 6.1.

Here, I argue there are two alternative interpretations. The first is that Valter is effectively negating increasing to mean 'can never increase', i.e.

Definition (false): A sequence $(a_n)_{n=1}^{\infty}$ is *not increasing* if and only if $\forall n \in \mathbb{N}$, $a_{n+1} \leq a_n$.

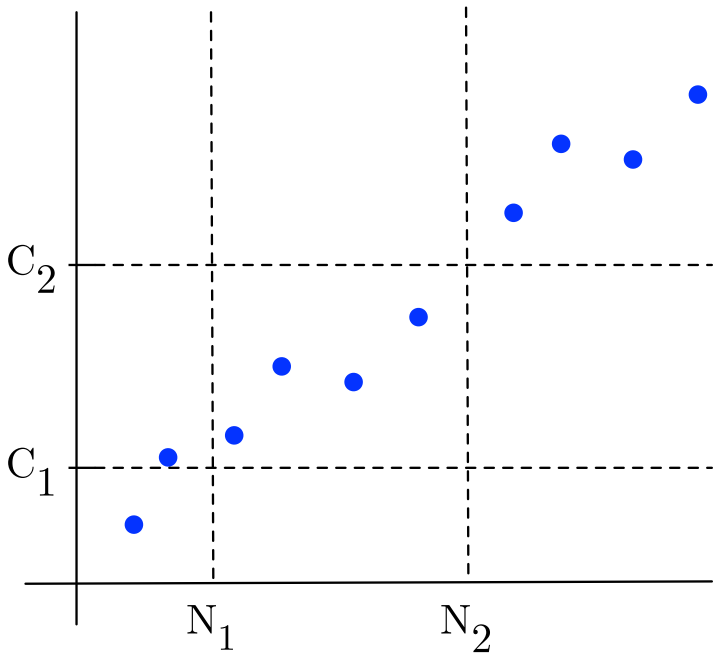


FIGURE 6.1: An illustration of an alternative (false) definition of increasing, i.e. that the sequence tends to its supremum.

and so his version of the definition of a 'not increasing' sequence is a T-T property. An alternative interpretation is that Valter is using 'increasing' to mean the L-T property (illustrated in Figure 6.1):

Definition (false): Let $X = \sup\{a_n, n \in \mathbb{N}\}$. A sequence $(a_n)_{n=1}^{\infty}$ is *increasing* if and only if $\forall C < X$, $\exists N \in \mathbb{N}$ s.t. $\forall n > N$, $a_n \geq C$.

This definition is perhaps more in keeping with the everyday meaning of increasing (the sequence is 'going up' over time), but it includes sequences which formally are not increasing. Despite the similarity with the everyday spontaneous conception, this incident was not categorised Def-B because Valter is doing more than referring to definitions, he is taking elements from the definition of tending to infinity and attempting to use these with the definition of increasing.

For the incident discussed here, and in earlier in Def-B4, it is often difficult to clearly categorise clearly into either Def-B or Def-C. This is because from one perspective (Def-B), these incidents are similar to Incident Def-B1/2 where spontaneous conceptions from everyday usage and other mathematical knowledge mean that students misuse the definitions. However, it is possible to interpret some passages as illustrations of Def-C, as we did here, because in some sense the students are using the way one deals with a certain definition type, but for a type which is dissimilar.

6.2.4 Def-D. Manipulates definitions

In this category, definitions are used as objects than can be used in conjunction with each other and restated in different ways or from different perspectives whilst not adding to, or modifying their meaning. Differences and similarities between definitions may be explored. The types of incident categorised as Def-D were:

- Making distinctions between similar definitions
- Negating definitions
- Combine definitions and using them to draw conclusions
- Reformulate definitions in everyday language or in formal mathematics, without changing meaning

Incident Def-D1: Making distinctions between similar definitions

It could be argued that the design of the task encourages students to focus on this type of activity. For instance, in order to move successfully from Question 1 to 2, a students needed to make some attempt to compare the definitions of increasing and strictly increasing, and many students did this explicitly. Two such students were Ben and Guan, who discussed the distinction between increasing and strictly increasing when answering Question 2, and appreciated that not all terms in the sequence need to be repeated:

> *Q2. An increasing sequence that is not strictly increasing*
>
> **Ben:** Again, these [sequences] are ones that can be equal to the ones before because that's the difference in the definition, so you could have say 1,1,2,2,3,3. Any combination of those really.
> [Answer given: $a_n = 1, 1, 2, 3, 3,$ (sic)]

> **Guan:** I'm thinking about the definition of increasing sequence and strictly increasing. So the difference is that strictly increasing the next term must be greater than the previous one, while increasing can be greater than or equal to the previous one. So it can be a constant sequence.
> [Answer given: $a_n = 2, 2, 2, 2, 2, \dots,$]

Incident Def-D2: Negating definitions

Successful completion of the task sheet required students to work with the negation of some definitions. Whilst some students did not manage to do this, instead taking the everyday negation of the statements (sometimes successfully and sometimes unsuccessfully), other students described a more formal negation, such as Valter who was asked why his answer to Question 4 was correct:

> *Q4. A sequence that is neither increasing nor decreasing.*
> [Answer given: $a_n = (-1)^n$]
>
> **Valter:** This follows straight from the definition of increasing sequence. The condition is that each term a_{n+1} is greater than or equal to the proceeding one a_n, which must be true for all n. But we can see that this is not true for all n. We take this term, a_2 and the next term a_3 is less than, so the sequence is not increasing. And by similar reasoning we can conclude that it is also not decreasing.

Incident Def-D3: Combining definitions and using them to draw conclusions

It was rare for a student to combine definitions at a level where they made clear which aspects of the definitions were being combined. When a students did attempt this type of detailed examination of a definition it was usually tough to extract meaning from their often meandering explanations (c.f. Incident Def-C3). Sometimes students combining sequence properties with little reference to the formal definitions (Incident Def-C1). However, Oksana came close to using definitions as part of a formal mathematical argument:

Q10. A sequence that tends to infinity that is not bounded below

[Answer given: impossible]

Oskana: The definition of tending to infinity is that you can find a, that there's always an N such that $a_N > C$ given a C, but you know that there's always going to be a negative term after you find— say you say a C. There's no way there's not going to be a negative term after the a_N term, so you're not going to be able to find a_N term such that it's greater than C for all $a_n > N$. I'm sorry, I'm definitely sure that it's not possible.

Because these were spoken rather than written there is a degree of interpretation as to whether "n" was n or N. In the above quotation, for the categorisation Def-D, Oskana has been given the benefit of the doubt, and her statement is assumed to be equivalent to the more more formal:

- From the definition of increasing to infinity, we know that for any $C > 0$, we can find N, where $a_n > C$ for all $n > N$.
- Starting at this N, we know that there must be a natural number $m > N$ where the corresponding sequence term $a_m < 0$, because otherwise the sequence would be bounded below.
- Therefore the sequence can't both tend to infinity and be bounded below.

If this is a valid interpretation of Oskana's argument, then she is using the definitions of tending to infinity and bounded below together, selecting the parts which contradict each other.

Incident Def-D4: Reformulate definitions in everyday language or in formal mathematics, without changing meaning

In incident Def-C3, Valter used the definition of tending to infinity to construct the increasing sequence (N_i). This incident was categorised Def-C because he went on to use this reformulated definition to conclude that the elements of the sequence (a_n) were increasing. However, if we focus on the reformulation part alone, Valter's reformulation lost none of the meaning of the definition of tending to infinity.

6.2.5 Commentary on definition usage

As was discussed in the introduction to this dimension of variation section, because students were presented with the definition sheet before attempting the task, the information that the task can provide with regard to how students are spontaneously aware about mathematical definitions is limited. From a phenomenographical perspective, the sheet has immediately shifted students' focus towards property definitions before they had even begun the task, and so it is not immediately clear that drawing conclusions about students awareness of property definitions is completely valid in this context. Furthermore, the task sheet also implicitly encouraged formal definition use via its content, and by the request that students 'think-aloud'. However, despite these concerns, some students were reluctant to consider definitions during the task until prompted (Def-A), and several relied on spontaneous conceptions rather than use or refer to the formal definitions (Def-B, Def-C). Few students treated them as objects that they could manipulate, combine, and reformulate without losing meaning (Def-D).

It is also worth noting that most of the incidents reported here focus on the sequence properties of increasing and decreasing, rather than boundedness. This does reflect the data, where most of the questions involving boundedness did not evoke statements or actions that referenced the corresponding definitions. Perhaps this is a reflection on the similarity between the everyday meaning of 'bounded above' and its formal mathematical counterpart, so maybe the difficulty with applying a definition is directly related to how similar it is to everyday spontaneous conceptions (this hypothesis is discussed further in Section 9.5).

Excerpts from the data that correspond to this dimension of variation can be tough to categorise because, in an example generation interview situation, implicit use of examples is superficially similar to not using definitions at all. There are three specific instances where the categorisation of a quote relies on a high degree of (possibly alternative) interpretation.

1. If a student makes no reference to definitions then this is not necessarily an indication that at that time that their behaviour is characteristic of category Def-A. They could be using definitions internally or implicitly in a way which, if articulated would be categorised Def-C or Def-D.

2. There were times when students claimed (a) the sequence satisfied the properties requested or (b) that there was no sequence satisfying the properties "by definition." This could be the result of using definitions in a either a way consistent with (formal) mathematics (Def-D), or in a more colloquial sense (Def-B).

3. As illustrated by Incident Def-C3, it is not clear whether a misconception related to a definition is the result of a spontaneous conception (which may result from the everyday meaning of a word, or prior mathematical knowledge), or treating different definition types (T-T, S-W and L-T) interchangeably.

Within a phenomenographic account these three issues are less troublesome. This is because the second research question is not concerned with labelling specific incidents, but describing the possible ways of experiencing, which the 'using definitions' dimensions has done.

6.3 Representation of sequences (Rep)

This dimension of variation considers students' awareness of the ways students chose to represent their answers, the types of objects students considered to be representations of sequences, and which, if any, representations are seen as 'better' answers to the sequence example generation questions. From the perspective of the second research question, these three areas are very similar, if not the same. For instance, if a student answers a sequence example generation question by choosing a type of mathematical object (which may or may not be a sequence), then it is possible to be aware of that object as a sequence. Similarly, if a student believes a certain type of mathematical object to represent a sequence, then it is possible such an answer can be given as a response to a sequence example generation task.

The dimension begins with category of description Rep-A *Any Representation is Suitable*, where students do not place any constraints on the way their sequence is represented. Even mathematical objects which are not functions from the natural numbers to the reals might be considered as sequences within incidents from this category of description, which has a clear link to the **INS** answers given by students in the two pilot studies in Chapter 4. In the next category of description, Rep-B *One Representation is Superior*, students recognise that certain objects are not sequences, but have a preference of one particular representation of sequence for their answers. Such students sometimes can generate a sequence, but cannot write it down in the required form, and so may leave the question blank. In the most sophisticated category of description, Rep-C *Any Well-Defined Representation is Suitable*, students are flexible in the way they write down their answers, but ensure that their answers are sequences.

An outline of the dimension is given in Table 6.2.

TABLE 6.2: Dimension of Variation: Representation of sequences.

	Description
Rep-A.	*Any Representation is Suitable*
Rep-B.	*One Representation is Superior*
Rep-C.	*Any Well-Defined Representation is Suitable*

6.3.1 Rep-A. Any representation is suitable

In the pilot study tasks it was common for students to give examples of mathematical objects which were not sequences but possibly did have the properties requested by the question in some loose sense. Such answers were classified **INS** in the coding scheme from the pilot studies. These objects fall into this category; they are representations of objects given when students were attempting to provide a example of a sequence satisfying certain properties.

Compared with the pilot studies, there were considerably fewer answers of the **INS** type from students in the main study (3 from 151 non-blank answers), which is remarkable given that the interview style, structure and content was relatively unchanged. Despite few students giving **INS** answers, several students did consider briefly offering objects which were not sequences, but corrected themselves before writing down their answer. Overall, however, there was far less variation observed amongst the **INS** answers, in particular there were no 'mathematical nonsense' answers, or objects which were not functions (such as intervals and sets).

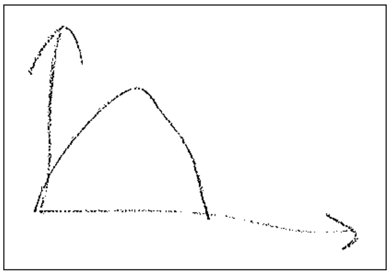
The types of incidents categorised as Rep-A were:

- **INS** answers generally[2]
- Changing the domain of a sequence
- Considering giving an answer of **INS** type

Incident Rep-A1: Giving a finite sequence

As described in the previous paragraph, it was rare for a student to settle on an **INS** response as a final answer but when such an event happened it was because the student had changed the domain of a sequence. One student, Anna, gave an answer which was a real function over a finite interval of the reals. The rest of her answers were well-defined sequences (or she stated that the combination of properties was impossible). During the discussion period this issue was raised:

Q3. A sequence that is both increasing and decreasing
[Answer given: $a_n = \sin n$, $0 \leq n \leq \pi$, with sketch (below)]



Interviewer: Out of interest, some of your sequences have gone on forever and some of them have stopped. I'm interested in why some of them did that and some didn't.

Anna: Like where?

Interviewer: Your sine one, that stopped, and I think the rest carried on, is that right? Apart from the ones which were impossible.

Anna: I just showed that it was increasing there and decreasing there [refers to the increasing and decreasing sections of the graph].

[2] as noted above, there were few **INS** answers in the main study, but the prevalence of these answers in the pilot studies warranted inclusion of this incident type.

In examining this incident within the Rep dimension of variation we must focus on the representation of Anna's answer, bracketing other dimensions of variation, in particular the fact that even if we consider the extension of the function to all reals, and then the projection on to the natural numbers it is still not increasing and decreasing.

Anna may or may not have considered her answer to be a sequence in its own right. She wrote "a_n", which may suggest that she was using $n \in \mathbb{N}$ as was the convention in her course. Later however, she drew the equivalent continuous function over the interval $[0, \pi]$. In the discussion period she demonstrated that she believed that this answer characterised the important features of a sequence answering this question, namely that the graph increases for a region and then decreases for a region, and so (within her frame of reference) it is an increasing and decreasing sequence.

From a phenomenographic perspective, by giving this answer to a sequence example generation task, Anna has demonstrated that some students may believe a sequence can be represented in a way which is not equivalent to a function from the natural numbers to the reals.

Incident Rep-A2: Changing the domain to $\mathbb{Z}$

As noted in the pilot study, some of the task questions which request combinations of properties that are impossible within formal mathematics can be considered possible if the definition of a sequence is altered to include double-sided sequence, that is a function $f : \mathbb{Z} \to \mathbb{R}$. In the interviews, Ian did precisely this:

> *Q10. A sequence that tends to infinity that is not bounded below.*
>
> **Ian:** Actually I suppose just $a_n = n$ is going to be, it's not bounded below because if you just include all the- if n can just be an integer. Yeah, if n is just an integer, then it's not going to be, there's no lower bound for the sequence. That will tend to infinity. And also negative infinity as well. Yeah, I'm just going to go with that though.
> [Answer given: $a_n = n$, *for all integers*]

Incident Rep-A3: Considering changing the domain

Several students implicitly considered using a domain other than ℕ when answering questions, but decided that it was not permissible. The following two quotes illustrate this:

> *Q10. A sequence that tends to infinity that is not bounded below.*
>
> **Mike:** I suppose we could have something like, no - because we can never have a negative n can we? I was thinking we could have something like $a_n = n^3$, something like that and that wouldn't be bounded below, and it would tend to infinity, but then we'd have to have a negative n, which we're not allowed.

> *Q7. A bounded, monotonic sequence.*
>
> **Ken:** Well I was trying to think of something that would tend towards the limit, actually I think I've already answered this question, bounded monotonic sequence - yeah, it is for natural numbers, isn't it?
>
> **Interviewer:** What do you mean?
>
> **Ken:** You're doing natural numbers for the sequences aren't you, so not negative?

These quotes illustrate something not seen in the pilot studies, they are incidents where students who briefly consider and then reject **INS** responses.

6.3.2 Rep-B. One representation is superior

For some students that were interviewed, certain representations of sequences were seen as superior. Such students would typically work with some representation of their sequence initially: a graph, list of numbers, or a mental object of some type, but no final answer would be given until this representation had been converted to a certain type, sometimes at the expense of not giving a final answer at all. For the students in this study, the desired representation was, without exception, a formula.

The types of incidents categorised as Rep-B were:

- Sequences are not valid unless they are represented with a formula
- Knowing an answer but unable to write it in a 'correct' way
- Unsure of correct notation to use

Incident Rep-B1. The only valid representation is a formula

In the interview data collected, if a student remarked that they felt there was only one valid representation of a sequence it was that of a formula, "$a_n = f(n)$". This incident selects some of these comments which were made. The first quote illustrates this belief manifesting as an obstacle for Nicola, who knew what she wanted her sequence to do, but ran into trouble writing the sequence as a formula:

> *Q2. An increasing sequence that is not strictly increasing*
>
> **Nicola:** If you have every two successive terms are the same, then that would be increasing. Think how to write that. Well I could write it as a list of numbers. That's actually a viable sequence?

Comments of this type often occurred during the discussion period when I was discussing questions the students had left blank. Where a student gave the impression that they could think of a sequence but couldn't write it down I suggested to students not to worry about finding a formula. The following two quotes illustrate this:

> *Q2. An increasing sequence that is not strictly increasing*
>
> **Interviewer:** You don't have to find a formula. You can just put a list of numbers if you prefer.
>
> **Phalgun:** That's not really right is it? It's writing numbers.

> *Q9. A sequence that tends to infinity that is not increasing*
> [Answer given: *Impossible*]
>
> **Interviewer:** So all we need to do is go down a few times, or just once, then go up forever.
>
> **Anna:** But how would you write that?

Interviewer: In terms of your drawing you could do down, then up, like a tick.

Anna: Yeah, do you know any like 'that' [points at her answer a different question, given as a formula]

Incident Rep-B2. Clarification of the 'correct' representation

One student, Anna, asked for clarification at the beginning of the task as to how to represent sequences in the task:

Anna: Do I need to write down the formula or draw the graph?

Other students did not ask for clarification until they had begun the task, seeking confirmation as to what was an acceptable way of representing their answer only when they were in possession of some sequence (either on paper or mentally):

Joe: I don't have to write like, brackets $a_n = n$ or anything?

Nicola: I'm thinking about a_{2n}. That's a sub-sequence. Am I allowed a sub-sequence?

Ken: Well a strictly increasing sequence has every term greater than the last one, so
[Writes answer: (3)]

Ken: Is that notation ok?

Laura: Strictly increasing sequence, so every term has to be bigger than the one before, so you could just have $1, 2, 3, 4$. Do I need to put like up to n or whatever?

The last quote from Laura also illustrates a cross-over between concerns about suitable notation (this Incident) and wanting to represent the sequence as a formula (Incident Rep-B1). Laura is unsure whether to represent the sequence "$(a_n) = n$" as:

(a) $1, 2, 3, 4, \ldots$

or (b) $1, 2, 3, 4, \ldots, n$.

6.3.3 Rep-C. Any well-defined representation is suitable

In the interviews, eight students (out of 15) gave at least one answer in terms of a list of numbers and at least one answer as a formula. Of these eight students, two (Phalgun and Ian) only wrote down a list of numbers after prompting from me. The remaining students typically provided the representation which was 'easiest' to write down, as will be illustrated in Incidents Rep-C1 and Rep-C4. What makes this category distinct from Rep-B is that in the incidents categorised here students are content that the object they give is a valid representation sequence, even if it is not the most desirable representation for them. Students did not typically comment on alternative representations of their sequences, but one such case is illustrated in Incident Rep-C2.

The types of incident categorised as Rep-C were:

- Using a representation when unable to use another
- Treating two representations as equivalent
- Switching between representations
- Using the representation which is easiest to write

Incident Rep-C1. Using a representation when unable to use another

Of the eleven questions on the task, Laura gave seven examples, stating that the combination of properties of four questions were impossible. Of the seven examples she gave, four were lists of numbers and three were formulae. During the discussion stage I asked if there was a difference for her in the answers:

> **Interviewer:** Another interesting thing, for me at least, is that you've given a list of numbers for some of them and for other ones you've given a formula. Did you just fancy a variation, or are these just more complicated perhaps, what were you thinking?

> **Laura:** I don't actually know. Mainly for ones like that I wouldn't know how to write it, well no you'd— I could but I'd have to think to do that, it's more. Yes I guess some of them, most of the ones that I wrote as a list it's because I couldn't think immediately how to put them as like '$a_n =$'.

Another student, Nicola, remarked when answering Question 2 that writing a list of numbers was a way of 'getting around' writing a formula:

> *Q2. An increasing sequence that is not strictly increasing*
>
> **Nicola:** I get around this by not writing this as a formula. 'Cus I want something that I want to dip at some point, but still going to be going properly up like that.

Nicola later modified her answer to include repeated terms (rather than dip), but kept the representation of the example as a list of numbers rather than a formula (all her other answers were given as a formula).

Incident Rep-C2. Treating two representations as equivalent

In one case a student remarked that two representations were equivalent:

> **Guan:** A sequence that tends to infinity. So that can just be $a_n = n$. Which equals $1, 2, 3, 4, 5$.
> [Answer given: $a_n = n = 1, 2, 3, 4, 5,$]

Note that in Guan's answer the equivalence is also illustrated in a way which is mathematically imprecise; she has used the equals sign in an inconsistent way (both between her two uses and relative to formal mathematics). Within this dimension of variation this is bracketed, however, and the written answer is interpreted as indicating equivalence of sequence representations.

Incident Rep-C3. Switching between representations

In this incident a change in representation is prompted by the interviewer:

Mike: A sequence that tends to infinity that is not increasing, I didn't think we would be able to do that

Interviewer: And why's that?

Mike: Actually yes we could — well, yes we could actually because, We want something that goes like that. [draws sketch, shown below]

Mike: So it's not increasing as there is a decrease between those terms, but it is increasing to infinity. So I'll write it, we want something like $a_n = 10, 9, 20, 19,$

Interviewer: It's quite tricky to get a formula for that

Mike: Yes I think so, we probably want something a bit simpler.

Interviewer: Can you think of how to make it simpler?

Mike: Well we could pick out something in here, we've got the 10, 20 and the 30. So I'm not sure how I'd notate this but something like a_n equals, lets say, if n is odd, $10n$. n is even, $a_{n-1} - 1$. Err, which would give us this sequence think.

A statement such as this demonstrates that Mike is comfortable, when prompted, to represent the sequence in different ways. Note that here the choice to represent as a formula was directed by the interviewer.

Incident Rep-C4. Using the representation which is easiest to write

One student, Oksana, switched between different sequence representations in her series of answers. For the first six questions, and later in Question 9 her answers were given as a list, and for Questions 7 and 8 she she wrote a formula. This was brought up in the discussion period:

Q1. A strictly increasing sequence

[Answer given: $1, 2, 3, 4,$]

Q7. A bounded, monotonic sequence

[Answer given: $\frac{1}{n}$]

Interviewer: Some of them you gave a list of numbers for, and then when you got to Question 7/

Oksana: Yeah [laughs]. Because I found it, I couldn't think of a, yeah actually I don't know why—

Interviewer: That's fine, it's just interesting

Oksana: Because I like the idea of "$1, 2, 3, 4,$" but the writing of "$1, \frac{1}{2},$" sound more complicated so I thought, I thought why not?!

Oksana has moved flexibly between different representations of sequences, and based on her response to the interviewer's question, she did not consciously decide to do so. On reflection, she notes that her answer to Question 7 would have been more cumbersome to write as a list of numbers, and hypothesises that this may be the reason why. This idea perhaps can be continued to explain her returning to writing a list of number for Question 9:

Q9. A sequence that tends to infinity that is not increasing

[Answer given: $231564897\ldots$]

This sequence is easier to write as a list of numbers than in a sequence.

6.3.4 Commentary on representation of sequences

The students who volunteered for the main study gave far fewer answers which were not sequences compared to the pilot studies, in terms of number of **INS** responses. For instance, no student gave an interval of the real line, something which was most common in the pilot interview study. The rare instances of **INS** responses were categorised Def-A, along with certain comments and answers which suggested that certain students considered representations which gave the 'essence' of a sequence as valid sequences in their own right (see, for instance, Incident Rep-A1). It was far more common for a student to to briefly consider altering the domain when faced with a combination of properties they were struggling with. Incident Rep-A3 presented two passages where students considered extending the domain to $\mathbb{Z}$ in order to answer a question involving boundedness.

The lack of **INS** responses did not mean that students were comfortable with multiple representations of sequences, however. Incidents which were categorised Rep-B typically involved students who knew, to some extent, what the behaviour of their chosen sequence was or perhaps what it would look like in a sketch, but they were unsatisfied with their answers until their sequence was represented as a formula in the format "$a_n = f(n)$". Even amongst students who presented their sequences in multiple representations, lists of numbers were typically seen as less desirable than formulae for many students, as is demonstrated in Incident Rep-C1.

6.4 Sequence construction strategies (Con)

This dimension looks at the strategies students use when they generate sequences. As discussed in the literature review in subsection 2.2.6, in this context the word strategy is used only loosely. The students did not typically comment on an example-generation strategy explicitly, but did often comment on what they were attempting to do when answering a particular question. It is the interpretation of such comments, together with observations from what students did during the task, that form the basis of this dimension of variation, addressing the second research question in terms of the possible strategies a student might be aware of, and therefore act upon.

When the heading *construction strategies* emerged from the open coding of my data it was compared with the classification made by Antonini (2006). Although Antonini did not frame his research within a phenomenographic framework, his categorisation of example generation strategies into three types — trial and error, transformation, and analysis — fitted neatly with incidents coded under sequence construction strategies, and so these strategies can be found within my dimension of variation under categories of description Con-B, Con-C, and Con-D, respectively. The remaining category of description Con-A *Generic Initial Approaches*, outlines routines and rituals performed by students typically at the start of the question, such as underlining key words in the question.

Aspects from this dimension of variation have been published in Edwards and Alcock (2010a).

An outline of the dimension is given in Table 6.3.

TABLE 6.3: Dimension of Variation: Sequence construction strategies.

	Description
Con-A	*Generic Initial Approaches*
Con-B.	*Trial and Error*
Con-C.	*Transformation*
Con-D.	*Analysis*

In this section, the categories within the dimension of variation are first described, postponing arguments that the categories with higher letters are more sophisticated until the discussion section.

6.4.1 Con-A. Generic initial approaches to questions

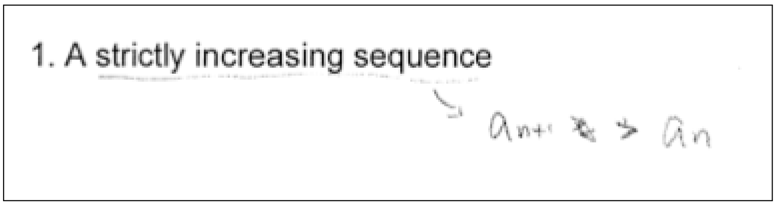
Some students had a routine that they followed when attempting a new question, regardless of the properties asked for in the question. These routines might be to refer to the appropriate definition, or underline parts of the question. These initial approaches to generating examples are considered as less sophisticated than the other strategies in the dimensions because they manifested as routines. Of course, such routines might be useful or even desirable when generating examples; by initially familiarising themselves with the appropriate definitions at the start of a question, students may be more successful in applying one of the other strategies (c.f. Incident Def-C2).

The types of incidents categorised as Con-A were:

- Underlining properties
- Writing down the definitions
- Repeating the question over and over

Incident Con-A1. Underlining properties and writing down definitions

Haaroon chose to underline in each question the properties which were requested, then to look at the definition sheet and copy down parts of the appropriate definition that were salient to him. The following scan of his task sheet illustrates this.



Phalgun also chose to begin by write down the definitions:

> **Phalgun:** Erm, basically I'm just writing the definition of increasing sequence so that I could think of an example.

Writing down relevant definitions is probably a very good starting point when generating examples, but such mechanised behaviour independent of the question asked is labelled Con-A.

Incident Con-A2. Repeating phrases or definitions

Some of the more sophisticated strategies that will be outlined later in this dimension of variation (specifically Con-B and Con-C) require an initial example to be generated. These initial examples may, or may not, be valid examples. For some students, finding any initial example at all to check or manipulate was problematic. Phalgun demonstrated two ways to gain inspiration, repeating the question aloud and writing over the question.

> **Phalgun:** I just keep writing over it 'cus I can't think. And [if] I can't do stuff I just keep writing over it and then hopefully.

Such routines appear to be very unsuccessful in providing inspiration. In a similar incident, Haaroon recalled he had seen a sequence which satisfied Question 9, and spent some time trying to recall it:

Q9. A sequence that tends to infinity that is not increasing.

Haaroon: I remember doing this one actually. Hum, I remember doing this one.

Interviewer: Was this a question on one of the assignments?

Haaroon: Yeah, I remember doing this one. So it's not increasing, so it's going up and then down, up and then down essentially. I think so yeah. . . . I remember doing this one—I should know this one, it's so poor. I'll have to leave it.

6.4.2 Con-B. Trial and error

The second category in the Sequence Construction Strategy dimension of variation is trial-and-error type strategies. Antonini (2006) described the trial and error strategy as follows:

> The example is sought among some recalled objects; for each example the subject only observes whether it has the requested properties or not. (Antonini, 2006, p.58)

Antonini found that this type of strategy was the most frequent used in his study of graduate students of mathematics, and in my data this is also the case. In a Con-B strategy, an initial example is generated and then either accepted as correct, or checked to see whether it satisfied the definitions specified by the question. In the most trivial case, the initial example is one which the student already knows is correct from past experience. The types of incidents categorised as Con-B were:

- Trying simple (prototypical) examples
- Trying examples seen in class
- Trying examples already used on the task

Incident Con-B1. Trying simple (prototypical) examples

Some students began questions by considering simple examples. A student who did this was David:

Q1. A strictly increasing sequence

David: Ok. So the first one is pretty simple. Just straightforward progression of the natural numbers.
[Answer given: $a_n = n$]

Q5. A sequence that has no upper bound

David: I'm going to use, again, the simpler, just the natural numbers again.
[Answer given: $a_n = n$, $n \in \mathbb{N}$]

For these two questions, David has noted that the simple sequence $(a_n) = n$ will work. Many students gave this sequence to the three 'warm-up' questions on the task sheet: Question 1 (93%), Question 5 (80%), and Question 8 (47%).

Incident Con-B2. Trying examples already used on the task

Some students took a sequence which they had already given to an earlier question and checked whether it also worked for a new question. No student remarked that they were using their prior answers in any kind of structured way, for instance looking at questions they had answered previously which contained similar properties to the question they were trying to answer. It is possible that this was implicitly the case for some students however, and some incidents could be interpreted as such. Take, for instance, the following incident from when Guan was attempting to give an example for Question 9. Earlier in the interview she had answered Question 6 as follows:

Q6. A sequence that has neither an upper nor a lower bound
Answer given: $a_n = (-1)^n n$

She was now attempting to answer Question 9:

Q9. A sequence that tends to infinity that is not increasing

Guan: I think it can just be this one [points to her answer to Question 6], because—oh, that doesn't tend to infinity[. . .] this is just a sub-sequence tending to infinity but the whole sequence doesn't.

Her answer to Question 6 was not increasing and contained a subsequence which tends to infinity, which is superficially similar to the question she was attempting. Later on in the task she used trial and error with an example she remembered from class to answer Question 9, and this is discussed towards the end of Incident Con-B3.

Incident Con-B3. Trying examples seen in class

It was common for students to refer to sequences they had seen in their lectures and problem classes. This was often because the material was fresh in their mind:

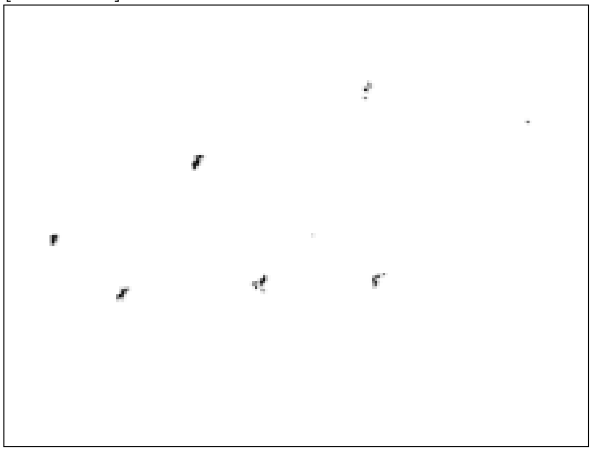
> **Haroon:** As you see I've been using $\log(n)$, because I've just had a tutorial and I've just been doing logs. So it's been drilled in to you.
>
> **Ben:** Well to be honest in the series that we were doing this morning this sequence came up.

It is interesting to note that the phrasing of this incident indicates that the reference to the assignment sheet is also being used to justify the correctness of the answer (see the dimension of variation Justifications in Section 6.5). This suggests that for some students, a trial and error strategy evoking examples the student has already met brings with it a justification as to why the example is correct (see Incident Jus-B1 in subsection 6.5.2).

In the previous incident (Con-B2) we saw Guan using her previous answers as candidates for latter questions. She too used examples she had seen in class:

> *Q9. A sequence that tends to infinity that is not increasing.*
>
> **Guan:** Yes, I think I've found an example for nine. It's, I remember the lecturer talked about a sequence which is [like] this:

[sketches]



Guan: So it tends to infinity but it's not increasing because this term [points an even term] is obviously not greater than the previous term. So, we can write it as $1, 0, 2, 0$.
[Answer given: $a_n = 1, 0, 2, 0, 3, 0, 4, 0, \ldots$]

Bracketing the fact that the example generated to satisfy Question 9 is incorrect, we can see that this quote, together with the ones presented in the last incident, shows that Guan had used various sources to obtain initial solutions for her Con-B trial and error strategy, including answers given to previous questions on the task and sequences she had seen previously in class.

6.4.3 Con-C. Transformation

Recall that Antonini (2006) described the transformation strategy as follows:

> An object that satisfies part of the requested properties is modified through one or more successive transformations until it is turned into a new object with all the requested characteristics. (Antonini, 2006, p.59)

A transformation strategy requires an initial example, or representation of an example, in order to make (possibly successive) transformations. For the sequence generation questions in the task, no student made more than one transformation, but there was variation in the source and representation of the initial example. The types of incidents categorised as Con-C were:

- Transforming an answer seen before
- Transforming via another mathematical object
- Transforming a graphical object

Incident Con-C1. Transforming an answer seen before

When answering Question 6, several students described a strategy where they began with an oscillating sequence of fixed magnitude such as $(a_n) = (-1)^n$, and increased the magnitude of successive terms. No student did the reverse: taking an increasing sequence and introducing an oscillatory factor, although this would also be a way in which a transformation strategy may give a correct answer to Question 6:

> *Q6. A sequence that has neither an upper bound nor a lower bound.*
>
> **Valter:** We can recall from the assignment 4 that the sequence $a_n = (-1)^n$ was neither increasing nor decreasing, it was oscillating between -1 and 1. But its absolute norm was 1 so we want to change this 1 to another number, say a number larger than 2. So this would still oscillate about the x-axis but would be, the odd terms would, the sub-sequence of the odd terms would diverge to minus infinity while the sub-sequence of the even terms would diverge to plus infinity. So it has no bounds.
> [Answer given: $(a_n) = (-2)^n$].

> *Q6. A sequence that has neither an upper bound nor a lower bound.*
> [Answer given: $a_n = n(-1)^n$]
>
> **Interviewer:** So how did you come up with that one?
> **Nicola:** I thought of $(-1)^n$, and thought that is bounded, so I needed to change it at each n, so just timesed it by that [n], and it'll be different every time.

Sometimes in the discussion part of the interview I prompted students to modify an attempt given previously:

Q6. A sequence that has neither an upper bound nor a lower bound.

[Answer given: $-1, 1, -1, 1$]

Interviewer: So you're kinda on the right lines with that number 6, so thinking about what you've done. Is there any way you can modify the sequence you've just wrote to make it get as big and as small as you like?

Edha: Oh, that's $a_n = (-1)^n$, if we multiply it by something I guess. So if we multiply it by $n + 1$. Then if it's 1, it's -2, and then 3, and then -4, yeah that's an example.

[Changes answer to: $a_n = (-1)^n(n + 1)$]

Sometimes, students began with an answer, then modified it syntactically:

Q6. A sequence that has neither an upper bound nor a lower bound

Phalgun: A sequence that has neither an upper bound nor a lower bound.

[writes $a = (-2)^n$]

Phalgun: $a_n = (-n)^n$. That's not right is it.

Here Phalgun begins with the prototype answer $a = (-2)^n$, and for some reason is not happy with it, so modifies the value inside the bracket. He is still not happy with his answer, but leaves it and continues to the next question. Later in the interview, it becomes clear that Phalgun is attempting to generate more esoteric answers by this method:

Q8. A sequence that tends to infinity

Phalgun: A sequence that tends to infinity

[pause of 20 seconds, during which Phalgun writes the example $(a_n) = 2^n$]

Interviewer: So you're after a better one again, yeah?

Phalgun: Yeah. I mean these are pretty obvious, I'm trying to think of a good example.

[Phalgun writes down $(a_n) = e^n$ and moves onto the next question]

The notion that although equally correct, some examples are better than others, ties in with the general (undefinable) notion of mathematical elegance. Some authors argue that it is a good pedagogical strategy to ask students to generate several examples, or requesting that students try to provide example which they believe few others will generate (e.g. Watson and Mason, 2005).

Incident Con-C2: Transforming via another mathematical object

This incident looks at where Valter begun with an **INS** object and then transformed that object into a sequence, before deciding that the sequence did not satisfy the question. Such events were categorised Con-C2, as an object is being transformed in an attempt to answer the question:

> *Q10. A sequence that tends to infinity that is not bounded below*
>
> **Valter:** An example of that would be '$a_n =$', no that actually would be. I was thinking about graphs as functions.
>
> **Interviewer:** Do you want to sketch what you were thinking of if you can't think of a formula?
>
> **Valter:** I was thinking of the graph of a function $\log(x)$, as a sequence we define it, it keeps on increasing tending to infinity, but it's bounded below.

Here, Valter has first considered the function

$$f(x) = \log(x), \quad f : \mathbb{R} \to \mathbb{R}$$

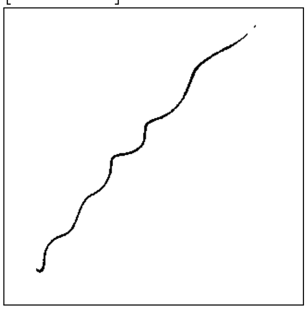
and noted that it tends infinity as $x \to \infty$ and is not bounded below. When restricting the function's domain to $\mathbb{N}$, he notes that the resulting sequence also tends to infinity, but is now bounded below.

Unlike the Incident Con-C1, the behaviour of the sequence is not transformed, it is the domain of the mathematical object that is adjusted. Although Valter's transformation did not produce a valid answer to the question, this type of strategy will sometimes be successful. For instance it could reasonably be argued that students who give sequences based on trigonometric functions are using this strategy implicitly.

Incident Con-C3: Starting with a graph

Some students began with a sketch of a function, and attempted to find a sequence which behaved in the same way. This type of incident is similar to Incident Con-C2, above, but in that incident the mathematical object was a real-to-real function, and the graph was a representation of that object. Here the difference is that the graph is the only given representation of the mathematical object.

During the discussion phase of Ken's interview, he was told that Question 9 was not impossible. He then had some inspiration, and sketched the shape of a function of which the corresponding sequence would have the required properties:

> *Q9. A sequence that tends to infinity that is not increasing.*
>
> [Answer given: Impossible]
>
> **Ken:** Ah, actually, yeah that one is possible. You could have something like a sine curve doing that.
>
> [sketches]
>
> 
>
> **Ken:** I need to think of the equation. Yeah, something like $n + \sin(n)$.

Here, Ken has taken his sketched graph, and tried to find a function which describes it. Superficially, the sketch looks as though it may be a sine graph rotated to be along the line $f(x) = x$. He then attempted to find a formula for the type of function he sketched, and suggested $a_n = n + sin(n)$. This transformation has not been unsuccessful, however, because the sequence represented by this formula is increasing (one way to see this is to observe that $f'(x + sin(x)) \geq 0 \ \forall x$, and so the sequence $(a_n) = f|_{\mathbb{N}}$ must be increasing).

In this incident Ken has taken a object represented jointly as a sketch and in relation to a known function, $\sin(x)$, and transformed it into an **IS** answer. This incident therefore demonstrates that a transformation strategy may not always provide a valid solution to a question, and such strategies may also remove desirable features from the object generated initially.

6.4.4 Con-D. Analysis

Recall that Antonini (2006) described the analysis strategy as follows:

> Assuming the constructed object [exists], and possibly assuming that it satisfies other properties added in order to simplify or restrict the search ground, further properties are deduced up to consequences that may evoke either a known object or a procedure to construct the requested one. (Antonini, 2006, p.59)

Recall that in his study with graduate students of mathematics, Antonini found fewer incidents of analysis than other strategies. Similarly, in their example generation study with undergraduates, Iannone et al. (2009) found only one case where a student's reasoning was consistent with an analysis strategy. In the main study data here there were also fewer instances of analysis compared with the other strategies, but still several incidents of note. Such incidents were typically where a student's construction strategy was analysis in that assumptions and deductions were made in order to evoke an example, but it was often the case that these individual assumptions and deductions within this strategy were mathematically incorrect.

The types of incidents categorised as Con-D were:

- Making deductions to clarify the example space
- Making assumptions to reduce the example space

Incident Con-D1. Making deductions to clarify the example space

When answering Question 11, Phalgun deduced that if a strictly increasing sequence does not tend to infinity, it must converge:

> *Q11. A strictly increasing sequence that does not tend to infinity.*

Phalgun: A strictly increasing sequence that does not tend to infinity. Increasing, converges to a limit.

Here, Phalgun's deduction has clarified a feature of the examples that satisfy Question 11 without changing the meaning of the question, and so it is an example of a Con-D *Analysis* strategy. Arguably this extra information could have resulted in him evoking a richer example space for this question. However, despite this deduction, Phalgun could not answer Question 11, and in the course of attempting it he did not make reference to this deduction again.

Sometimes deductions which serve to clarify the extent of the example space were incorrect, and actually expanded the example space to include sequences which do not satisfy the question. Such a deduction was made by Ian:

Q5. A sequence that has no upper bound.

Ian: I'm not sure you can say that but any strictly increasing sequence will do that. So $(a_n) = n$ diverges to infinity so it has no upper bound.

This chain of statements contains deductions which are mathematically incorrect. It is not the case that any strictly increasing sequence will have no upper bound, and so within this analysis strategy incorrect deductions have been drawn. The next statement, that a sequence which diverges to infinity will have no upper bound, is correct and the answer given, $(a_n) = n$, is also correct, however.

In the discussion period, Ian came close to the type of reasoning that if present during example construction would be labelled Con-D *Analysis.* Whilst justifying why his answer to Question 6 was correct he noted that:

Q6. A sequence that has neither an upper bound nor a lower bound.
[Answer given: $a_n = (-2)^n$]

Ian: It's going to increase, a subsequence is going to increase and decrease too. An increasing subsequence, decreasing subsequence, and for each term is going to exceed the upper bound and the lower bound.

This is an account of why Ian's answer was correct, rather than an account of a construction process or strategy, so it may not be clear why this passage was coded Con-D. However, this incident can be interpreted as the student indicating the features of answers which satisfy Question 6. In other words:

- If a sequence (a_n) has neither an upper bound nor a lower bound it must contain an increasing subsequence and a decreasing subsequence.
- For any candidate upper bound, there must be a term in the increasing subsequence that exceeds it (or we can modify the increasing subsequence accordingly), and a similar property for candidate lower bounds and the decreasing subsequence.
- $(a_n) = (-2)^n$ has suitable subsequences.

A chain of reasoning such as this would demonstrate a clarification of the type of sequences which satisfy Question 6.

Incident Con-D2. Assuming further conditions

When a further condition is assumed, the example space is restricted. This can be of benefit when the further condition serves to make the reduced space more accessible, and the examples in it more salient. Within the data from this example generation task, such further conditions were typically that an answer could be written in a particular way. For instance, Haaroon decided to assume that his answer could be written in terms of two subsequences:

> *Q10. A sequence that tends to infinity that is not bounded below.*
>
> **Haaroon:** But not bounded below, so it'll be one of those like a_{n+1} [sic] and a_{2n}

This assumption did not help Haaroon provide an example, because there are no sequences which satisfy the conditions, so imposing further constraints on the representation of answers is unhelpful in this case.

Sometimes the assumption that the sequence could be represented as two subsequences reflected the belief that the entire sequence would somehow carry-over the desirable features of each subsequence. For instance, Ian began down this line of reasoning, but realised it would not work:

> *Q10. A sequence that tends to infinity that is not bounded below.*
>
> **Ian:** I'm trying to think of the example $(-2)^n$ again, where it's got subsequences. One can tend to infinity and one can tend to negative infinity. It's not going to be bounded below, but the whole sequence isn't going to tend to positive infinity.

Therefore, for the students in this study a Con-D *Analysis* construction strategy was not always successful, especially when further conditions were assumed.

6.4.5 Commentary on sequence construction strategies

When the heading *sequence construction strategies* emerged from the open coding of my data it was compared with Antonini's (2006) classification. Almost all incidents coded in this category fitted neatly within Antonini's classification, and those which did not were typically generic initial ways of approaching questions; actions that were performed to several questions in succession (categorised as Con-A *Generic Initial Approaches*). The remaining categories took Antonini's classification and labelled in the same order he presented in the paper presented: Con-B *Trial and Error*, Con-C *Transformation*, and Con-D *Analysis*. This subsection now proposes that this ordering of the dimension of variation increases in terms of mathematical sophistication, then goes on to relate the dimension to previous research on Anonini's classification.

Con-A is clearly the least sophisticated strategy; it is the application of one or more of a set of routines that are independent to the content of the question. These routines are arguably not strategies to answer the question as much as a way a students can familiarise themselves with the question and its associated content. Such routines may be beneficial if used in conjunction with another strategy, but for some students, this type of behaviour was the way they predominately went about tackling the questions.

Incidents categorised Con-B are more sophisticated than those classified by Con-A because an example is generated. The validity of examples produced by trial and error is typically unknown until they have been checked, and so this strategy is less sophisticated than Con-C *Transformation* or Con-D *Analysis*.

The most sophisticated strategies are those which either take an example, consider what it is lacking and transform accordingly (Con-C), or deal implicitly with the set of all examples which satisfy the given properties, clarifying or restricting this set to make example generation easier (Con-D). It is very difficult to compare these two strategies in terms of sophistication, but I believe that deducing or assigning properties in the abstract is more sophisticated than taking with a concrete example and modifying it. One can easily imagine a very sophisticated transformation, and a very simplistic analytic deduction or assumption, and so I would not argue that all Con-D incidents will be clearly more sophisticated than all Con-C incidents; there is some degree of overlap.

Antonini (2006) and Iannone et al. (2009) both reported that trial and error was the most common strategy seen in their participants. This is notable because the participants from the two studies had different levels of mathematical experience (graduate students and undergraduate students of mathematics, respectively). The undergraduate mathematicians in this study also predominately used strategies labelled Con-B, but this prevalence in the data is possibly due to two confounding issues. First, students were encouraged to think aloud; it is much more likely an event is self reported as 'I thought of this example' rather than a more detailed account of a more sophisticated example generation strategy. Second, generating examples of sequences which satisfy combinations of properties is perhaps less conducive to Con-C and Con-D strategies compared with the mathematical objects Antonini's participants were asked to generate.

When students in my study did use strategies which were categorised Con-C and Con-D, they typically did not produce valid examples. This may be because the type of mathematics involved within these strategies is too sophisticated for the students; novice mathematicians are more likely to make incorrect assumptions and deductions in a Con-D strategy, and make errors transforming an example in a Con-C strategy. By the very nature of the trial and error strategies categorised as Con-A however, errors can only occur when checking to see if an initial example has the required properties.

6.5 Justifications (Jus)

This dimension looks at how students justified their confidence in their answers, and how they justified the deductions they made during the interviews. Incidents presented from the categories of description in this dimension originate from two parts of the interview therefore: comments students made when attempting the questions during the task period, and comments made when students were reflecting upon, and discussing their answers in the discussion period. In terms of the second research question, the dimension is richer than just incidents where a student did, or did not justify their answers. It also looks at the sophistication of these arguments, and considers students' level of confidence in their justifications.

The category of description Jus-A *No Justification Attempted* presents incidents where students did not justify their answer. Such students may have been unsure that their answer was correct so did attempt to justify it, and perhaps others were so confident in their answer that they felt no justification was appropriate. Incidents where students sought justification, clarification or confirmation from an external source are classified Jus-B, *Appeals to an External Authority.* An attempt at a justification, but one based on intuition, or otherwise without reference to any mathematical content is classified Jus-C, *Informal Justifications.* Finally, incidents where a student offered a justification which was based on suitable mathematical warrants were classified Jus-D, *Formal Justifications.* These mathematical Jus-D classifications bracket whether the mathematics is valid relative to formal mathematics, they are based more on the type of justification offered.

An outline of the dimension is given in Table 6.4.

TABLE 6.4: Dimension of Variation: Justifications.

	Description
Jus-A.	*No Justification Attempted*
Jus-B.	*Appeals to an External Authority*
Jus-C.	*Informal Justifications*
Jus-D.	*Formal Justifications*

6.5.1 Jus-A. No justification attempted

There were many incidents where students made no attempt to justify why their answer was correct, or why a chain of reasoning was correct. Incidents categorised as Jus-A range from when students made a statement with reason given as to why it was correct, to remarks about the obviousness of an answer or conclusion, to those students who remarked that they were unsure of their answer. In summary, the types of incidents categorised as Jus-A were:

- No reference to justification
- Reference to an answer being obviously correct
- Students unsure of the validity of their statement

Incident Jus-A1. No reference to justification

There were many times a conclusion was drawn, or an answer given, with no justification at all. Such statements may have been correct, or incorrect, relative to formal mathematics as is illustrated by the following two extracts, respectively:

> **Nicola:** If it oscillates it's not increasing or decreasing.

> **Mike:** If it's not bounded below, then that means it goes to infinity below.

Incidents of this type typically stated their conclusion as if they were facts.

Incident Jus-A2. Reference to an answer being obviously correct

In lieu of justifying why an answer was correct, some students appealed to the obviousness of their answer. Consider the following four extracts which claim, respectively, that an example is easy to generate, that a question is likely to be impossible via intuition and common sense, that an example is trivially correct, and that a question is impossible by definition:

Q4. A sequence that is neither increasing nor decreasing.

[Answer given: $1, 1, 1, 1, 1, 1, \ldots$]

Edha: That's easy.

Q9. A sequence that tends to infinity that is not increasing.

Valter: Intuitively I'd say 9 was one that was impossible. Just appealing to common sense.

Q4. A sequence that is both increasing and decreasing.

Phalgun: Both increasing and decreasing, a trivial one $a_n = 1$

[Answer given: $(a_n) = 1$]

Q9. A sequence that tends to infinity that is not increasing

Ken: I think that's impossible. By definition.

These four justifications are grouped here because they are similar, not because they necessarily should all belong to Jus-A. Note that the phrases "trivial" and "by definition" are often used legitimately within a formal mathematical framework, and so it is a subjective matter whether the use of this word in a given context should be classified Jus-A *No Justification Attempted* or Jus-D *Formal Justification.* In particular I believe by considering the structure of the justification, rather than the mathematical truth, Edha's comment should be classified Jus-A, Valter's Jus-C, and Phalgun and Ken's Jus-D.

Incident Jus-A3. Students unsure of the validity of statements

Some incidents were marked Jus-A when a student noted they were unsure if their answer, or a conclusion they were drawing was valid:

Valter: I'm not sure you can say [this], but any strictly increasing sequence [will have no upper bound]

6.5.2 Jus-B. Appeals to an external authority

In this second category of description, incidents are presented where students did not justify their answers or conclusions themselves, but instead referred to an external authority. The external authority may be what they have been told in lectures, what they remember from class problem sheets, or from work they have otherwise done previously. Such incidents were not categorised as Jus-A because here a student is providing a justification, all be it one based on some other source. Several of the events categorised Jus-B can also be found in incident Con-B3 *Trying Examples Seen In Class.*

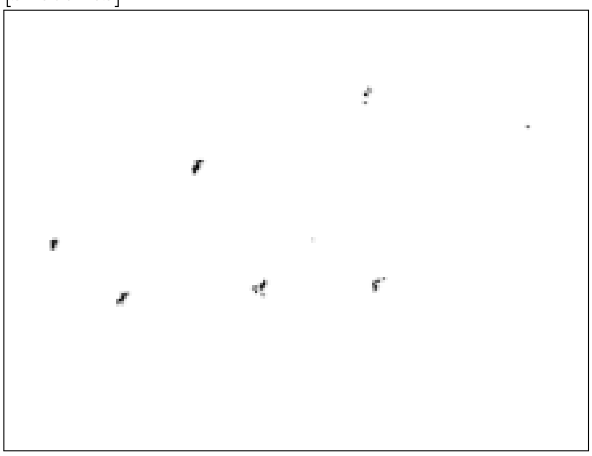
The types of incidents categorised as Jus-B were:

- Referring to lecturers comments
- Recalling past assignment sheets
- Asking the interviewer for input

Incident Jus-B1. Referring to lecturers' comments

The following quote from Guan was previously given as part of Incident Con-B2. Here Guan believed she had remembered her lecturer giving an example of a sequence which would satisfy Question 9:

> *Q9. A sequence that tends to infinity that is not increasing.*
>
> **Guan:** Yes, I think I've found an example for nine. It's, I remember the lecturer talked about a sequence which is [like] this:
>
> [sketches]



> **Guan:** So it tends to infinity but it's not increasing because this term [points an even term] is obviously not greater than the previous term. So, we can write it as $1, 0, 2, 0$.
>
> [Answer given: $a_n = 1, 0, 2, 0, 3, 0, 4, 0, \ldots$]

In this quote after providing the justification that the sequence was mentioned by her lecturer, she went on to provide justification why the sequence is not increasing ("because this term is obviously not greater than the previous term"), but did not continue to give a reason why the sequence tends to infinity. Indeed the sequence does not tend to infinity as every second term returns to zero.

Joe similarly made reference to his lecturers' comments, but decided to overrule them:

> *Q4. A sequence that is neither increasing nor decreasing.*
>
> [Answer given: $a_n = 3$]
>
> **Interviewer:** Why is that [neither increasing nor decreasing]?
>
> **Joe:** Because it's just like a flat line. But then in those cases we're told that it's increasing and decreasing, but then it's doing neither as well. So I'll just put that for that.

Joe's recollection that his lecturer stated that a constant sequence is increasing and decreasing is at odds with his spontaneous conception that "it's doing neither as well", but he is unwavering in providing the constant sequence $a_n = 3$ for Question 4 (A sequence that is neither increasing nor decreasing), and the sequence $a_n = (-1)^n$ for Question 3 (A sequence that is increasing and decreasing).

Incident Jus-B2. Recalling past answers

As an illustration of a student who justified an statement by appealing to previous work they had completed, a quote from Valter, previously given in Incident Con-B3, is repeated. Note that here "assignment" refers to a question already answered on the task sheet, rather than a class assignment.

> **Valter:** We can recall from the assignment 4 that the sequence $(a_n) = (-1)^n$ was neither increasing nor decreasing.

Incident Jus-B3. Asking the interviewer for input

On rare occasions, students asked the interviewer if their answers were correct:

Q7. A bounded, monotonic sequence.

Ben: Bounded monotonic sequence. if it's bounded then it's bounded above and below. It's monotonic so it's increasing or decreasing. So the bounds. Yeah, if you just have a normal monotonic sequence that is the same. Monotonic, yeah you can have them both. Then that's bounded above and below at 1, and it's monotonic.

[Answer given: $a_n = 1, 1, 1, 1, 1$]

Ben: Is that ok?

Q2. An increasing sequence that is not strictly increasing.

Joe: I don't know, is this like one of those ones where you've got a_n, where the odd numbers equal zero and the even numbers equal n.

[Answer given: $[a_{2n}] = 0$ $[a_{2n-1}] = n$ (sic)]

Ben: Does that count for this one?

When this did happen, I explained that I didn't want to say until they had finished the task, but was happy to go over answers afterwards.

6.5.3 Jus-C. Informal justifications

To be classified as Jus-C, a student will have provided a justification as to why their answer, or chain of reasoning is correct, but that justification will not be based in (formal) mathematics. What exactly should constitute a justification calling on formal mathematics or otherwise will depend on context, but for the sequence generation task given to students the types of incidents categorised as Jus-C were:

- Justifying with non-mathematical terminology
- Referring to intuition or common sense
- Unable to think of any alternatives

Incident Jus-C1. Justifying with non-mathematical terminology

In this type of incident students justify their answer or chain of reasoning, by using terms that have not been defined. Often it is very clear what the student means such as Nicola's use of "plateau":

> *Q2. An increasing sequence that is not strictly increasing*
>
> **Nicola:** I was thinking for this one you need one that's going to, at some point, plateau, stay at the same point. Cus it's not strictly increasing.
>
> [Answer given: $a_n = 1, 1, 2, 2, 3, 3, \ldots$]

Although Nicola's notion of a sequence plateau can easily be formally defined (perhaps $\exists N \in \mathbb{N}$ s.t. $a_n = a_{n+1}$), one can never be sure that a term phrased in everyday language will be similarly understood by everyone. The phrase "stay at the same point", again phrased in everyday language, has less ambiguity and helps clarify Nicola's meaning.

Similarly, the treating an increasing sequence to be the same as a sequence that "goes up" has been a commonly referenced spontaneous conception in this chapter. Using such language as a justification is therefore labelled Jus-C:

> *Q3. A sequence that is increasing and decreasing*
>
> **Edha:** Sine or cos[ine] curve?
>
> **Interviewer:** What made you think of the sine or cos[ine] curve?
>
> **Edha:** It just keeps going up and down, but I don't think that's right.

Incident Jus-C2: Referring to intuition or common sense

An extract already presented as part in Incident Jus-A2 for comparative purposes was labelled Jus-C:

> *Q9. A sequence that tends to infinity that is not increasing.*
>
> **Valter:** Intuitively I'd say 9 was one that was impossible. Just appealing to common sense.

Here Valter's justification that no sequence can both tend to infinity and not increase is based on his common sense. Such statements have been labelled as Jus-C, as opposed to statements such that the answer is "trivial" or true "by definition", which have a more valid, if ambiguous, footing in formal mathematical reasoning.

Incident Jus-C3: Convincing oneself of the answer

Those students who justified their answer by stating they had convinced themselves were also categorised as Jus-C:

> *Q9. A sequence that tends to infinity that is not increasing.*
>
> **Nicoa:** So I'm looking for a strictly increasing sequence that's bounded above, because that would mean it wouldn't go to infinity.
>
> **Nicola:** Yeah, I've convinced myself that, maybe, logs aren't bounded, because you can have have any positive number greater than 1 and you'll keep getting an answer.

Nicola's statement, that the logarithm function is unbounded is correct (and so she correctly reasons that it is not bounded above), but her justification: that it is defined for any number greater than 1 (or alternatively that it is increasing with the domain) is invalid relative to formal mathematics.

Incident Jus-C4: Unable to think of any alternatives

In a similar vein to the last incident, the next extract taken from Ian's interview reached a situation where he reached a conclusion because he could not think of any other examples:

> *Q9. An sequence that tends to infinity that is not increasing.*
> [Answer given: *Impossible*]
>
> **Interviewer:** So could you explain why, in wavy-handy terms, why you think [this question is] not possible?

> **Ian:** Erm, because if it's going—I don't know—I suppose, by the definition of infinity, because there's always going to have to be terms that exceed a limit so you just always seem to be increasing, I don't know, all the terms that I think of are going to infinity in other ways, always increasing. I can't think of anything that's just constant or decreasing that's going to infinity.

It is probably the case that the interviewer's request for an informal justification (intended to put Ian at ease, indicating I didn't want a formal proof) contributed to the nature of the response, but the last comment: that the student couldn't "think of anything that is constant of decreasing that's going to infinity" was labelled Jus-C. Note that Ian's error stems from negating increasing to mean constant or decreasing, but in this section we bracket this spontaneous conception related to definition-type.

6.5.4 Jus-D. Formal justifications

The final category of description includes those incidents where a student's answer or reasoning was backed by a mathematically-based justification. To be classified as Jus-D the mathematical content need not be correct. It is the style of argument offered by the student on which this dimension of variation focuses. The types of incidents categorised as Jus-D were:

- Logical deductions with a mathematical warrant
- Use of definitions to support claims
- Use of accepted mathematical phrases

Incident Jus-D1. Logical deductions with a warrant

Students quoted in Incident Jus-A1 typically stated their conclusion without any justification, as if their conclusions were facts that required no further warrant. Where students did provide such a warrant, the passage was labelled Jus-D:

> *Q5. A sequence that has no upper bound*
>
> **Valter:** $a_n = n$ diverges to infinity, so it has no upper bound.

Ian: That's the same [sequence] as Question 1 I think [$a_n = n$]. For every upper bound you can always have terms above that upper bound.

In both of these quotes, the student has reused their answer to Question 1 when answering Question 5, and provides a reason why this is appropriate. The following extract from Guan's interview also has this structure (so is labelled Jus-D), but the warrant is no longer valid:

Q5. A sequence that has no upper bound

Guan: So it can be [a] strictly increasing sequence, so every term is greater than the previous term which [has] no upper bound. So it can be the same as the first one.

Incident Jus-D2. Use of definitions to support claims

Many of the incidents categorised Def-D *Manipulates Definitions* contain justifications based on the definition of sequence properties. The extracts given in Incident Def-D1 are restated here to illustrate this:

Q2. An increasing sequence that is not strictly increasing

Ben: Again, these [sequences] are ones that can be equal to the ones before because that's the difference in the definition, so you could have say 1,1,2,2,3,3. Any combination of those really.
[Answer given: $a_n = 1, 1, 2, 3, 3,$]

Guan: I'm thinking about the definition of increasing sequence and strictly increasing. So the difference is that strictly increasing the next term must be greater than the previous one, while increasing can be greater than or equal to the previous one. So it can be a constant sequence.
[Answer given: $a_n = 2, 2, 2, 2, 2, \ldots,$]

The students in both of these extracts justify why their answer to Question 2 is correct by drawing attention to the difference in the definitions of an increasing sequence and a strictly increasing sequence (i.e. that the constraint $a_n < a_{n+1}$ is relaxed to $a_n \leq a_{n+1}$).

Incident Jus-D3. Accepted mathematical phrases

In Incident Jus-A2, two interview extracts were presented that were similar to those which make reference to an answer being 'obviously' correct, but used the mathematical terms 'trivial' and 'by definition':

> *Q4. A sequence that is both increasing and decreasing.*
>
> **Phalgun:** Both increasing and decreasing, a trivial one $a_n = 1$
> [Answer given: $(a_n) = 1$]
>
> *Q9. A sequence that tends to infinity that is not increasing*
>
> **Ken:** I think that's impossible. By definition.

These forms of justification are common in mathematics, and although usually reserved for statements and claims that are mathematically simple to show (relative to the audience), the students' use of these terms are attempts at using mathematical warrants for claims.

6.5.5 Commentary on justifications

This section has outlined the dimension of variation Justifications, which has focused on the types of arguments students provided as to why their answers and chains of reasoning were correct. Throughout the dimension, the validity of a statement relative to formal mathematics has been bracketed, and ways in which these statements were justified examined. In Jus-A incidents, students provided no justification, and in Jus-B justification was sought from an external authority. When students provided justifications themselves, these were categorised as either informal (Jus-C) or formal (Jus-D). The distinction the section presented between informal and formal was subjective, but based on the structure of the argument presented: if the structure of an arguments was compatible with formal mathematics then it was marked Jus-D, otherwise Jus-C.

Ambiguity in this dimension of variation is likely to arise from different interpretations of what exactly constitutes a formal mathematical justification framework. The categorisation presented here makes the distinction that deductions involving warrants drawn from mathematics are classified as Jus-D, bracketing the truth of those deductions. This is different to a distinction between arguments that use logic and those that do not. The distinction is whether a claim is backed up in a mathematical way.

Considering briefly the times when students chose to spontaneously give reasons for their answers and thinking, this dimension of variation corresponds well with Asghari's (2005) account of students' checking procedures for the 'mad dictator' task (see Section 2.2.6). Recall that Asghari noted that students would either use their concept definition when generating the example, thereby establishing the example's validity relative to that concept definition (conceptual generating), or else check to see if the answer was correct after generating it (figural generating). In this task the timing of (spontaneous) justifications was similar: students would either generate the example using the think-aloud protocol to provide reasons why their comments were correct (relative to their concept definitions), or else they would give an example and then justify why it was correct. Note that some of the justifications found in this dimension may not have been classified by Asghari as checking procedures, such as appeals to an external authority.

6.6 Summary of chapter

This chapter has addressed the second research question by presented the dimensions of variation that emerged from the phenomenographic data analysis of the main study of this thesis. The chapter began with a brief outline of the variation perceived between the definitions and the questions in the task, from the perspective of this researcher. A distinction was drawn between definitions that compared sequences term-by-term, definitions that gave a constraint that had to be met by all terms in a sequence, and definitions that governed the long-term behaviour of a sequence. Respectively these three definition types were labelled *term-by-term*, *sequence-wide*, and *long-term*. It was noted that the questions on the task then asked students to combine different sequence types.

Thereafter, four dimensions of variation were presented in Sections 6.2–6.5. The categories of description for each dimension are given in Table 6.5. Within each section, each category of description was illustrated by a collection incident types, where extracts of interview transcripts were presented. In keeping with the phenomenographic account of the data, in this chapter it was common to bracket two features: the validity of the extract's content relative to formal mathematics, and the 'location' of extracts from the perspective of the other dimensions of variation.

TABLE 6.5: Summary of the different dimension of variation outlined in this chapter.

Using Definitions (Section 6.2)	
Def-A.	*Unaware of Definitions*
Def-B.	*Refers to Definitions*
Def-C.	*Uses Definitions*
Def-D.	*Manipulates Definitions*

Representation of Sequences (Section 6.3)	
Rep-A.	*Any Representation is Suitable*
Rep-B.	*One Representation is Superior*
Rep-C.	*Any Well-Defined Representation is Suitable*

Sequence Construction Strategies (Section 6.4)	
Con-A	*Generic Initial Approaches*
Con-B.	*Trial and Error*
Con-C.	*Transformation*
Con-D.	*Analysis*

Justifications (Section 6.5)	
Jus-A.	*No Justification Attempted*
Jus-B.	*Appeals to an External Authority*
Jus-C.	*Informal Justifications*
Jus-D.	*Formal Justifications*

The first dimension of variation, Using Definitions, reported students' awareness of definitions in the context of such a task. In the least sophisticated category of description, Def-A *Unaware of Definitions*, students relied on spontaneous conceptions based on the everyday use of term. Def-B *Refers to Definitions* provided instances of students who made reference to (formal) definitions without using them as such, Def-C *Uses Definitions* included times when students used the content of the definitions in the example generation process, and in the most sophisticated dimension, Def-D *Manipulates Definitions*, students manipulated the definitions, perhaps negating or combining two definitions.

The second dimension of variation, Representation of Sequences, contrasted students' awareness of different ways in which to write down their sequence (typically in the form of a list of numbers, or a sequence). The least sophisticated category of description, Rep-A *Any Representation is Suitable*, provided instances where students presented answers without considering the representation chose of the sequence. When students had a strong preference for a particular type of representation (almost always as formula), this was categorised Rep-B *One Representation is Superior*. Here students were aware of different possibilities for sequence representation, but inflexible in the representations they provided. Finally, in category Rep-C *Any Well-Defined Representation is Suitable*, students were aware of different possible representations for their sequences, and moved between them with ease choosing the most suitable for a particular purpose.

The third dimension of variation, Sequence Construction Strategies, focused on the ways in which students approached the example generation task. In the least sophisticated category of description, Con-A *Generic Initial Approaches*, students performed routines and rituals prior to attempting to answer the question, repeating phrases aloud, writing down definitions and sketching graphs. The remaining three categories of description followed Antonini's (2006) example generation classification: Con-B *Trial and Error*, where students tried examples one-by-one until they found one which satisfied the required properties, Con-C *Transformation*, where an initial object not satisfying the requested properties was modified in stages until it did, and Con-D *Analysis*, where deductions were made about the properties an correct example might or should have, which eventually evoked a known sequence or a procedure to construct one.

The final dimension of variation, Justifications, looked at the different ways students justified their final answer or chains of reasoning. In the least sophisticated category of description, Jus-A *No Justification Attempted*, students presented their answers or gave logical statements with no warrant for their validity. Then, in the category of description Jus-B *Appeals to an External Authority*, students asked the interviewer for confirmation, or stated that they had relied on what they had been told in class. In the final two categories of description, Jus-C *Informal Justifications* and Jus-D *Formal justifications*, students provided reasons why their answer or chain of reasoning was valid. The distinction between an informal and a formal justification was drawn at whether the framing of, and technical terms in, a statement could be found in mathematics.

Chapter 7

Inter-Coder Validation Exercise

This chapter outlines an inter-coder validity exercise, which aims to consider how well the second research question has been addressed by the dimensions of variation reported in the previous chapter. The exercise aimed to explore both the communicative and pragmatic validity of the dimensions of variation that emerged from the data analysis. In the validation exercise, two colleagues independently coded extracts of interviews from a set of data that was not used in the main study. They were then asked to comment on how consistent and applicable the categories within each dimension were, and to discuss how complete a (phenomenographical) picture of the data was achieved.

7.1 Aims of the validity exercise

The data analysis procedure described in Chapter 5 aimed to produce a phenomenographical description of the different ways in which students experience the sequence example generation task. The outcome of the analysis was the emergence of a set of dimensions of (possible) variation. As described in the methodology of the main study (Chapter 3), the validity of such an analysis can be considered from two angles. The first angle is in terms of communicative validity, which addresses the question of how persuasive the interpretation of the data has been, whether the dimensions of variation are internally consistent and whether they are applicable to new data. The second angle is in terms of pragmatic validity, which explores “the extent to which the research outcomes are seen as useful” (Åkerlind, 2005c, p.331).

The validation exercise described in this chapter gave two mathematics education researchers interview extracts and asked them to consider the communicative and pragmatic validity of the dimensions of variation outlined in the last chapter.

7.2 Origin of the data used in the validation exercise

The data used in the validation exercise originated from interviews with two first year mathematics undergraduates at Leicester University. In common with the students from the main study, the topic of sequences and their properties was taught to the Leicester students in their first year (in this case in the module MA1151 Introductory Real Analysis). All students registered on this module were contacted at the beginning of the course inviting them for interview. Three students volunteered, and each took part in interviews in the same format as in the main study. During the interviews the students were given the same task sheet and definition sheet as the students in the main study (the same notation was used in each module).

For the validation exercise, two of the three interviews were selected and transcribed. The student in the interview which was not selected for the exercise had given few examples on the task, instead giving a short sentence for each answer. For instance, for Question 4 she had answered:

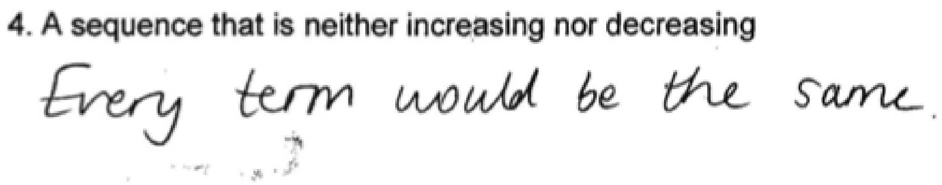


FIGURE 7.1: An answer to Question 4, given by one student at Leicester whose answers were not selected for the validation exercise

Although this type of answer is interesting, and perfectly classifiable within my coding system[1], these answers were so different to those given by students in the main study I felt it was more appropriate to give my colleagues tasksheets and transcripts from students who had attempted to give sequences to the questions.

[1] For instance it could be coded as Con-D. (Analysis)

The other two interviews were transcribed starting from when the student began the task, and ending when the discussion phase had evolved into more of a tutorial. The transcripts of these two students (from here labelled Student A, and Student B) can be found in Appendix B. It may be helpful for a reader of this thesis to code a portion of them in conjunction with the dimensions of variation given in Chapter 6 before continuing with this chapter.

7.3 Format of the validation exercise

7.3.1 The coders

Two colleagues agreed to code the extracts, and shall be known hereafter as Coder 1 and Coder 2. Coder 1 has had much experience with qualitative analysis and coding using a grounded theory approach, but little practice using Atlas.ti. Coder 1 was also familiar with my research in general. Coder 2 has had less experience with qualitative data analysis, but during the last twelve months has frequently used Atlas.ti to code data. Coder 2 was less familiar with my research and had not seen the task sheet or outlines of dimensions of variation before.

7.3.2 Outline of the exercise

At the start of the validation exercise Coder 2 went through the task sheet to familiarise herself with its layout and content. I then outlined my interpretation of a phenomenographic approach to coding, in particular the second order perspective, and bracketing of content outside a dimension of variation's focus. I then presented each coder with a summary sheet for each dimension of variation, and talked through each sheet's content. Each summary sheet contained the table outlining the layout of the dimension (for instance see Table 6.1 in the main study chapter), and the short introductions which described each category of description within a particular dimension of variation (see the text between the category title and the first incident in each category). The incidents and quotes presented in the main study chapter were not included. Any questions about the dimensions or the validation task were answered and I briefly explained the aims of the validation exercise. This introductory phase lasted around 45 minutes.

Each coder was then given a hard copy of the transcripts and task sheets for Students A and B, and an Atlas.ti file which was set up with the transcripts and categories for each dimension of variation. Each coder was asked to code the transcripts for salient features in conjunction with my dimensions of variation, but also to note anything that was related to a dimension but did not fit, and also anything salient that seemed not to fit with any dimension. This stage of the exercise lasted around 90 minutes. Coder 1 coded both transcripts in this time. Coder 2 coded the transcript of Student A and around a fifth of the transcript of Student B. Neither coder asked any additional questions during the coding.

After the coding, we discussed as a group the validation exercise for around 45 minutes. We first compared the dimensions that each coder had chosen, and then the appropriateness of the dimensions of variation structure for each extract, and any other general thoughts. Both coders gave their permission for this portion of the validation exercise to be audio-recorded. I listened to the recording several times, writing notes about the themes which emerged.

7.4 Outline of outcomes emerging from the validation exercise

This section describes the salient points made during my discussion with Coder 1 and Coder 2 in the final part of the validation exercise. In general both coders felt able to apply codes to the data, and differences of opinion were typically focused on the distinction between codes and the intended purpose of the various dimensions. Although the discussion was centred around the choice of categories to code a passage, emerging themes could be split into three areas. They were: (1) General coding issues and technique, (2) The phenomenographical approach to data analysis, and (3) The meaning of specific categories within a dimension of variation. I will outline the concerns, suggestions and comments of the coders in each of these areas, interjecting *in italics* my reflections. I will then finally relate these comments to communicative and pragmatic validity.

7.4.1 General coding issues and technique

Some of the issues discussed related to general coding choices and themes which may be true of any interpretive coding scheme.

Length of passages to code

The two coders had made different choices when deciding appropriate lengths of transcript to attach codes. Coder 1 followed a style that was similar to how the main study data has been initially coded. She selected salient passages related to any category within a dimension of variation, and these passages were labelled with a code. In a different style to this, Coder 2 had split the text up question by question assigning a code from each dimension to each question.

This resulted in Coder 2's transcripts being more heavily coded than Coder 1's transcripts. It also resulted in Coder 2 feeling that for sections of text associated with some questions there was not enough data to code for each dimension.

> **Coder 2**: I thought I had to put one of each [dimension] for each question. For each question I put four [codes]. I decided to see which one it was I thought fitted.
> **Coder 1**: I didn't do that.
> **Coder 2**: But there was an answer where she didn't say much, I think the first four are almost identical [in respect to the codes applied to Student A]

When a researcher is coding a transcript for dimensions of variation, it is not clear which length of text is appropriate for a single incident. Coder 1's style was in keeping with the way a researcher would generate codes when initially coding transcript, but perhaps this is not so appropriate when coding within a preexisting framework?

I believe an appropriate approach is to consider first what a coding framework is trying to achieve. The dimensions of variation presented in Chapter 6 aim to describe students' awareness of example generation of sequences, not to provide a framework that can classify each answer with respect to each dimension. So although a question-by-question approach may be an appropriate level of detail in which to code in some cases, the framework is not intended to code every answer with each dimension of variation.

Naming of categories

The practice of labelling each category within a dimension with a short title sometimes resulted in a discrepancy between the category's intended use, and alternative interpretations of the category's meaning.

> **Coder 2**: I was sticking with it because I decided, you know once you decide that 'unaware' [of definitions] only means a certain thing you will stick with it until it's completely different, and so it's one of these coding experiences.

Coder 2 also found that once she had decided on the meaning of a particular category, the meaning of other categories would then be extrapolated based on comparisons between different parts of the transcripts, rather than the descriptions of the categories.

Both of these issues are inevitable for any classification scheme that is interpretative and hierarchical. I felt it necessary to have a short name for each category because it is quite dry to refer to categories by their classification only (e.g. Rep-C), but I acknowledge that there is a danger that short names do not completely reflect subtle distinction between similar sounding categories.

Coding of answer sheets as well as the transcript

A general issue that was raised by both Coder 1 and Coder 2 was that at times they felt that it was less appropriate to code part of a transcript, but instead more insightful to code a students' answer. This can be seen when Student A gave the sequence "$1, 4, 9, 16, 25, \ldots$" as an example of an increasing sequence. Both Coder 1 and Coder 2 felt that this was a prototypical example (square numbers), and therefore evidence of a Con-B (Trial and error) construction strategy. However, Student A had not articulated the terms of the sequence or verbally described her thoughts and so there was nothing to code on the transcript. Because Coder 2 had decided to try and assign a code from each dimension for every question on the task she had coded the corresponding question Con-B. Coder 1 agreed with Coder 2's interpretation of the answer, but decided the transcript did not have the required emphasis and so elected not to code for this interpretation.

This discussion point highlighted a more general question of what evidence, if any, a coder should consider as complimentary to the audio transcript. When coding the audio from an example generation task such as the one given to students, it seems appropriate to consider students' answers as codable entities also. During the analysis of the main task, almost all answers were commented upon at some point during the interview (either in the task or discussion phase), and so this issue did not emerge. In each incident analysed in Chapter 6, I included the student's answer, and where appropriate I included a scan of the answer.

Differences in interpretation of codes

Finally, there is the issue that with any interpretive coding scheme, there will be differences. A particular rich example of this is for the line when Student A remarked "No, I don't know that one". Depending on the context of the remark, and the interpretation relative to the dimensions of variation it could be coded:

- Con-A (generic initial approaches). Perhaps Student A tried to think of the example for each question and gives up if she can't think of one right away?
- Con-B (trial and error). Perhaps internally she is trying out prototypical sequences?
- Jus-B (appeals to an external authority). Perhaps she is justifying not giving an answer to the interviewer?
- Def-B (refers to definitions). Perhaps she is explaining that she doesn't know what the definition means, and so can't think of an example?

A qualitative coding system relies on coders interpreting both the codes, and the material to be coded. An individual incident may be difficult to define exactly within a particular dimension, and what is often more helpful is to compare a variety of incidents, determining which are more and less representative of particular categories within a dimension. The nature of the validation exercise described in this chapter meant that such a comparative process was not possible and so it is unsurprising that there were discrepancies in some cases between myself, Coder 1 and Coder 2. I outline some of these discrepancies in later sections of this chapter.

7.4.2 The phenomenographical approach to data analysis

During the discussion, three main comments were raised about the study's phenomenographical methodology, and the subsequent implications for data analysis. The first, and most general, was on how phenomenography appears to blur the distinction between what a student articulates and what can be inferred by a researcher based on what the students says and does. Coder 2 wondered if a limited series of interviews can yield enough information to truly explore how students think, especially if such a study is reliant on students articulating what they think.

As described in Chapter 3, a phenomenographic account of data describes the variation in the different ways students can experience a topic, and exploring the relationships between these dimensions of variation. Attributing a way of thinking to an individual, or labelling a student's way of thinking is not a primary issue. I have found that over the course of fifteen interviews evidence of a variety of dimensions, and a rich variation within each dimension.

The second issue raised was connected with bracketing. Both coders agreed they did not have difficulties bracketing a dimension from other dimensions.

> **Coder 1**: [There was] no real issue with regards to which dimension we were coding at a particular time.

However, both coders found it difficult to bracket mathematical correctness during the coding of the data. For instance, both transcripts contain incidents that could arguably coded Con-D (Analysis), but the statements themselves were incorrect relative to formal mathematics, or the students were dealing with concepts that were in some sense too easy to count as sophisticated. For this to be coded as the 'most sophisticated' in the dimension seemed to reward student for not understanding the material.

> **Coder 1**: I was in two minds of [deciding if the extract should be coded as] manipulating and using [definitions], because he's doing something with the definitions—combining them correctly—but not hard ones, I didn't really feel he deserved a [Def] D.

To bracket the content of the mathematics is difficult but it is certainly the case that a student can comment on a particular definition, representation, construction strategy or justification in a way that is sophisticated relative to the dimension, but with mathematical content that is unsound relative to formal theory. The problem with bracketing the mathematical correctness is an issue which reoccurs throughout this description of the validation exercise.

The third issue raised which was related to the phenomenographical approach was that some of the less sophisticated categories within some dimensions were negatively phrased. For instance, it is not clear how best to code the validation transcripts for Def-A (unaware of definitions). If a student declares "I don't know the definition here" they are referring to definitions, and so the extract should be coded Def-B. Exactly when is it appropriate to code for Def-A?

In some dimensions of variation there are negatively phrased categories which are included in part to complete the dimension, for instance Def-A's inclusion in the dimension using definitions. *This is not the same as these categories not being present in the data; sometimes a student's comments demonstrate the a spontaneous conception based on the everyday use of a word, thus they imply that the student is unaware of definitions at this point. Within such a transcript therefore inferences must be made as to when it is appropriate to label using a negative phrased category.*

7.4.3 Discussion related to the dimension *using definitions*

We now move on to discuss Coder 1's and Coder 2's comments which were related to individual dimensions of variation. Recall that in the last section, Def-A (unaware of definitions) was described as negatively phrased. When discussing with the coders what they felt this category meant, it was clear that initially Coder 2 saw the category as the same as 'getting the definition wrong' or 'uses definitions superficially'.

> **Student A, line 25**: "And here there's an L. So maybe that's like a_n's in between U and L. But I don't know if the U counts."
>
> **Coder 1**: The next one on line 25 [Student A], I also put [Def-B] refers to definitions
>
> **Coder 2**: I put [Def-A] unaware of definitions

> **Coder 1**: She seems to be talking about the U, the L and the a_ns, that's clearly from the definitions isn't it?
>
> **Coder 2**: Erm, because it didn't go more into it

This understanding of the role of Def-A is probably a result of the labelling. Clearly a student who misapplied a definition is unaware of the correct application. The intended use of the category was not for this purpose, for if a student misapplies a definition they are, to some extent, aware of how definitions could be used, and more generally aware of the role they play in mathematics. The category Def-A is intended for incidents where a student completely relies on everyday spontaneous conceptions, or answers a question with no reference to definitions. So I would personally code the incident as Def-C.

During the discussion, my interpretation of the boundary between Def-B and Def-C was neatly summarised by Coder 1:

> **Coder 1**: It's about referring to the content rather than [the] existence

Lastly, it was discussed whether Def-D (manipulates definitions) could occur when they are manipulated incorrectly, for instance combining two definitions improperly.

Because a phenomenographic account brackets how correct an answer is when considering a student's awareness of definitions, I consider this to be perfectly possible.

7.4.4 Discussion related to the dimension *representation of sequences*

Most of the discussion which was related to the categories within the representation of sequences dimension, focused on Student A's answers. Student A gave each of her examples as a (one-sided) list of numbers, and so arguably this dimension could not be richly coded for. For instance, giving all lists could be classified Rep-B (one particular representation is always superior), for instance there was one passage when both coders felt that the student was trying to find a list of numbers which corresponded to a graph she had sketched:

> **Coder 1**: It seems like she is looking for a different representation, but it doesn't seem that she thinks that's the only one that's superior.

It may be that only incidents where a student remarks that one type of answer is superior should be coded Rep-B, but it is also possible to infer that if a student gives examples in a certain representation and appears to want to give others in that representation also then this is also an incident that can be classified as Rep-B. In my interpretation of the interview with Student A there is too little comparison of sequence representations to determine the classification; it is not clear that she would be unsettled by giving an example in a different representation.

Student B was seen to have a Rep-B (any representation is suitable) approach throughout, and there was little discussion about this.

7.4.5 Discussion related to the dimension *construction strategies*

The first general comment that was made about construction strategies was that it was not clear if, to be coded as such, a strategy had to be commented upon or if it could be inferred from a series of actions or a final answer.

> **Coder 1**: For [Student A] there is naturally less on strategies because you don't hear her talking about what she's doing as she does it.

Incidents presented in the construction strategies section of the main study (Section 6.4) were usually based on the comments made by the students, rather than inferences made by the final answers given. In the transcripts used in the validation exercise there were fewer such incidents.

Both coders decided to code for inferred meanings and usually these corresponded well. Discrepancies included:

- Do Con-B (trial and error) and Con-C (transformation) strategies always need a sequence to start with?
- Must the representation of the sequences within a Con-C (transformation) strategy remain the same?
- Is transformation the same as thinking with (different) graphs?

These questions, and the coding of extracts based on them, are probably the result of not giving the coders examples from my data. My view of the answers to these discrepancies (yes, no, no) is clearer when the incidents from the main study data are read in conjunction with the category classifications within the sequence construction strategies dimension of variation.

7.4.6 Discussion related to the dimension *justifications*

The main concern from Coder 2 with regard to the justifications dimension was the scope of the justifications which were to be included. For instance, should a justification which is in relation to a strategy used (rather than the answer given) be included?

> **Coder 2**: At one point I asked myself: am I looking for reasons for the strategy used, or reasons for the answer given?

This concern is an unfortunate consequence with the naming of the dimension. When the validation task was completed the name of the dimension was 'justification of answers', which implied that the dimension was only focused on students comments related to the answers they gave, rather then their justifications more generally.

The second concern was again based in interpretation of the category names, in particular the informal/formal distinction. For instance, was this the same case as informal/formal mathematically? Also, in the words of Coder 1, 'did formal justification have to be right?'

The way I envisioned the dimension, formal justifications were those judged by an expert to be suitable, relative to the statement which is being justified. So a justification can be in a suitable format, but with an incorrect warrant, for instance. I also think it is possible to have a formal justification which uses a mathematical basis, but contains little formal mathematics (for instance, "the sequence 1,2,3,... is increasing, because the definition says that each number must get bigger"), whereas an informal justification may be something where a meaning is given, but it is not related to formal mathematics (such as, "the sequence 1,2,3,... is increasing, because it clearly is").

An interesting comment was made in relation to the separation aspect of dimensions of variation. When a student 'uses' or 'manipulates' a definition, it is sometimes for justification purposes, and so it may be the case that these categories often coincide.

Coder 1: When they refer to the definitions and they, at the same time, give an informal justification, they often came together as a natural thing.

While I feel this is true, I still believe the categories are separable; although it may not be possible to justify an answer without using a definition, other justifications (for instance targeted at the strategy used) are possible with no reference to definitions.

7.5 Implications for communicative validity

Recall that communicative validity addresses to what degree a coding framework is internally consistent, whether it offers a persuasive interpretation of data, and whether it can be applied to new datasets. Both coders felt able to code the transcript of student A and B with the dimensions of variation outlined and described in Chapter 6. Each believed the structure of the dimensions allowed them to be consistent in their coding both within and between the two transcripts. This implies that the dimensions can be applied to new data unrelated to the main study.

The exposition in the previous section has focused primarily on the discrepancies between the coders' choices, and such discrepancies have typically been at the level of distinguishing between categories within a dimension of variation. Coder 2's decision to associate a category from each dimension of variation to every question in a transcript resulted in some discrepancies in her choice of codes compared with Coder 1, but once the coders had discussed the meaning of the different categories, they found they could agree on the meaning of each category, and to what data it is most applicable. Moreover, these meanings are consistent with my interpretation of the codes also. There was discussion on the boundaries of the categories within some dimensions, but this is inevitable when the distinctions are fine and subtle. Overall I believe this indicates that the dimensions of variation are internally consistent.

Neither coder felt that new categories need be added to any dimension, but both coders felt that it would be beneficial to code answer sheets as well as transcripts for a data analysis to be more complete.

7.6 Implications to pragmatic validity

Recall that pragmatic validity addresses to what extent the research outcomes are seen as useful. This is arguably tougher to address than communicative validity; it depends on a reader's view of educational research and the application of a phenomenographic approach to data analysis. The two coders that took part in the validation exercise were both familiar with qualitative approaches to data analysis, and felt that the phenomenographic approach was appropriate for the data.

I would argue that the validation exercise reported in this chapter is not a very good way of addressing pragmatic validity in general. Although both coders agreed without prompting from me that they had a greater insight into example generation and students' awareness of sequences after the exercise, this can not necessarily be attributed as a result of coding for the dimensions of variation. For instance, such insights may just be a consequence of reading through the transcripts.

7.7 Summary of chapter

This chapter has presented the results of a validation exercise which addressed the communicative validity, and to a lesser extent pragmatic validity, of the dimensions of variation framework presented in Chapter 6. In the validation exercise, active researchers in mathematics education were presented with new data to code within the framework, and then were asked to reflect on their choices of codes and the validity of the framework. Both researchers felt that the structure and content of the dimensions were appropriate to the data, internally consistent, and provided insight into the data.

Chapter 8

Using Rasch Analysis to Validate the Task

This chapter uses a statistical technique — Rasch Analysis — to help determine how characteristic the students interviewed in the main task were, relative to the population of first-year undergraduates from which they were recruited. First, Section 8.1 introduces some of the theory behind Rasch Analysis, describing briefly its use in educational research to date. This section can be omitted for readers who are familiar with the technique, although please note that the notation and terminology used in this chapter can be found in Subsection 8.1.1.

Next, Sections 8.2 and 8.3 describe the origin of the data and its suitability for the Rasch Model, respectively. Then Section 8.4 describes how Rasch Analysis has been used to (1) objectively rank the questions in terms of difficulty relative to my population, (2) compare the students I interviewed in the main task with the general population, and (3) determine how characteristic (in a statistical sense) the interviewed students are in relation to the general population. This last use helps address, to a certain extent, concerns that are inherent with self-selecting or 'opportunistic' samplings of my population, and whether the think-aloud protocol may have affected student performance on the task.

Some of the material in this chapter has been published in a similar form (Edwards, 2010; Edwards and Alcock, 2010b), in particular the introduction to Rasch Analysis, but the material has been edited for inclusion in this thesis.

8.1 Description of Rasch Analysis

The Rasch Model and its application to data, Rasch Analysis are Item Response Theory (IRT) models. Rasch Analysis aims not only to rank questions by their difficulty, and rank students in terms of their performance on the test, but also to measure how likely a particular pattern of answers many be (for instance it is unlikely a student scoring well on a test will answer an easy question badly). The Rasch Model is one of the more commonly used models in IRT (Baker and Kim, 2004), in part because the routines associated with it can be shown to converge to a unique solution if given 'well-conditioned' data (Fischer, 1981, see also Section 8.3 of this chapter).

The use of IRT models has recently increased in popularity due to the introduction of affordable technology capable of quickly running the routines (Kline, 2005, p.107), and Rasch Analysis has been used as the basis of analysis for a diverse range of studies. These include the analysis of Likert-scale questionnaires to determine course satisfaction (Waugh, 1998), and attainment (Haines and Crouch, 2001). Rasch Analysis has been used both to analyse (Ryan and McCrae, 2006), to construct (Chen et al., 2005), and to use (Dobby and Duckworth, 1979) banks of test items.

Within mathematics education, the Rasch Model is "being used increasingly as a research tool by 'mainstream' researchers rather than merely by the sophisticated psychometricians involved in large-scale achievement testing"' (Callingham and Bond, 2006). This quote was taken from the editorial of a special issue on Rasch Analysis of the Mathematics Education Research Journal. Articles in the journal included Watson et al.'s (2006) longitudinal study of students' understanding of chance and probability, and Bradley et al.'s (2006) use of Rasch Analysis to explore Likert data of student's conceptualisations of quality mathematics instruction. Other articles in this special issue found their data was unsuitable for Rasch Analysis (Stacey and Steinle, 2006), or combined Rasch Analysis with other statistical techniques, such as Factor Analysis (Grimbeek and Nisbet, 2006).

In terms of using Rasch Analysis to determine the likelihood of a particular patterns of answers (in other words how characteristic is a set of answers?), Ryan and Williams (2007) have used this idea to produce personalised assessment feedback to trainee teachers, highlighting which areas of mathematics needed more attention. Other authors have explored the 'person-fit' statistics (these are a measure of the likelihood of a set of responses) via additional statistical techniques (Emons et al., 2005; Reise, 2000). As far as I am aware, no author has used Rasch Analysis to compare an interview cohort with the associated wider population of students.

8.1.1 Terminology

The Rasch Model assumes that for each person taking the test there is a parameter measuring their ability (or performance on the test), and for each item on the test there is a parameter measuring the item's difficulty. The notation used for these parameters differs from text to text (Fischer and Molenaar (1995) use ν_j, β_i whereas Baker and Kim (2004) use η_j, δ_i). In this section, and beyond, I shall be using the following notation and terminology:

Item A question on a test. The item in focus at any particular time is given the label i, and there are always j items in a test.

Person An individual taking the test. The person in focus at any particular time is given the label p, and there are always n people in the test.

Test A set of j items, attempted by n persons. It is not necessary that each person answers each question, but in my task this was the case.

Task When I refer to the task outlined in the main study chapter, it is 'the task', with $j = 11$ and $n = 147$ (or $n = 162$ if interviewed students are included). When I am referring to test instruments in general I use 'a test', as outlined above.

Answer The word 'answer' is used interchangeably with the score (of 1 or 0) person p obtained for item i. It is often written as $x_{p,i} \in \{0, 1\}$. The matrix of test answers (persons as rows, items as columns) is the $n \times j$ matrix $\mathbf{X} = ((x_{p,i}))$.

Total scores The number of students answering item i correctly is the item total, $s_i = \sum_p x_{p,i}$. The raw score of a person is $r_p = \sum_i x_{p,i}$.

Difficulty parameter Associated to every item, $i = 1, \ldots, j$ there is a difficulty parameter, δ_i. This is an idealised property that can only be estimated. We call these estimates D_i (*estimated difficulty of an item*).

Ability parameter Associated to every person, $p = 1, \ldots, n$ there is an ability parameter, ξ_p. This is an idealised property that can only be estimated. We call these estimates A_p (*estimated ability of an person*).

The distinction between δ_i, D_i (and ξ_p, A_p) is an important one, because in order to estimate the parameter δ_i we make certain assumptions for D_i. For instance, in the analysis which follows, we arbitrarily centre D_i so that the mean across all items is equal to zero.

8.1.2 Definition of the Rasch Model

The Rasch Model, given in Equation (8.1), gives a probability that person p answers item i in a test correctly, provided certain assumptions hold (these assumptions are discussed in the next section). This success-probability is a function of A_p and D_i alone:

$$P_{p,i} = P(x_{p,i} = 1) = \frac{\exp(A_p - D_i)}{1 + \exp(A_p - D_i)} \tag{8.1}$$

The graph of this success-probability against $A_p - D_i$ is drawn in Figure 8.1. Notice that because the Rasch Model gives the probability for person p answering item i correctly, the equation only makes sense when we are considering the difference between A_p and D_i, rather than just considering one of these quantities.

It is this distance between a person's ability and an item's difficulty that is a measure of how well-suited the item is for the person. The value $A_p - D_i$ is defined because the parameters are both on the same scale. For instance, if a person and item are judged to have the same parameters (so $A_p - D_i = 0$), the probability that the person answers the item correctly is defined to be 0.5. The choice of scale for the parameters is arbitrary; all that is required is that for a fixed distance between a person and an item there should be a fixed probability of answering the item successfully.

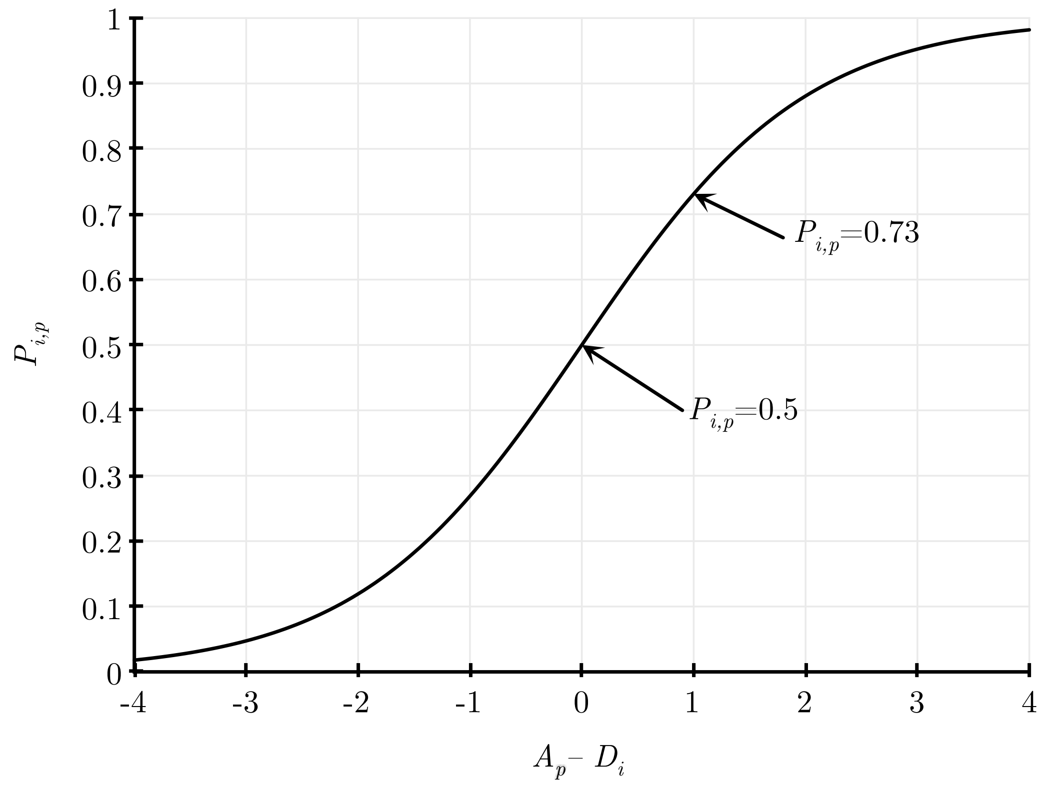


FIGURE 8.1: The shape of the Rasch Model, defined by Equation 8.1.

The Rasch Model uses a logit (*log*istic un*it*) scale for A_p and D_i. To explain how this scale works, note that by taking natural logarithms of Equation 8.1 we obtain Equation 8.2.

$$\log_e\left(\frac{P_{p,i}}{1-P_{p,i}}\right) = A_p - D_i \tag{8.2}$$

This value, $\log_e(P_{p,i}/(1-P_{p,i}))$ is the logit of $P_{p,i}$, which represents the log-odds of answering a question correctly. By expressing A_p and D_i in logits we can make direct comparisons between different person and item parameters.

A more difficult item or a more able person's parameters are larger positive numbers, and easier items/less able persons have larger negative values. In the task used in the main study (and arguably most tests in general) it is desirable for almost all persons to have a success-probability between 0.05 and 0.95 for each item, or else the items are much too difficult or much too easy for the population. Using Equation (8.1), these success-probabilities correspond approximately to $A_p - D_i$ lying between -3 and $+3$ logits.

Given Person 1 and Person 2, with associated parameter estimates of $A_1 < A_2$, the success-probability for Person 1 will be lower than Person 2 no matter which item is under consideration. This does not mean that the model of success-probabilities is linear, however, as illustrated in Table 8.1. In this table we fix two A_p and two D_i, but note that the success-probability from an easier to a harder item decreases more for one of the persons. This table is best read in conjunction with Figure 8.1.

TABLE 8.1: The success-probabilities of persons for different values of A_p and D_i. Note that Person 1's success-probability decreases by a larger amount than Person 2's success-probability.

		Item 1 $D_1 = -1$	Item 2 $D_2 = 0$	Success-probability difference $D_1 \to D_2$
Person 1	$A_1 = 0$	0.731	0.500	-0.231
Person 2	$A_2 = 1$	0.881	0.731	-0.150

8.1.3 Assumptions of the Rasch Model

As described in the last section, the central assumption of the Rasch Model is that when a person is faced with an item in a test they have a certain probability of answering that item correctly, which is based solely on the parameters δ_i and ξ_p. This means it is relatively quick to compute the A_p and D_i estimates of the parameters, and it is easy to interpret the results, but it does prompt questions as to how it is possible to claim that a person has a latent trait called 'ability' and an item has a latent trait called 'difficulty'. One way of addressing this claim is to emphasise that the Rasch Model is a stochastic[1] model of what happened on one particular test, not an attempt to describe actual reality in any philosophical or psychological sense (Rasch, 1980, p.110). However, the assumption that there is only one parameter per item or person does mean that the test should be unidimensional, that is the test should only assess a single construct (Kline, 2005, p.109). This assumption also means that the Rasch Model is unsuitable for tests where another factor may have an influence, such as the inclusion of a 'guessing' factor.

[1]i.e. a model which does not predict the outcome of an event, but focuses on the probability of an outcome's occurrence

The assumption of unidimensionality in a Rasch Analysis may seem incompatible with the multiple dimensions of variation that formed the outcome of the phenomenographic analysis of interview data presented in Chapter 6. This is not the case, however, because there is a difference in the use of the word 'dimension' in the two contexts. In Rasch Analysis, 'dimension', or more specifically the assumption of unidimensionality in the data, refers to the presence of a single unknown parameter that predicts how well a student might perform on a task. When estimating this parameter, we consider only whether questions were correctly or incorrectly answered, and there are various statistical tests that can be performed to suggest if the data is unidimensional (see Section 8.3). In the context of the phenomenographic analysis of the interview data (Chapter 6), 'dimensions of variation' are a lens with which to describe different aspects of how the cohort of students interacted with the task. The dimensions do not aim singly or collectively to measure how good a student is at answering a specific set of questions; they instead consider aspects of students' reported interactions with the task. It is, therefore, perfectly possible to have a unidimensional set of questions from the perspective of a Rasch Analysis, but where students' interaction with the task can be described phenomenographically by a set of dimensions of variation.

A further assumption of the Rasch Model is local (stochastic) independence of the test items, that is, that the probability of answering one item correctly should not be dependent on the answers to other questions. This assumption considerably simplifies both the use of the model and the estimation of the parameters δ_i, ξ_p. The dichotomous Rasch Model is therefore not suitable for tests where the questions are structured into parts, with the latter parts relying on the earlier ones (although there are ways around this). A full, detailed discussion of the assumptions of Rasch Analysis is given by Fischer and Molenaar (1995).

8.1.4 Validating the model

In a review of the Rasch literature Tennant and Pallant (2006) found that there are three main approaches to assessing (uni)dimensionality.

1. Using classical approaches (e.g. principal component analysis)

2. Those which consider that once the Rasch Model has been applied, a good fit with the model reflects unidimensionality

3. Those which involve some post-hoc testing, assuming that there is a good fit to the model

My own literature search confirms Tennant and Pallant's: different authors are concerned with the assumption of unidimensionality to different extents. Some authors make no reference to the unidimensionality of their data, implicitly assuming that if the analysis shows a good fit with the Rasch Model then their data is unidimensional (Chen et al., 2005; Lawson, 2006; Misailidou and Williams, 2003; Waugh, 1998). This is also reflected in the following quote from Bond and Fox (2007):

> Generally, practitioners of Rasch measurement rely on the indicators of misfit to reveal the extent to which any item or person performance suggests more than one underlying latent trait is at work. (Bond and Fox, 2007, p.251)

Some authors have offered qualitative arguments for the unidimensionality of their data (Coe, 2008), whilst others argue that classical approaches can be used successfully in conjunction with a Rasch Analysis. For instance Smith (1996) examined simulated data sets and found both principal component analysis and methods examining the fit of the Rasch Model work in a variety of situations. Green (1996) argued that when the data is unidimensional then both principal component analysis and a Rasch Analysis are successful, but when items are less closely related, the correspondence is imperfect, and the different analyses resulted in different definitions of scale.

In section 8.3 I examine the validity and unidimensionality of my data by looking at Cronbach's α and a principal component analysis.

8.1.5 Applying the model to test data: Rasch Analysis

Given a test that produces dichotomous data (answers are either correct or incorrect), and assuming the assumptions of the Rasch Model, a Rasch Analysis of the data is conducted in two stages:

1. Produce parameter estimates A_p and D_i for each person and item;

2. Test fit of model to the original dichotomous data.

It has been shown that if the data is 'well-conditioned', there will be a unique A_p and D_i for the data, and that procedures to estimate the parameters will converge to these values (Fischer, 1981). In Section 8.3.4, it will be shown that the data from the example generation task is well-conditioned. For now we assume that the data is well-conditioned.

Estimating the parameters of the model

In the first stage, persons with zero or all items correctly answered, and items on which every person was (un)successful are removed from analysis. This is because most methods of estimating the ability/difficulty parameters of such persons/items would result in parameters of $\pm\infty$. This is undesirable; for instance taking the case of assigning $A_p = +\infty$ to someone scoring full marks would imply that the test contains the hardest question that it is possible to write. In most test data (including my data), there are many more persons than items, and so it is rare for an item to be removed at this stage.

After such items and persons are removed, there is the possibility that new items and persons now score zero or perfectly, and so these must also be removed from the initial analysis. In reality, this occurs rarely (Fischer and Molenaar, 1995, p.42), and such persons and items removed at this and previous stages are re-introduced after the other persons and items have had their parameters estimated.

There are a variety of ways of proceeding with the parameter estimation from this point and whole books have been written on the different choices (for instance see Baker and Kim, 2004). Some procedures take an iterative process where parameters are first essentially guessed, then these guesses are used as estimates in a procedure to produce a second set of estimates and so on. Others attempt to jointly estimate both sets of parameters.

The data analysis presented in this chapter uses a combination of the 'Normal Approximation Algorithm' PROX to produce the initial estimates and then from these estimates run the 'Joint Maximum Likelihood Estimation' (JMLE) algorithm to give more precise estimates. This combination is popular in the literature (Bond and Fox, 2007; Wright and Masters, 1982), and is the default option on the computer program used to run the analysis: Winsteps (Linacre, 2009).

The PROX routine takes as an initial guess for each item the logit of the proportion of persons who answered it correctly, and does similarly for the abilities of persons. Then in an iterative process described by Wright and Masters (1982, p.64) it re-estimates the item difficulties based on the estimates of person ability (so that the greater the variance of person ability, the greater the variance of item difficulty), and vice-versa. When the rate of increase of range of A_p and D_i slows to a certain level (which is set to 0.5 logits per iteration), the procedure ends.

The JMLE routine takes as its guess estimates the output from PROX. It applies the Rasch Model (Equation 8.1), to each $\{p, i\}$ pair, and then sums these expected values to obtain item and person estimated raw scores. It compares these predicted raw scores with the observed scores, making minor alterations to A_p and D_i depending on how the estimated and observed scores differ. This results in new estimates of A_p and D_i, and the procedure is repeated. Fischer (1981) proved the existence and uniqueness of a JMLE solution (i.e. that this iterative process converges), dependent on the necessary and sufficient condition that the data is *well-conditioned*. I show in Section 8.3.4 that my data is well-conditioned, and so a unique solution exists for D_i and A_p using JMLE estimation. This is not the same as assuming that the unique solutions are equal to δ_i and ξ_p.

Testing the fit of the model

After the parameters D_i and A_p have been estimated, the Rasch Model (Equation 8.1) is used to produce a matrix of success-probabilities for each $\{p, i\}$ pair:

$$\mathbf{S} = ((P_{p,i}|A_p, D_i))$$

It is then possible to make comparisons between the data matrix $\mathbf{X}$ and the success-probability matrix $\mathbf{S}$. The closer the fit between these two matrices, the better the Rasch Model is at modelling the data. To compare how well the data fits person-by-person, we can compare corresponding rows of the matrices. To compare how well the data fits item-by-item we can compare corresponding columns of the matrices. Finally, we can compare on a person-item level by comparing individual cells of the matrices. The most common comparisons ('outfit' and 'infit') are based on chi-squared tests, and are outlined below. We say that an item or person underfits the Rasch Model when the response string is noisy or erratic (for instance a student might score highly but answer an easy question incorrectly). A person or item overfits the model if the response string is almost too good to be true, which may indicate a lack of local independence of the items (Bond and Fox, 2007, p.241).

To obtain a person's *outfit* to the model, the expected score of this person answering each question is calculated from Equation 8.1, and subtracted from the observed score to obtain a residual. These residuals are squared and standardised, and then an average of the squared standardised residuals is calculated (Wang and Chen, 2005). These *outfit mean square* values can then be transformed into a t statistic that follows approximately the standard normal distribution (Wilson and Hilferty, 1931). Outfit statistics are outlier-sensitive, which means they tend to be dominated by persons(items) which have unexpected responses for the items(persons) deemed much tougher(more able) or easier(less able) compared with the person(item)'s measure (Wright and Masters, 1982, p.99).

A person's *infit* to the model is information-weighted, that is, each standardised residual is first weighted according to its variance. These weighted standardised residuals are then summed, and divided by the sum of the variances. Similarly to the outfit value, these *infit mean square* values can be transformed into a t statistic. Infit statistics tend to be dominated by persons(items) which have unexpected responses for the items(persons) with similar measures.

As described by Bond and Fox (2007, p.239), an infit or outfit mean square value of v corresponds to $100(v-1)\%$ more variation between the observed and the estimated values than expected by the model, and infit or outfit t-values greater than $+2$ or less than -2 can be interpreted as having not being compatible with the model ($p < .05$).

8.2 The data from this study

8.2.1 Data collection

Participants

The participants in the validation study reported in this chapter were first year students at Warwick University. These students were in the same year as the students interviewed for the main study (see Chapter 5) and they took the task in class between one and three weeks after the final interview took place.

The first Rasch Analysis of data only includes students that completed the task in the classes (rather than in interviews). There were 147 such students, once those that had taken the task both in an interview and class were removed. For the Rasch Analysis which also included the 15 interviewed students (Section 8.4.3 onwards), the size of the dataset increases to 162.

The definition and task sheet

Each student was provided with the same definition sheet as given to the students interviewed during the main study.

The same questions that were asked during the main study interviews were included on the task sheet. The ordering of questions was unchanged, but the space between questions was reduced so that they could fit on a single page. A box was added at the top of the page asking students to indicate if they had already completed the task sheet in an interview (tutors were asked to remind students to tick the box if appropriate). Four students in the sample of 151 students indicated they had previously answered the questions in an interview and were removed from the analysis.

There were 11 classes in total, and 7 classes returned sets of completed sheets. A breakdown of the number of students, the mean score and standard deviation of scores is given in Table 8.2. Most classes had 15-35 students present, and the average scores for each group were high ($> 8/11$ questions answered correctly). It is also worth noting that, in general, students did better on the task in class than they did in interview.

Table 8.2: Descriptive statistics for each class group, with the interviewed students' statistics included for comparison.

Class	No. students	Mean score	SD
A	38	9.26	2.10
B	30	8.83	2.48
C	25	10.04	1.02
D	4	10.00	0.82
E	31	9.06	1.79
F	19	8.53	2.50
Total	147	9.19	2.05
Interviewed	15	7.40	2.38

8.2.2 Marking the task sheets

Each answer was marked in conjunction to the taxonomy given in Section 4.2 (i.e. each question was categorised **C**, **B**, **IS** or **INS**), and the results tabulated in a spreadsheet. This data can be found in Appendix C. For the Rasch Analysis, dichotomous (correct or incorrect) data must be entered so the categories **IS**, **INS**, and **B** were combined so that any student answering in these ways would have $x_{p,i} = 0$; correct answers were coded as $x_{p,i} = 1$.

8.2.3 Ethical considerations

As with all the studies in the thesis, Loughborough University's ethical checklist was completed beforehand. No question on this checklist raised further issues.

Similar to the ethical considerations discussed in the context of the first pilot study (Section 4.1.3), an important ethical consideration with this study was that of using student problem class time to collect research data. Initially contact was made with the course lecturer, and permission was granted to give the task during problem classes. The lecturer did not consider the content of the task to be greatly different from the tasks given routinely in the classes, either in terms of the type of questions asked, or the difficulty of these questions. Furthermore we both agreed there were clear pedagogic reasons to encourage students to complete and reflect upon such tasks. Each problem class tutor was briefed on the questions students typically found difficult, and it was suggested that the tutors go through these questions in more detail with the students.

By doing this, the students would benefit from the activity in the same way they would from the standard problems given out each week.

8.3 Tests of validity and suitability

As described in Subsections 8.1.3 and 8.1.4, before conducting a Rasch Analysis of dichotomous data there are a variety of tests that can be performed to check that the data is suitable for the procedure (Fischer, 1981; Glas and Verhelst, 1995; Kline, 2005; Tennant and Pallant, 2006). Following such recommendations, this section explores the data's internal validity and dimensionality first by calculating Cronbach's α and its associated error in Subsections 8.3.1 and 8.3.2, and then by performing a principal component analysis in Subsection 8.3.3. Then, in Subsection 8.3.4 the data is shown to be well-conditioned (in the sense of Fischer, 1981), so that the JMLE routine to estimate the A_p and D_i parameters will have unique solutions.

8.3.1 Cronbach's alpha

Cronbach's alpha (Cronbach, 1951), referred to as alpha, is a measure of the reliability of a set of items (i.e. their internal consistency, which refers to how interrelated the items in a test are). It is considered a persuasive index, so much so that it has almost become synonymous with reliability (Kline, 2005, p.174). Alpha is related to a test of reliability called 'split-half reliability', where the dataset is split randomly into two halves (by persons, not items), scores for each item are calculated in each half, and the correlation is calculated between the two halves. One would expect a good test to have a high correlation. The alpha statistic is equivalent to taking the mean of all the possible split halves (Cronbach, 1951), and can be defined as

$$\alpha = \frac{j}{j-1}\left(1 - \frac{\sum_{i=1}^{q} \sigma_i^2}{\sigma_T^2}\right)$$

where j is the number of items on the test, σ_i is the variance of item i, and σ_T is the variance of the entire test, that is

$$\sigma_T^2 = \sum_{i=1}^{j} \sigma_i^2 + \sum_{i \neq k} \sigma_{ik}$$

For my data, $\alpha = 0.792$, which is above the 0.7 threshold considered reliable (Cohen et al., 2007, p.506). It is close enough to 0.7 to consider alpha's standard error (see next section).

Whilst alpha is a valid statistic for reliability, it should not be considered as a statistic which measures the unidimensionality of items (Cortina, 1993; Field, 2009). In fact, alpha is an underestimate of reliability unless the data is unidimensional (Schmitt, 1996, p.350).

That alpha is not a measure of unidimensionality can be demonstrated by constructing sets of data with two or more underlying dimensions and then showing that alpha for such data is above 0.7 (Field, 2009). So the alpha for my task indicates that the items are reliable, which is a good thing, but the alpha does not tell us that the test is unidimensional.

8.3.2 Standard error of Cronbach's alpha

It is a well-known fact that the value of alpha increases with the number of items on a test (Cortina, 1993, p.101), and so it is appropriate to also consider the standard error of alpha, which is the error in the calculated alpha relative to the true alpha. Various methods have been proposed to calculate this error (Duhachek and Iacobucci (2004) list seven). The following will be used in this thesis:

1. Bootstrapping
2. Cortina's formula
3. Duhachek and Iacobucci's formula.

The second and third methods were published in papers that were concerned with the misuse of alpha (Cortina (1993) and Duhachek and Iacobucci (2004), respectively), and the first was suggested to me by a colleague at Loughborough.

Bootstrapping is a computer-intensive method where the standard error of a statistic is estimated by treating the data as a population and taking repeated samples (with replacement) from the data set (Efron and Tibshirani, 1986). A Matlab routine was written (see Appendix D.1) which took a sample of size $n = 147$ (with replacement) from my data and calculated Cronbach's alpha for this sample. The code was set to execute a million times, and from each run alpha was calculated. The variance of these million alphas was calculated to be $SE_{bootstrap} = 0.0494$. A histogram of the distribution can be found in Figure 8.2.

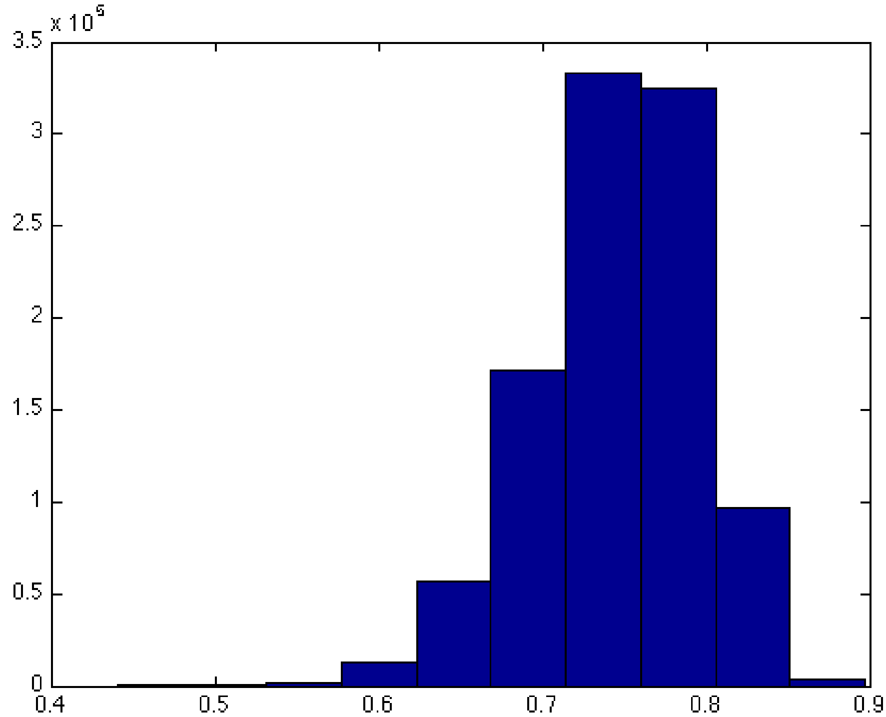


FIGURE 8.2: Bootstrap histogram for 1,000,000 values of alpha for my task data.

Application of the standard error formula given by Cortina (1993),

$$SE = \frac{SD_r}{\sqrt{0.5j(j-1)-1}}$$

where SD_r is the standard deviation of the item intercorrelations (0.149), gives $SE_{Cortina} = 0.0202$

Finally, the standard error formula given by Duhachek and Iacobucci (2004), based on van Zyl et al.'s (2000) work, was used:

$$SE = \left(\frac{2j^2}{(p-1)^2(d'Vd)}\right)\left((d'Vd)(trV^2 + tr^2V) - 2(trV)(j'V^2j)\right)$$

where d is a $j \times 1$ vector of ones, and V is the population covariance matrix among the items. The formula was applied via a Matlab code (for code, see Appendix D.2), and it

TABLE 8.3: 95% Confidence intervals for Cronbach's alpha, based on three methods of calculating alpha's standard error.

Method	SE of α	95% confidence interval
Bootstrap	0.0494	$0.774 < \alpha < 0.810$
Cortina (1993)	0.0202	$0.780 < \alpha < 0.804$
Duhachek and Iacobucci (2004)	0.0688	$0.770 < \alpha < 0.814$

gave the value $SE_{Duhachek} = 0.0688$.

All these measures of the standard error of alpha assume that it is normally distributed, and so we can calculate 95% confidence intervals for alpha based on the standard formula

$$\alpha \pm 1.96 \left(\sqrt{\frac{SE}{n}} \right) \tag{8.3}$$

Table 8.3 lists the SE and confidence intervals for the different calculations for standard error, and by each method we can see that alpha is comfortably above 0.7, the level associated with reliable internal consistency of a test.

8.3.3 Principal component analysis

The previous section's analysis of Cronbach's alpha gave a measure of the reliability of the test, suggesting how the task's might vary if it were asked on repeated attempts. The next analysis of the data prior to running the Rasch Analysis is a principal component analysis (PCA). This is a technique which will be used to explore the dimensionality of the data. Recall that to be suitable for a Rasch Analysis, items should be unidimensional. This means that a student's total score could reasonably interpreted as a measure of a single latent trait, in this case perhaps 'a student's ability to generate sequences which satisfy certain constraining properties'.

A PCA of the data looks to maximise the variance of linear combinations of the items (Rencher, 2002, p.380). The initial task items ($j = 11$) could be considered as eleven principal components which collectively describe all the variance in the data. The key idea of a PCA is is that if a linear combination of these items can explain much of the variance of the data then this principal component is in some ways a better descriptor of the data than the individual items. If a PCA gave four components which could account for a good proportion of the variance, then it may be the case that it is better to look at these four principal components than the original eleven items. I use the word 'may' because there is no statistical test to help a researcher make this decision. Instead, qualitative judgements must be made after the PCA.

PCA can be used as a preparatory tool to examine dimensionality of data (Tennant and Pallant, 2006), again subject to qualitative judgements on both the method of rotation[2] and the interpretation of results. Manly (2005, p.101) and Field (2009, p.642) both suggest the 'orthogonal varimax' rotation as a good choice for exploratory data analysis and so I have used this type of rotation (furthermore, running alternative rotation methods did not give me very different results). For each PCA reported in this section, I have used the program SPSS to run the analysis (SPSS, 2010).

If the data is unidimensional a PCA would find that there is one component with an eigenvalue much larger than the rest. Rencher (2002, p.397) suggests a variety of methods to decide what 'larger' means in this context, but notes that it is common to use a scree plot to look for a natural break or point of inflection in the eigenvalues.

Before the first PCA of the data, the Kaiser-Meyer-Olkin measure was used to verify the sampling adequacy (i.e. to check that the items are correlated enough to perform the PCA) for the analysis (Kaiser, 1970). For the complete data set with interviewed students removed $n = 147$, $KMO = .72$ ('good' according to Field, 2009, p.659), and all KMO values for individual items were $> .65$, which is above the acceptable limit of .5 (Field, 2009). Bartlett's measure of sphericity was used to test the null hypothesis that the correlation matrix was null (in which case it would not be suitable for a PCA). For my data, $\chi^2(55) = 536.26$, $p < .001$, which indicates that correlations between items were sufficiently large for PCA to be meaningful.

The PCA was conducted on the 11 items with orthogonal rotation (varimax), to obtain eigenvalues for each component in the data. The rotated component matrix for the first three components can be found in Table 8.4, and a scree plot of their eigenvalues in Figure 8.3.

The scree plot indicates there are two components before the point of inflection on the graph. The rotated component matrix suggests that the first component is an overall

[2]Because a principal component is a linear combination of the items, the problem of calculating components with maximum variance is equivalent to rotating matrices of correlations.

Table 8.4: Rotated Component Matrix. Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

	Component		
	1	2	3
Q1.	.169	.704	.206
Q2.	−.055	.410	.654
Q3.	.747	.063	.359
Q4.	.710	−.126	.242
Q5.	.177	.899	.033
Q6.	.616	.270	−.040
Q7.	.594	.291	−.340
Q8.	.153	.853	.011
Q9.	.285	−.009	.684
Q10.	.658	.197	.095
Q11.	.602	.152	.041

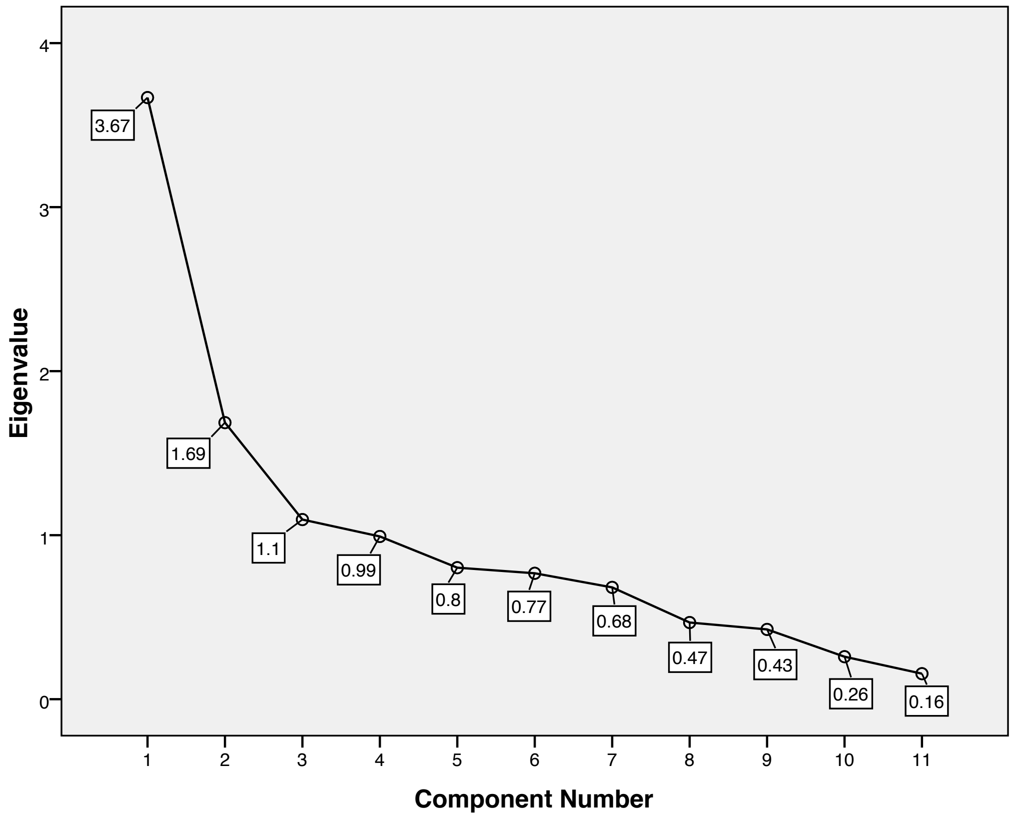


Figure 8.3: Scree plot for the PCA of all questions.

measure of performance on the test; the loading is positive for all but one question, and for the one question that is negatively weighted (Q2), its loading is much smaller than the rest. The second component is weighted highly for Questions 1, 5, and 8, and less so for the other questions. It is also worth nothing that for these three questions the loadings on the first factor is smaller. In summary, Component 2 can be considered a measure of how well students performed on Question 1, 5, and 8 and Component 1 can be considered to measure of how well students performed on the remaining questions. Recall that Questions 1, 5 and 8 were deliberately chosen because they were easy 'warm-up' questions. The PCA's second component may reflect the outcome that students who could not answer these three questions correctly would not score highly on the test overall, and so for this component, performance on the other questions has less of an impact (for the unrotated component matrix the values for these questions were negative).

It is not immediately clear whether the PCA has indicated that the data is unidimensional. If the reasoning in the previous paragraph is sound then the PCA's two components have highlighted the easier and harder questions on the task, both of which still are asking students to do the same thing. Looking at the components in more detail, the first component's eigenvalue is over twice the second, and this is reflected in the percentage variance explained by these components (33.35% and 15.33%, respectively). Moreover, if the proposed explanation for the second component is legitimate, then it too is measuring students' ability to produce examples in this setting, only that if a student is unable to answer easy questions they will do badly overall. This does not seem to be a further dimension in the data.

To clarify the situation, the PCA was re-run with Questions 1, 5, and 8 removed (but otherwise with the settings unchanged). This examines the task's dimensionality without these easier questions. Table 8.5 and Figure 8.4 present the rotated component matrix and scree plot for this second PCA. As before the Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, $KMO = 0.73$, ($> .63$ for individual items) and Bartlett's measure of sphericity $\chi^2(28) = 258.40$, $p < .001$ indicated that correlations between items were sufficiently large.

In this second PCA, the eigenvalue for the first component (which again can be interpreted as overall measure of performance on the task) is smaller (2.63 compared to 3.67),

TABLE 8.5: Rotated Component Matrix. Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

	Component		
	1	2	3
Q2.	−.032	.052	.779
Q3.	.854	.217	.197
Q4.	.880	.098	.032
Q6.	.083	.817	.175
Q7.	.208	.696	−.159
Q9.	.232	.099	.719
Q10.	.215	.712	.259
Q11.	.502	.388	.042

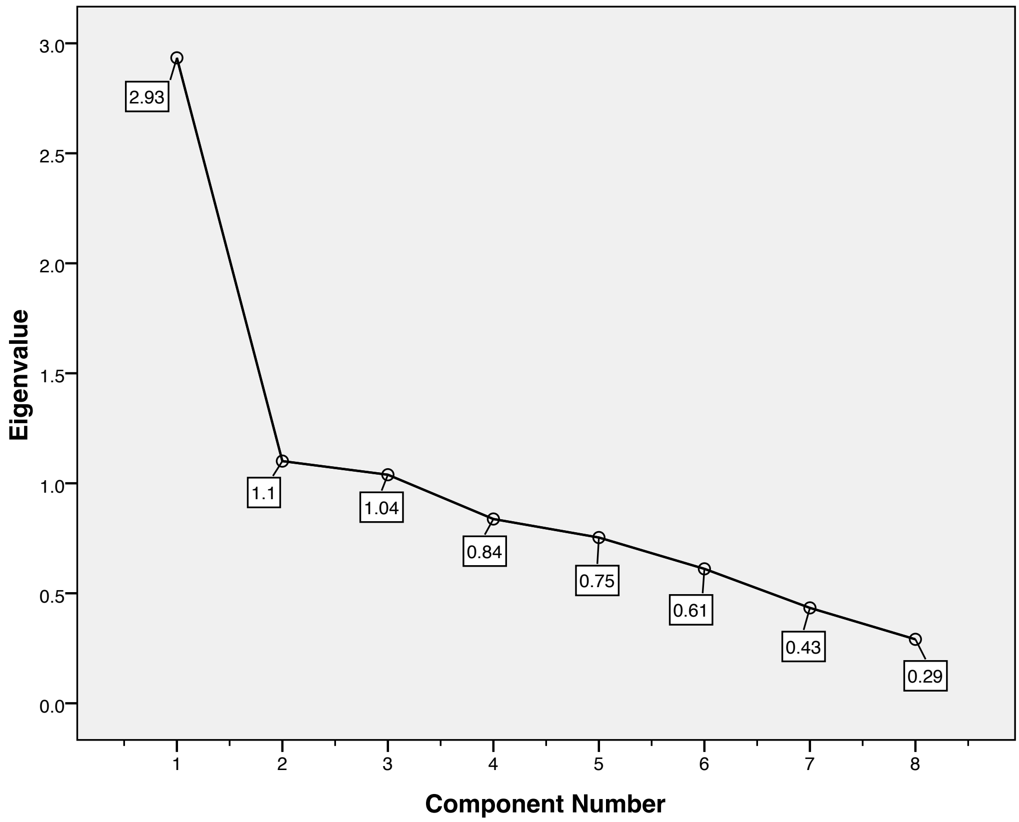


FIGURE 8.4: Scree plot for the PCA of the reduced set of questions.

but the percentage variance explained by this factor has increased slightly (36.67% compared to 33.35%). The eigenvalue of the second component is much smaller now that Questions 1, 5, and 8 have been removed (1.1). It now positively weights Questions 6, 7, and 10, but accounts for less overall variance (13.76%).

For the Rasch Analysis, outlined in the next section, all questions have been included in the dataset. This is for several reasons:

1. A principal component analysis will always give j components, and a qualitative judgement must be made as to whether the components identified suggest the data is multi-dimensional.
2. The second component (which weighted Questions 1, 5, and 8 more than the other questions) accounted for only 15% of the variance of the task
3. It is not clear that the presence of easy questions makes a test multidimensional; scoring badly or well on these items is still an indication of a student's ability to generate example of sequences
4. Removing these items served to highlight further items which made up the second component.

Other authors have suggested that if residuals are within acceptable limits after the Rasch Analysis (i.e. if the outfit and infit measures are t-values are between -2 and $+2$), then the Rasch Model is appropriate for the data (Tennant and Pallant, 2006). Subsection 8.4.2 will show that this is the case for the questions on the task.

8.3.4 Fischer's test of well-conditioned data

In the next section, the PROX/JMLE routine as outlined in Section 8.1.5 will be used to estimate the A_p and D_i parameters for each person taking the task and each item in the task. Fischer (1981) presented a necessary and sufficient condition for the existence and uniqueness of a solution for such routines (i.e. that they converge). The definition of this condition, that the data be *well-conditioned*, is not needed in this discussion, but can be found in Fischer (1981, p.59). In the same paper, Fischer proved the following lemma:

Lemma 8.1. *Let the items be ordered according to the number of persons answering them correctly,* $s_1 \leq s_2 \leq \cdots \leq s_k$*. Define* n_r *to be the number of persons with total score* r*. Then a n.s. condition for the data to be ill-conditioned is the existence of an index value,* k'*,* $1 \leq k' \leq j-1$*, such that*

$$\sum_{r=k-k'}^{k} n_r r = \sum_{i=1}^{k'} s_i + (k-k') \sum_{r=k-k'}^{k} n_r. \tag{8.4}$$

Proof. See Fischer (1981, p.63). □

To apply this lemma, first the dataset was reduced to those persons/items who did not score zero or a perfect score, because the PROX/JMLE routine is only run on the data with such persons and items removed. With this reduced dataset ($n = 100$, $j = 11$), n_r and s_i, were calculated (see Tables 8.6(a) and 8.6(b), respectively). Then for $k' = 1, \ldots, 10$, Equation (8.4) was calculated (note that Lemma 8.1 says that if the equation holds for any $1 \leq k' \leq j-1$ then the data is not well conditioned). Table 8.6(c) shows that for each k' the equation does not hold, and so the data is well-conditioned.

8.4 Applying the Rasch Model to my data

8.4.1 Comparing Rasch analyses with and without including the students who were interviewed

The first Rasch Analysis was run with only the students who were not interviewed. The dataset therefore consists of 147 strings of 11 dichotomous responses. Using the package Winsteps (Linacre, 2009), person ability and item difficulty parameters were estimated via the PROX/JMLE routines. The JMLE convergence criterion as set so that it would continue to iterate until the largest logit change was 0.0001.

The same procedure was then completed with the answers from the fifteen interviewed students included. This was to demonstrate that the item difficulty parameters from the first run and those from the second run were not significantly different. Table 8.7 gives the parameters for the two runs, together with the standard error for each estimate and

Table 8.6: Preliminary and final calculations for Fischer's (1981) equation for n.s. conditions of ill-conditioned data.

(a) Number of students answering r questions correctly

Total score r	No. of students n_r
1	1
2	0
3	0
4	2
5	5
6	6
7	14
8	9
9	25
10	38

(b) Questions ranked by the number of students answering them correctly, after students scoring zero or full marks have been excluded

Rank i	Qn	No. of students s_i
1	9	23
2	11	64
3	10	70
4	3	73
5	6	78
6	4	79
7	2	81
8	7	85
9	1	95
10	8	98
11	5	99

(c) Partial calculations of elements of Fischer's (1981) equation for ill-conditioned data, and the result of applying Equation 8.4

k'	$\sum_{r=11-k'}^{11} n_r r$	$\sum_{i=1}^{k'} s_i$	$(11-k')\sum_{r=11-k'}^{11} n_r$	Equation 8.4 holds?
1	380	23	380	No
2	605	87	567	No
3	677	157	576	No
4	775	230	602	No
5	811	308	516	No
6	836	387	485	No
7	844	468	396	No
8	844	553	297	No
9	844	648	198	No
10	845	746	100	No

the outfit t-statistic. The standard error is given by the given by formula

$$SE_{D_i} = \frac{1}{\sqrt{\sum_{p=1}^{n}(P_{p,i}(1-P_{p,i}))}} \tag{8.5}$$

where $P_{p,i}$ is the probability the model predicts person p answering item i correctly, as given by Equation 8.1. If we knew the 'true' item difficulty, δ_i, it would be expected that the estimate D_i to fall in the interval $(\delta_i - SE_{D_i}, \delta_i + SE_{D_i})$ 68% of the time. The outfit t-statistic was described in Section 8.1.5.

The difficulty estimate and standard error for the two analyses (the first four columns

TABLE 8.7: Estimated difficulty parameters for the data with(out) interviewed students included.

Item	Difficulty estimate		Model S.E.		Outfit (std)	
	without	with	without	with	without	with
Q1.	−1.65	−1.77	0.52	0.49	1.2	1.2
Q2.	0.28	0.34	0.29	0.27	1.4	1.2
Q3.	0.89	0.93	0.26	0.24	−1.9	−2.1
Q4.	0.44	0.69	0.28	0.25	0.9	1.6
Q5.	−3.85	−4.24	1.15	1.12	−1.7	−1.9
Q6.	0.52	0.27	0.28	0.27	−1.2	−0.3
Q7.	−0.10	−0.14	0.32	0.30	−0.4	−0.4
Q8.	−2.90	−2.75	0.83	0.69	−0.2	−0.4
Q9.	3.80	3.89	0.26	0.25	1.9	1.5
Q10.	1.09	1.28	0.26	0.23	−0.9	−0.3
Q11.	1.47	1.49	0.25	0.23	0.5	0.5

of Table 8.7) are plotted for comparative purposes in Figure 8.5, which shows that there is not a significant difference between the item difficulty parameter estimates when the interviewed students are included, although the difficulty estimate of Question 6 has decreased when the interviewed students were included (note that 10/11 interviewed students answered this question correctly).

In terms of fit to the model, only one item has outfit smaller than −2 or greater than +2, Question 3 after the interviewed students were included. The outfit mean square value for this question is 0.56, which corresponds to 44% less variation than expected by the model. In the first run of the model the question had 42% less variation than expected by the model, so in reality this is a modest change also.

Because the difficulty parameters of the items do not change significantly, for simplicity further analysis will be conducted with the dataset including all students.

8.4.2 Difficulty of items in the task

Using the item difficulty estimates in Table 8.7 we can order the items in terms of their difficulty, as presented in Figure 8.5. Note that the logit estimates reflect the metrics used in the pilot study evaluation (calculated in Table 4.1, and presented in task-sheet order in Figure 4.5), giving roughly the same ordering. The Rasch Analysis has indicated some differences from the first pilot study, notably that the pilot study found Questions

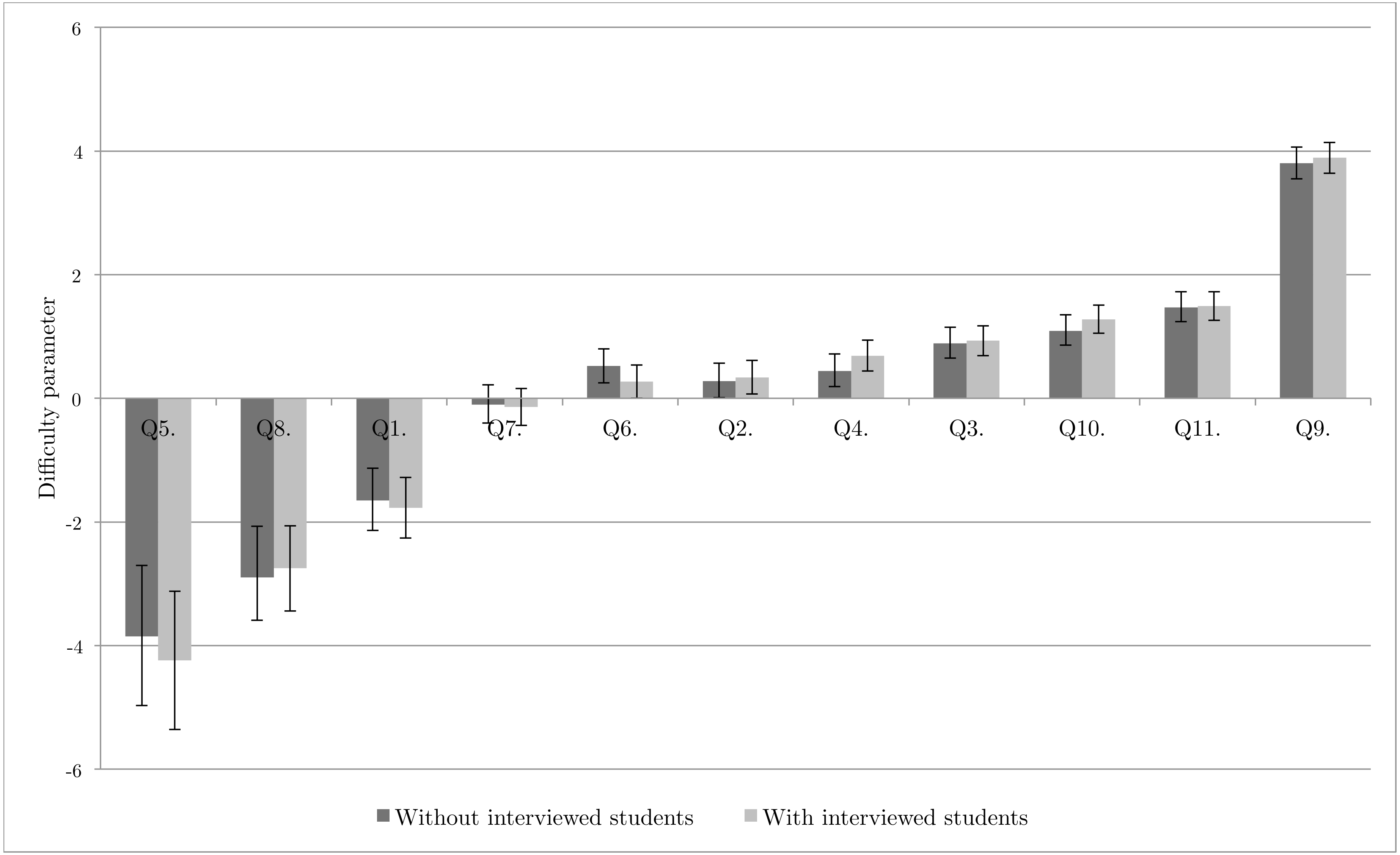


Figure 8.5: A comparison of the estimated difficulty parameters for the data with(out) interviewed students included, ordered by difficulty of question with all students included.

3 and 11 to have the same level of difficulty (0.87 and 0.86 respectively in the scale of the pilot study's metric), whereas the Rasch Analysis indicates that Question 11 is 0.56 logits more difficult.

The three questions which were included as easier warm-up questions, Questions 1, 5, and 8, have large negative D_i estimates. Their standard errors are large, which suggests that the majority of students' ability estimates were much higher than D_1, D_5, and D_8 (this is confirmed in the next section, Figure 8.7). Recall that $A_p - D_i > 3$ approximately corresponds to $P_{p,i} > 0.95$, and so if there are few students with ability parameter estimates within three logits of D_1, D_5, and D_8, then the accuracy of the estimation of these D_i parameters will be poor.

The remaining questions have estimates between -1 and $+2$ logits, with one question deemed much tougher than the rest at just under 4 logits. Note that exactly how much tougher is a question that can only be answered when we are considering a particular person, and so fixing a value of A_p in equation 8.1. For instance, the model predicts that a person with $A_p = 2$ is around five times as likely to answer Question 10 correctly than Question 9 ($P(Q10) = 0.673$, $P(Q9) = 0.131$), but this difference changes to almost eleven times for a person with $A_p = 0$ ($P(Q10) = 0.218$, $P(Q9) = 0.020$).

8.4.3 Rasch Analysis of interviewed students

I now turn my attention to the students interviewed in the main study. Their ability estimates, standard errors, outfit and infit statistics are given in Table 8.8. Note that each of the standardised outfit and standardised infit measures is in the $(-2, +2)$ region, which indicates each interviewee's series of responses is compatible with the Rasch Model (Tennant and Pallant, 2006). The range of ability parameter estimates for the interviewed students is less than for the population, but this is a reasonable difference when there were only 15 interviews.

The ability estimates, standardised outfit and standard errors are represented in a Rasch pathway diagram in Figure 8.6. Pathway diagrams are used to represent three-dimensional data, and authors such as Bond and Fox (2007) have used them to represent the output of a Rasch Analysis. This diagram plots a bubble for each interviewee. The bubble's centre is in plotted with the person's ability measure in the vertical direction,

TABLE 8.8: Rasch Analysis statistics for the interviewed students.

Student	Score	Ability	S.E.	Infit		Outfit	
				mnsq	std	msq	std
Anna	7	1.02	0.76	1.11	0.47	0.75	0.28
Ben	10	3.58	1.25	0.33	−0.8	0.11	−0.64
Valter	9	2.40	0.95	0.70	−0.35	0.41	−0.08
David	7	1.02	0.76	0.78	−0.74	0.53	0.07
Edha	2	−2.56	1.07	1.52	0.91	1.73	0.90
Phalgun	8	1.64	0.82	0.63	−0.97	0.39	−0.10
Guan	8	1.64	0.82	0.80	−0.41	0.53	0.06
Haaroon	4	−0.77	0.84	0.51	−1.06	0.34	−0.39
Ian	5	−0.12	0.78	1.02	0.16	0.86	0.22
Joe	6	0.46	0.75	0.90	−0.28	0.67	0.08
Ken	10	3.58	1.25	0.33	−0.8	0.11	−0.64
Laura	8	1.64	0.82	0.80	−0.41	0.53	0.06
Mike	10	3.58	1.25	1.86	1.09	1.07	0.52
Nicola	7	1.02	0.76	0.89	−0.31	0.67	0.20
Oksana	10	3.58	1.25	1.86	1.09	1.07	0.52

standardised outfit in the horizontal direction, and the bubble's width is plotted to reflect the standard error of A_p (the circle's diameter is set to equal the SE measure in logits). I will briefly discuss the diagram in conjunction with the response strings of the interviewed students before replotting the diagram to include the students who were not interviewed (in Figure 8.8). This is in order to give a comparison between interviewed and non-interviewed students.

Concentrating on the standard error (the size of the bubbles), we can see that the standard error is smaller for persons with ability estimates between $+2$ and -1 logits, and larger for more extreme ability estimates. This corresponds to the majority of items in the task having difficulty measures within this region (see the previous section), hence a better estimate for these students can be made.

Looking at the location of the bubbles, we see that at the top of the diagram there are four bubbles centred at a height corresponding to $A_p = 3.58$. This corresponds to answering 10/11 questions from the task correctly. Two bubbles slightly overfit the model (Ben and Ken are in the negative outfit region) and two slightly underfit the model (Mike and Oskana are in the positive outfit region). This is an indication of how likely the response strings are for each person; Ben and Ken answered all but Question 9 correctly and Mike and Oskana answered all but Question 10 correctly. Because Question 9 was

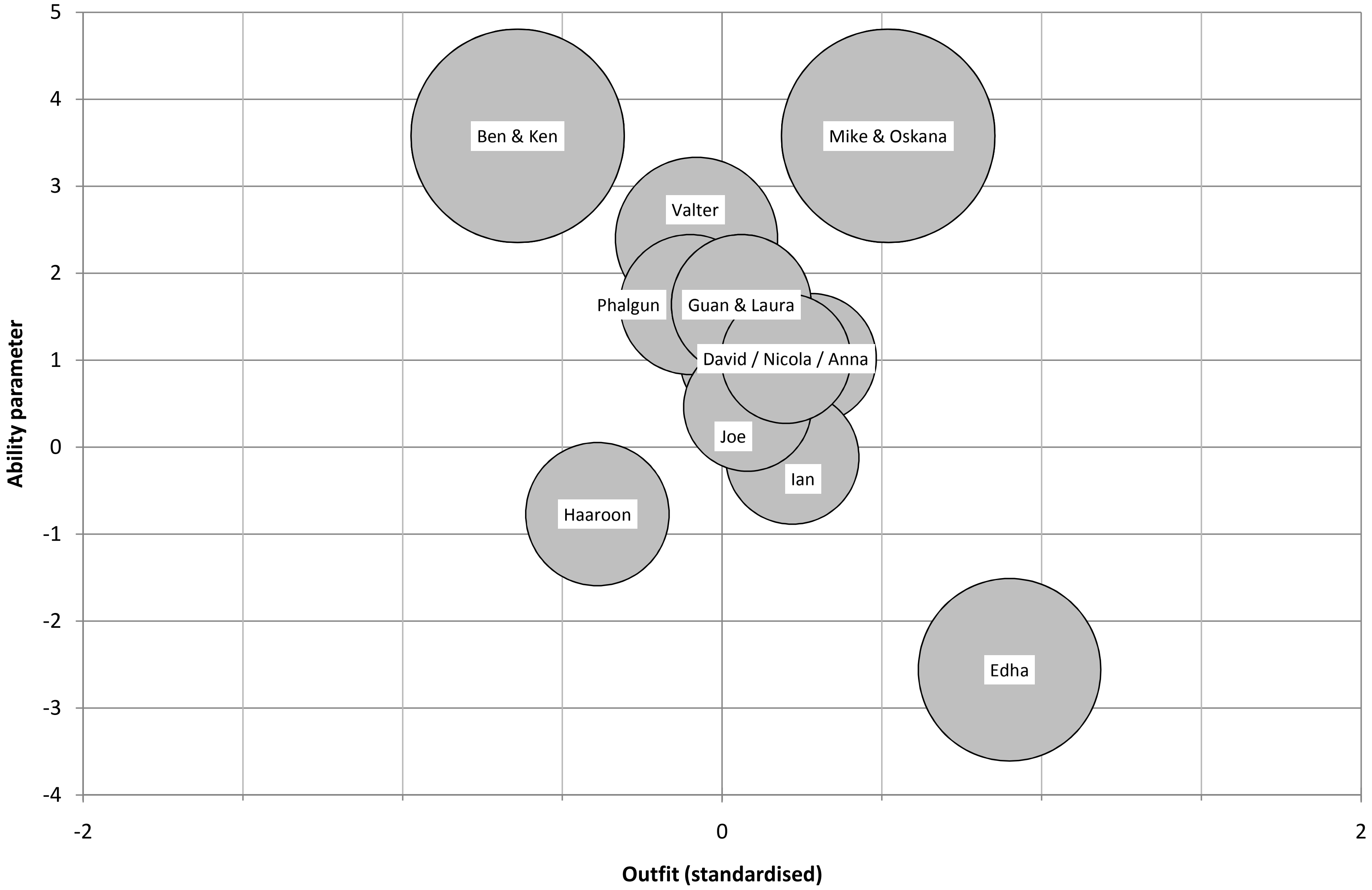


FIGURE 8.6: A Rasch pathway diagram for the interviewed students. Where two students are represented by the same bubble the '&' symbol separates them. Where two bubbles overlap too closely to label each individually, names are separated by '/' symbol(s) in the same order as the respective bubbles.

estimated to be more difficult than Question 10 (in terms of the difficulty parameters), Ben and Ken's string of answers fit the Rasch Model better, and so their outfit measure is less than Mike and Oskana's outfit measure.

At the bottom of the diagram we have Edha and Haaroon. Edha answered only 2/11 questions correctly ($D_{Edha} = -2.56$), but her two correct questions were 5 and 6. Other questions on the task which were estimated to be easier than Question 6 (such as Question 1 and 8) were answered incorrectly. The Rasch Model would expect a student scoring 2/11 to have answered these questions correctly rather than Question 6, and so Edha's standardised outfit to the model is relatively large at 0.9, although this is still well within the $(-2, +2)$ region which is deemed acceptable (Bond and Fox, 2007). Haroon answered four questions correctly, and these four were Question 1, 5, 7, and 8. These four questions were measured to be the easiest four questions on the task, and so Haaroon's standardised outfit is the negative value of -0.39. In this sense, the misfit measure (in this case, standardised outfit) is a measure of how characteristic a specific student's scores are in relation to the whole population (Edwards and Alcock, 2010b).

The rest of the interviewed students have ability parameter estimates between -0.5 and $+2.5$, and standardised outfit measures closer to 0. Comparing these students' outfit and infit measures, we see that both David and Phalgun have larger infit than outfit measures, which suggests that these students answered the easy questions correctly, the difficult questions incorrectly, but gave unexpected (or uncharacteristic) answers for questions with similar difficulty parameter estimates.

Recall that test suitability can be considered by counting how many questions have difficult parameter estimates within three logits of each person's ability parameter estimate. Doing this, we find that 9 questions were 'suitable' for Anna, David, Haaroon, Ian, and Nicola, 8 questions were suitable for Valter and Joe, 7 questions were suitable for Phalgun, Guan, and Laura, and 6 questions were suitable for Ben, Edha Ken, Oksana, and Mike (note that the questions may be different for each person). Thus the Rasch Analysis suggests that for all interviewed students at least half the questions were suitable, and for seven students at least eight of the questions were suitably targeted.

8.4.4 Comparing interviewed students with those completing the validation task

Now, differences between the students taking the task as part of the interview and in class are considered. This subsection first concentrates on the proportion of students answering each question correctly, noting that interviewed students generally answered fewer questions correctly. Then the ability parameter estimates of the interviewed students are considered, and it is noted that according to the Rasch Model, the difficulty of questions are targeted more towards the interviewed students. Finally, measures of fit are considered, and here also the Rasch Model suggests the task was more suitable for the interviewed students, in the sense that their response strings better fitted the model's expectation.

Comparing proportions of students answering each question correctly

Before we consider what a Rasch Analysis can tell us about how the interviewed students compared with all students, we will discuss Table 8.9, which recaps the proportion of students answering each question correctly in the validation exercise and main interview study. The table shows that the interviewed students performed worse than those who took part in the validation study; for most questions, the percentage of students answering correctly is lower amongst those from the interview cohort. An independent samples t-test was performed on each students' total score. On average students who took part in the validation task scored better ($M = 9.19$, $SE = .17$) than the interviewed students ($M = 7.40$, $SE = .62$). The difference was significant $t(160) = 3.18$, $p < 0.02$, and the effect size medium $r = .24$ (although only just: Cohen (1988) gives 0.24–0.36 as a medium effect size).

Below I suggest three hypotheses why there may be a difference in the two groups. Seminar tutors were asked to tell students to work with the task individually, and to collect sheets before going over the answers, so I will assume that these requests were followed. The first hypothesis to explain the observed difference is that the initial recruitment advertisement attracted those students who were struggling on the course. As discussed in the ethical issues section of the Main Study (Section 5.6), students were told that the interviews had the dual purpose of discussing how they found the

Table 8.9: Total correct answers given to each question, for the validation exercise (this Chapter, $n = 147$) and the main interview study (Chapter 5, $n = 15$).

	Question	Validation %	Interviews %
1.	A strictly increasing sequence	95.9	93.3
2.	An increasing sequence that is not strictly increasing	86.4	44.6
3.	A sequence that is both increasing and decreasing	81.4	12.9
4.	A sequence that is neither increasing nor decreasing	85.0	46.7
5.	A sequence that has no upper bound	98.6	100.0
6.	A sequence that has neither an upper bound nor a lower bound	84.4	93.3
7.	A bounded, monotonic sequence	89.1	80.0
8.	A sequence that tends to infinity	98.0	93.3
9.	A sequence that tends to infinity that is not increasing	46.9	13.3
10.	A sequence that tends to infinity that is not bounded below	78.9	40.0
11.	A strictly increasing sequence that does not tend to infinity	74.8	53.3

transition to university, and also to ask them some "maths questions related to their first year's study." Perhaps this wording encouraged students who were struggling with their course to attend, however when chatting to the students during the 'chat' period of the task (see Figure 5.3) none gave me the impression that this was the case. A second hypothesis to suggest such a difference is that the validation exercise took place one to three weeks after the interviews, giving students more time for the material to settle. A third hypothesis is that the 'think-aloud' protocol may discourage successful generation of examples, in keeping with Miller's (1956) famous claim that there are 7±2 items of working memory at any given time. Nisbett and Wilson (1977) argued that the verbal content of think-aloud task-based interviews is unreliable, so perhaps also there is a hit on a performance level too (Clark et al., 2006; Moreno, 2006). However various studies have suggested that think-aloud protocols make no difference in performance in well-designed interviews (e.g. Ericsson and Simon, 1980; Leow and Morgan-Short, 2004). In this context, well-designed refers to speech that is related to the thinking needed to complete a task, and there is no reason to suggest that such a condition was violated in these tasks.

A similar difference was observed between the quantitative first pilot study and the qualitative second pilot study. Table 8.10 recaps the first and second pilot studies' number of students, number of questions, mean student score and standard deviation. The table suggests that the difference in the performance of students taking the task

Table 8.10: A summary of average student scores in each study.

Study	No. students	No. questions	Mean score	St. dev
Pilot study 1	101	20	6.95 (34.7%)	3.98 (19.9%)
Pilot study 2	6	11	2.83 (25.8%)	2.23 (20.2%)
Main study	15	11	7.40 (67.3%)	2.30 (20.9%)
Validation study	147	11	9.19 (83.5%)	2.05 (18.6%)

during an interview and those taking the task during class might not be isolated to the main study.

Rasch Analysis' estimates of person ability

Now let us look at how a Rasch Analysis can complement this standard statistical analysis. Winsteps's output for the entire population is given in Appendix E. Before the interviewed students are compared with those who completed the task in class, the complete dataset (including both groups of students) will first be described. Looking at the entire population of persons taking the task, represented as a histogram in Figure 8.7, 28% of students provided a correct example for every question, and so scored perfectly. These students were not included by Winsteps when estimating the other person and item parameters; they were re-introduced by the program afterwards with a small value subtracted from their total score, and so the ability estimates for such students ($A_p = 5.21$) should be taken with greater caution (such estimates have a large standard error, in this case 1.98).

Almost all students not scoring perfectly have ability estimates between -1 and $+4$ (97%). By adding and subtracting 3 logits to each individual's ability estimate, we can conclude that 71 students have a suitable ability parameter estimate for at least 8 items on the task (recall that $|A_p - D_i| < 3 \implies P_{p,i} \in (0.05, 0.95)$). So, according to the model the task was correctly targeted for around half of the students, which is probably a result of the 46 students who answered every question correctly. Recall from the last section that when given in interview, the task was correctly targeted for 7/11 students.

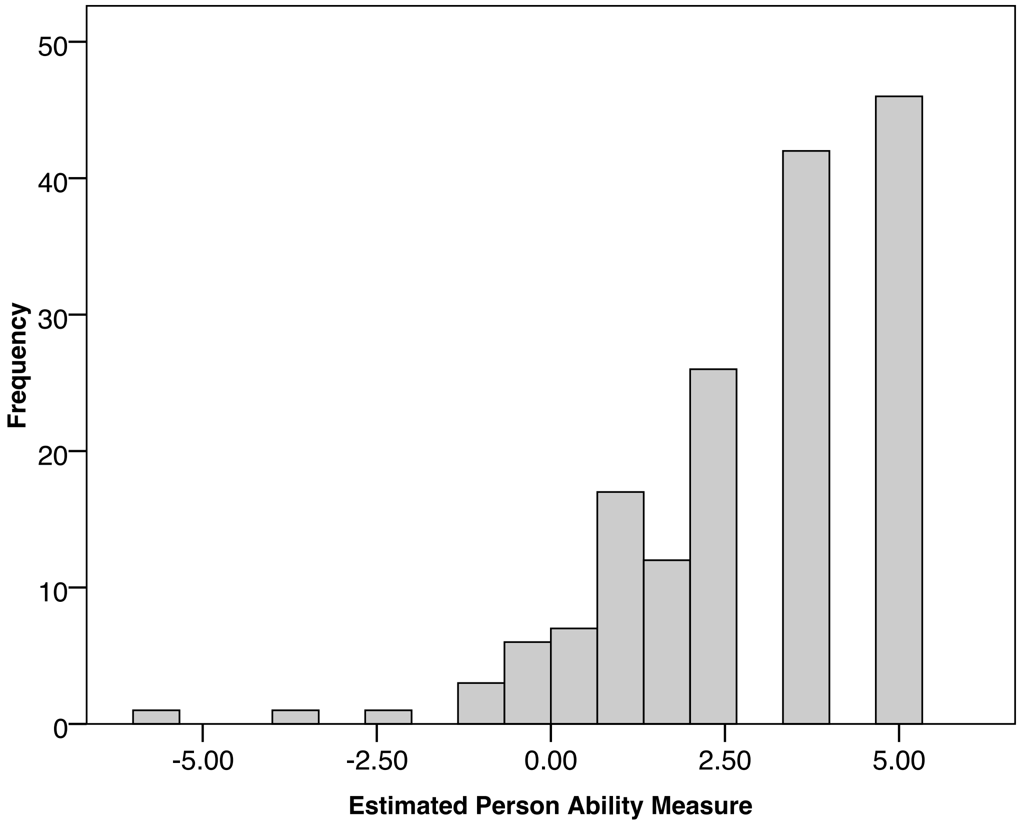


FIGURE 8.7: Histogram of estimated person ability (for the dataset with interviewed students included).

Pathway diagram and measures of fit

A Rasch pathway diagram for all students is given in Figure 8.8. On this graph, the bubbles which represent interviewed students are shaded. Bearing in mind that the bubble centred at $(5.21, 0)$ represents 47 people, the graph confirms the t-test indicating that the interviewed students performed significantly worse than those taking the validation task. The reasons for using a Rasch Analysis is not just to compare person abilities however, it is also to compare the nature of the response strings, as measured by the outfit values. When we compare the outfit values of the interviewed students with those scoring similarly in the validation task (by looking at the horizontal placement of the bubbles) we can see that the interviewed students' standardised outfits are between -1 and $+1$, but the outfit values of some of the students who took the task in class are larger than this. This may mean that the interviewed students' response strings fit the Rasch Model better than the overall populations' response strings, but it could also be because there were far fewer students who were interviewed; a further statistical test is needed to see if the difference is significant (see below).

There is also the possibility that between the two groups there is a different relation for infit compared to outfit. Recall that outfit emphasises discrepancies between predicted and observed scores for items with difficulty parameters that are far from a person's ability estimate, and infit emphasises discrepancies for items that have similar difficulty parameters to a person's ability estimates. For instance, if it is the case that asking an interviewee to articulate their thinking induces them to make errors on questions that the student might otherwise answer correctly, one would expect this effect to be larger for items with similar difficulty parameters to that person's ability parameter estimate. If this is the case we would find that interviewed students' infit measures are higher than non-interviewed students, relative to their respective outfit measures. The opposite scenario is also possible, that asking students to think aloud makes them focus more on the questions, although the significant result from the t-test on students' total scores suggests that this latter possibility had not been the case.

To address these concerns a one-way between-groups multivariate analysis of variance (MANOVA) was conducted, with the null hypothesis that there was no difference in the relationship between the dependent variables 'outfit' and 'infit' for the groups 'interviewed students' and 'non-interviewed students'. The MANOVA failed to reject the null-hypothesis, $F(2, 159) = 0.67$, $p < .516$ (Each of Pilli's trace $V = 0.008$, Wilks's statistic $\Lambda = 0.992$, Hotelling's trace $T = 0.008$ and Roy's largest root $\Theta = 0.008$ also failed to reject the null-hypothesis). This is not the same as claiming there is no difference between the groups[3], but it is evidence to support a claim that there is no significant difference in the outfit and infit measures for interviewed and non-interviewed students in this study. This indicates that the response strings of two students with the same total score are likely to be similar, even when one student completed the task in an interview and one student did not.

[3]For instance, it may be the case that both of the hypothetical scenarios described in the previous paragraph were true for different students and so there was no statistical effect for the entire group considered as a whole

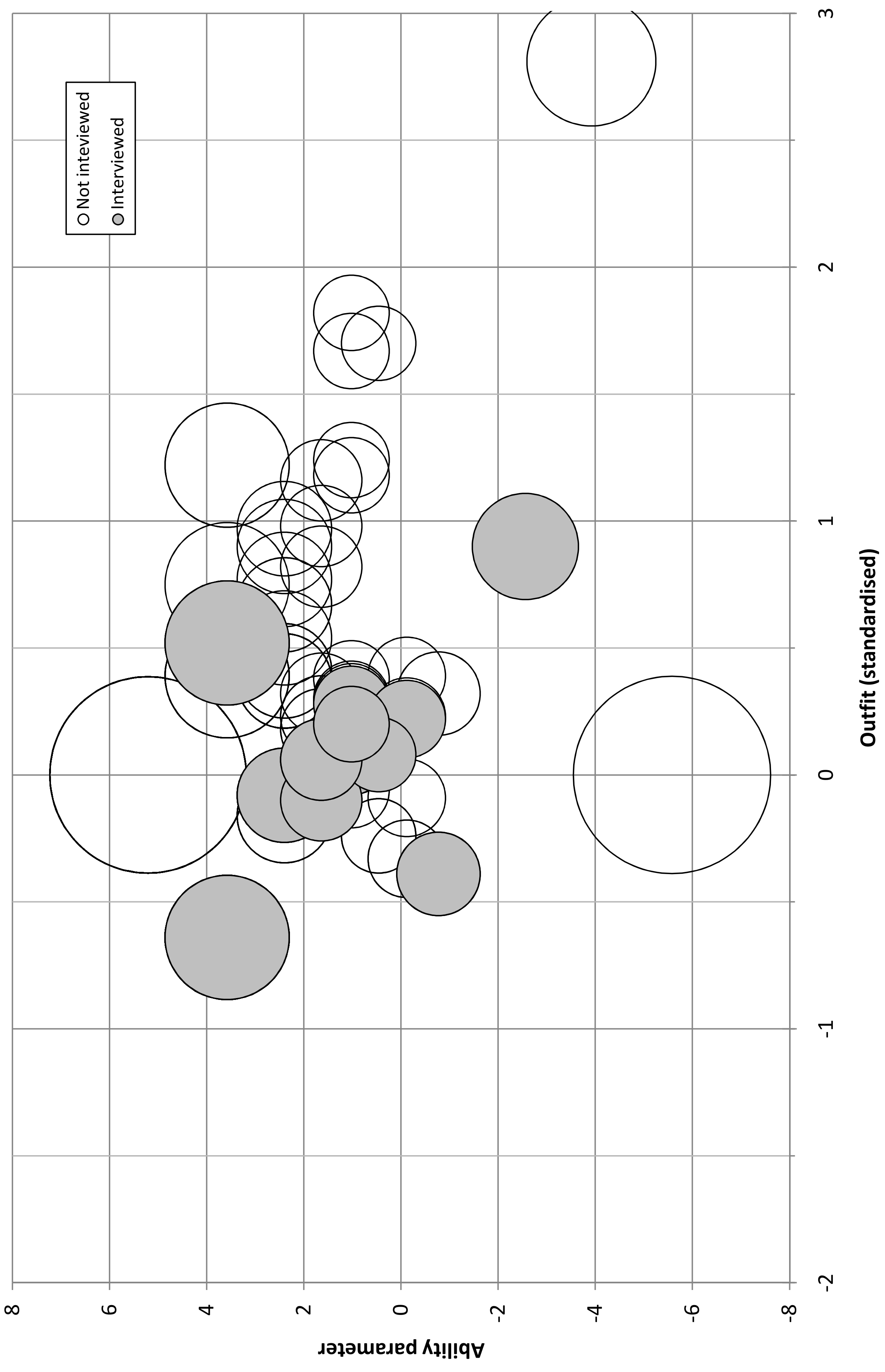


FIGURE 8.8: A Rasch pathway diagram for all students, with interviewed students indicated.

8.5 Summary of chapter

In order to compare the students interviewed in the main study with the wider population, the example generation task was given to the rest of the year group approximately three weeks later during problem classes. After removing those students who completed the task in class having already been interviewed, there were 147 students in the dataset. The statistical procedure of Rasch Analysis was used to analyse this dataset and the dataset with interviewed students included.

Once the data was shown to be suitable for Rasch Analysis, the computer package Winsteps (Linacre, 2009) was used to estimate the Rasch Model parameters for the data, and measure these parameters' fit to the model. First the parameters were used to objectively rank the questions in order of their difficulty, and second to make comparisons between the interviewed students and the general population. Then, fit statistics were used to determine how characteristic the interviewed students response strings were, relative to the general population.

It was shown that students who took the task during an interview performed worse compared to those that did not (on average answering 1.79 fewer questions correctly), and some hypotheses were suggested to account for this. When looking at the estimated parameters' fit to the Rasch Model, both interviewed and non-interviewed students response strings fitted the model well. Moreover, when comparing interviewed and non-interviewed students with similar ability estimates, there was no statistical difference between the response strings. These two results combine to indicate that, in terms of the Rasch Model, although the students who took the task during an interview scored lower, their responses were not uncharacteristic for students from the population who scored similarly.

Chapter 9

Discussion and Conclusions

This thesis has contributed to two growing areas of research: that which looks at students' understanding of the undergraduate module Analysis, and that which explores students' awareness in the example generating process. As well as advancing what is known in both these areas in isolation, the thesis has shown that an example generation task can be successfully used to explore students' understanding of sequences and basic sequence properties in Analysis. The thesis is also the first account to use an example generation task as the basis for a phenomenographic analysis of a topic.

This discussion and conclusions chapter takes as its starting point the research questions initially stated in the Introduction (Section 1.1), and restated here:

1. How successful are students at accurately generating examples of sequences satisfying certain combinations of properties?
2. What is the qualitative variation in students' experiences of sequence generation?

The first two sections of this chapter focus on the research questions in reverse order. Section 9.1 addresses the second research question by revisiting Chapter 6's dimensions of variation. Then, in Section 9.2, the first research question is addressed by considering all the studies in the thesis and relating these to the dimensions of variation where appropriate.

Section 9.3 then shifts the focus away from the research questions and considers the methodology of the main study, examining the extent to which this phenomenographic study of an example generation task has provided useful information about the generated concept outside of such tasks. Pedagogical implications are discussed in Section 9.4, and then Section 9.5 suggests possible future research stemming from all of these areas.

9.1 What is the qualitative variation in students' experiences of sequence generation?

This research question was addressed in the analysis of the main study, presented in Chapter 6. Four dimensions of variation emerged from the analysis of the example generation interview data, with each dimension containing three or four categories of description, ordered in terms of increasing sophistication. Recall that when analysing the data, the dimensions emerged by taking students' comments at face-value, bracketing where appropriate both the truth of the statement relative to formal mathematics, and the other dimensions.

Table 9.1 restates the categories of description for each dimension of variation that was previously presented as Table 6.5.

The first dimension of variation, Using Definitions, described qualitative differences in students' awareness of definitions. In this context, the definitions were those of sequence properties (rather than the definition of a sequence itself), and students' awareness refers to how students interacted with definitions during the task, to what extent they explicitly made reference to them, and whether parts of the definitions were used in the example generation process. Analysis of data collected in the main study suggested that, when students did not use definitions they instead relied on spontaneous conceptions (usually based on everyday use of words).

The second dimension of variation, Representation of Sequences, considered the different ways students wrote down their sequences, and the reasons the students gave for this choice. Typically students presented their sequences as a list of numbers or as a formula (indeed the definition sheet given to students suggested they do this), but this dimension of variation also captured behaviour where students wrote down as answers objects that were not sequences. Analysis of interview transcripts suggested that such students typically still regarded these objects as sequences. This dimension of variation also

TABLE 9.1: Summary of the different dimensions of variation outlined in Chapter 6.

Using Definitions (Section 6.2)	
Def-A.	*Unaware of Definitions*
Def-B.	*Refers to Definitions*
Def-C.	*Uses Definitions*
Def-D.	*Manipulates Definitions*

Representation of Sequences (Section 6.3)	
Rep-A.	*Any Representation is Suitable*
Rep-B.	*One Representation is Superior*
Rep-C.	*Any Well-Defined Representation is Suitable*

Sequence Construction Strategies (Section 6.4)	
Con-A	*Generic Initial Approaches*
Con-B.	*Trial and Error*
Con-C.	*Transformation*
Con-D.	*Analysis*

Justifications (Section 6.5)	
Jus-A.	*No Justification Attempted*
Jus-B.	*Appeals to an External Authority*
Jus-C.	*Informal Justifications*
Jus-D.	*Formal Justifications*

highlighted some students' belief that certain types of representations (e.g. a formula) were 'better' than others, which sometimes led to students being able to generate a sequence, but not feeling comfortable with writing it down.

The third dimension of variation, Sequence Construction Strategies, focused on the strategies students used when generating their examples. Various informal approaches were included as 'strategies' when exploring the extent of this dimension, such as repeating the question aloud, writing down relevant definitions and sketching a graph. Classed as more sophisticated were times when students followed the trial and error, transformation and analysis strategies previously identified in Antonini's (2006) research exploring how experts generated examples. This thesis contributes further to what is known about how undergraduates use these strategies, in particular transformation and analysis that contain flawed reasoning (see Section 6.4 and Edwards and Alcock, 2010a).

The final dimension of variation, Justifications, explored the different ways students justified their final answers and the chains of reasoning given as they followed the think-aloud protocol. The types of behaviour included in this dimension ranged from students who provided no justification for their answers, to those whose reasoning was based on arguments framed in formal mathematics.

The inter-coder validation exercise in Chapter 7 suggested that these dimensions of variations are generalisable to other groups of students taking the main study's task. Moreover, I speculate later in this discussion and conclusion chapter that the qualitative variation in students' experiences of sequence generation, as contained within these dimensions of variation, will also be present in most example generation activities at undergraduate level.

It is important to bear in mind that, even in the context of answering this research question, the dimensions of variation presented in Chapter 6 are not a classification system whereby each student can be labelled a 'Def-w|Rep-x|Con-y|Jus-z student'. In the course of the main study's interviews, there were cases where a student was labelled by different categories of a dimension of variation, depending on the portion of the interview considered (for instance, incidents from Valter's interview can be found in Def-A, Def-C and Def-D). The dimensions of variation instead help us reflect upon the different possible ways students can be aware of the different aspects of the example generation of sequences, and so they have addressed the second research question.

9.2 How successful are students at accurately generating examples of sequences satisfying certain combinations of properties?

This section considers the extent to which the thesis has addressed the first research question. To a certain extent, each study in the thesis has contributed towards answering this question, and so this section draws on aspects from each study, including relevant parts of the dimensions of variation. In order to justify that the thesis has answered the question and thus contributed to what is known about how successful students are at generating examples of sequences, the following areas will be discussed: the number of students giving correct answers in the studies, the types of examples students gave, how students have gone about example generation, and what the thesis has told us about students' concept images, example spaces, and spontaneous conceptions associated with sequences and sequence properties.

9.2.1 The number of students providing correct answers

This subsection addresses the first research question quantitatively. It collects the mean score of the students from each study, and suggests that the answer to the research question depends on both the cohort of students and the way in which the example generation task is delivered. It then considers a specific pair of questions: Question 1 and Question 3 as an illustration of a quantitative result that can be found across all studies.

Table 9.2 repeats Chapter 8's summary of the mean student scores from the various studies in the thesis. In the description that follows this table, the first research question is answered with respect to each individual study. Some comparisons are drawn between the mean percentage scores from the studies, but there are no in-depth comparisons. This is because differences between the studies make such a comparison not appropriate. For instance the first pilot study included more questions than the other studies, and the second pilot study's cohort was different from the main study's cohort. This is not a weakness in the design — each study's aim was not the same — and so it is unsurprising that results are not directly comparable.

TABLE 9.2: A summary of average student scores in each study. The dotted lines separate studies where direct comparisons cannot be made.

Study	No. students	No. questions	Mean score	SD
Pilot study 1	101	20	6.95 (35%)	3.98
Pilot study 2	6	11	2.83 (26%)	2.23
Main study	15	11	7.40 (67%)	2.30
Validation study	147	11	9.19 (84%)	2.05

In the first pilot study 101 students sat a 20-question example generation task during a lecture. All but one student had studied a module on Sequences and Series the previous year, and so it was surprising that the mean student score was low (35%). An indication of this study was that students were struggling with the content of the task, rather than the task format. For the second pilot study, a similar task (but with nine fewer questions) was given to six students who were currently attending the Sequence and

Series module. In this pilot study the percentage of correct answers was even lower than in the first pilot study (26%).

There was a higher mean percentage of correct answers in the main and validation studies. These used the same question set as the second pilot study, yet there was a marked increase in total marks. This difference may be a result of the higher entry requirements required by the second institution, or a reflection that the second institution placed more emphasis on formal definitions earlier in the degree. Subsection 8.4.4 reported that an independent samples t-test indicated a significant difference between the students' scores in the main study and the scores in the validation study, with the interviewed students scoring less well (note Table 9.2 indicates this was also the case for the pilot study). This subsection also discussed hypotheses for this discrepancy. In terms of the mean score of students, therefore, the answer to the first research question is that the success of students answering the type of example generation exercises varies depending on the cohort and situation.

Across each study, it was noted that there were many students who gave a correct example for Question 1 (a strictly increasing sequence), but who did not for Question 3 (a sequence that is both increasing and decreasing). In the first pilot study 97% of students answered Question 1 correctly but only 13% answered Question 3 correctly. For the second pilot study these percentages were 67% and 0%, in the main study they were 93% and 60%, and in the validation study 96% and 81%. To find this result across each study suggests that, in terms of answering the first research question qualitatively, students are less successful when combining multiple definitions. The Rasch Analysis of the validation study also indicated that questions that combine multiple definitions are more difficult (see subsection 8.4.2).

9.2.2 The types of examples given by students

This subsection considers the different types of examples students gave when answering the tasks' questions. The discussion here has a clear link to such students' evoked example spaces (i.e. all examples to which a student has access at a particular time, including both mathematical and non-mathematical representations). This thesis has contributed to what is known about students' example spaces by identifying a reoccurring phenomenon where students treat classes of objects which are not sequences as sequences in this type of task.

Each question on the task(s) asked students to give an example of a sequence satisfying particular properties, and indeed each question contained the word sequence. In order to be successful in any question on the various tasks therefore, a student needed to present their answer in a form that was a valid representation of a sequence, i.e. a mathematical function $f : \mathbb{N} \to \mathbb{R}$, or an object that could be interpreted as such (a more thorough description of what is meant by this was presented in Section 4.1.4).

It was noted in the first pilot study that some students presented their answers in forms which were not sequences (this was presented as a conference paper in Edwards and Alcock, 2008). In this thesis such answers have been classified and referred to as incorrect non-sequences (**INS**). These answers typically retained some of the constraints asked for by the question, such as a student giving the unbounded interval $(-\infty, \infty)$ when the question asked for an example of an unbounded sequence. Such answers were also seen in the second pilot study, where it was noted that some students tended to give **INS** to several questions, whilst some questions prompted **INS** responses from several students. It was also noted in this second pilot study that one of the students who gave many **INS** reported the belief that any sequence could be written as both a formula and as an interval of the real line.

In the main study, answers that were **INS** and statements related to these answers were included within the dimension of variation Representation of Sequences (see Section 6.3). Category of description *Rep-A. Any Representation is Suitable* (Section 6.3.1) contained incidents where students believed there was no restriction on the way their answer could be represented, including the **INS**-type answers. Examples given here included students who gave a continuous function whose domain was finite, and those that gave a double-sided sequence (i.e. a function $f : \mathbb{Z} \to \mathbb{R}$). When such students discussed these answers with the interviewer, it was unclear whether they believed these answers to be sequences (as was clear for the student from the second pilot study), but the students did report that they believed the examples to satisfy the properties requested in the question. Although the validation study did not focus on the nature of incorrect answers, a browse of Appendix C shows that some students here too gave several **INS** answers (e.g. student WB09).

The above discussion has highlighted that students across each study provided answers that were not of the requested object type. Given that the studies' cohorts span two universities, and at least two year groups in the case of the pilot studies, these results strongly suggest that such answers may be a common feature to example generation exercises, or at least ones where the object type is fixed and the properties requested by the questions are loosely applicable to other object types.

When a student answered a question correctly, the example given was typically written as a list of numbers or as a formula. When a student reflected on the source of their examples (usually when they were stuck and searching for inspiration), it was sometimes remarked that they had seen such an example in class or in a lecture. Instances where a student appeared to construct their example during the interview are discussed in the next subsection.

9.2.3 The process of example generation

In relation to the first research question this thesis has also contributed to what is known about how undergraduate students go about the process of example generation.

The two pilot studies did not explicitly consider what students did, or what students reported they did, as they were constructing examples. The main study and its analysis, however, did consider what the example generation task suggests about these areas. As outlined in Section 6.4, the data fitted very well with Antonini's (2006) three categories of example generation (Trial and error, Transformation and Analysis). This implies that Antonini's categories apply more widely (to undergraduates as well as postgraduates), and with tasks that ask a set of questions all on the same topic (rather than from different aspects of mathematics). As well as contributing to what is known in this regard, this thesis has also highlighted that when undergraduate students use a transformation or analysis strategy, sometimes the logical deductions which underpin these strategies are false relative to formal mathematics.

A further addition to what was previously known about students' example generation strategies is a type of behaviour not noted by Antonini (2006) or subsequent studies using Antonini's strategy as their basis. In the least sophisticated category of the Sequence Construction Strategies dimension of variation, students would perform rituals and routines that were seemingly unrelated to the question asked. Such routines were not always undesirable, for instance writing down any definitions in the questions is probably a very good idea, but others such as repeating the question over and over again waiting for inspiration were seen as 'strategies' less sophisticated than even trial and error.

9.2.4 Concept images and spontaneous conceptions

The various studies presented in this thesis have also contributed to what is known about students' concept images in relation to sequences and sequence properties. The previous discussion about students' example spaces is relevant here, provided one considers the example space to be a subset of the concept image (as was suggested in Section 2.2.2), as it suggests which objects students treat as part of their concept image for sequences.

Another aspect of the concept image that the studies in this thesis can shed light upon is that related to the definitions of sequence properties. In each task, students were given lists of sequence definitions, but the ways students dealt with definitions varied (see dimension of variation Using Definitions in Section 6.2). In the studies where the task was given to students in class (the first pilot study and the validation study), little can be deduced about students' concept images as data consists solely of the completed task sheets. In the other, interview-based, studies, only the main study focused (in part) on the ways in which students interacted with the sequence definitions. Therefore what this thesis has told us about students' concept images and spontaneous conceptions comes primarily from the dimension of variation Using Definitions (see Section 6.2).

In this dimension, it was noted that many students did not make reference to formal definitions, instead relying on what they thought the definitions meant (in Tall and Vinner's (1981) terminology such students relied on reasoning from their concept images rather than working with formal concept definitions). Students who did this typically interpreted the everyday meaning of a definition, such as "essentially going up" for an increasing sequence. While spontaneous conceptions such as these were successful in answering some questions, students who relied on this type of reasoning typically did not provide correct answers for questions that combined properties: "essentially going up and down" is a less helpful concept image for a sequence that is both increasing and decreasing. Although when Cornu (1991) described spontaneous conceptions they could be drawn from both everyday language and previous mathematical experience, most of the identifiable spontaneous conceptions from students came from the everyday meaning of words.

Section 6.1 hypothesised that there were essentially three different types of formal sequence property definitions in the task: those that compared sequence elements pairwise term-by-term, those that gave a rule for every sequence element to follow, and those that controlled the behaviour of the sequence in the long-term. It was noted in subection 8.4.2 that, according to the estimates of difficulty parameters in the Rasch Analysis, the most difficult questions (in terms of number of students answering incorrectly) were those that combined the definition types, for instance asking a student to give a sequence that satisfied both term-by-term (e.g. increasing) and a long-term (e.g. tending to infinity) properties. This finding, combined with the main study's observation (in subsection 6.2.1) that many students who struggled with the questions combining sequence types were working with everyday concept definitions rather than formal ones, provides more backing for the claim that a question is more difficult because it requires reasoning using formal definitions rather than common spontaneous conceptions present in concept images.

In terms of the first research question, this thesis has suggested that when a student struggles with the type of example generation activity given in the various studies, it may be because they reason primarily with their concept images rather than formal concept definitions. As noted in the literature review, most studies considering undergraduate students' concept images of formal definitions have focused on definitions such as limiting behaviour and continuity of sequences, and so this thesis has furthered what is known by demonstrating results that are consistent with earlier observations, but for simpler definitions that are met by students earlier on in their courses (a similar conclusion was made by Alcock and Simpson (2009b), but via a classification task rather than by example generation).

9.3 Using example generation as a phenomenographic research tool

As well as addressing the two research questions, as discussed above, this thesis has been the first study to use the combination of an example generation task with a phenomenographic analysis to draw conclusions about students' understanding of a concept. Various authors have suggested that studying students generating examples is a good research tool (e.g. Bills et al., 2006; Zazkis and Leikin, 2007), but few have speculated on the best way to do this beyond Zazkis and Leikin's (2007) suggestion that researchers explore the correctness of the examples, how they are produced, how they vary, and how general or specific they are.

In phenomenographic studies to date, typically a concept or object is introduced to students, and then the students are asked to reflect on that topic through a series of open interview questions. For example Marton and Säljö's (1976) distinction between deep/surface approaches to learning was the result of studying students' interaction with a reading task, and the same students' reflections on that task. Renström et al.'s (1990) exploration of students' conception of matter partly involved asking students what they thought matter was. In this thesis's main study, rather than asking students what they thought a sequence was, students were instead provided with a sheet of formal definitions and asked to complete a structured example generation task, thinking aloud as they went along, then afterwards discussing the task with the interviewer. What follows here is some thoughts on the extent to which this thesis can be considered a phenomenographic study.

By basing the student interviews around the context of an example generation task, the main study in this thesis has introduced students to the topic in question, and encouraged the students to get involved with the topic in a way that would be difficult, if not impossible, via a series of open questions. In this sense it is similar to the type of studies above. However, unlike these studies, the focus of discussions with the students has been more narrow than it would have been had the interviews been focused around the question "what is a sequence?" On reflection, I believe that if the students had been asked this question during the interview, it would either have had an influence on the type of behaviour seen during the task (if students were asked before the task), or implied to students that some of their answers were not sequences (if they were asked after the task), thus interrupting the flow of conversation of the discussion period and creating ethical difficulties. In hindsight, however, I would have liked to ask this question at the end of the task phase of each interview, although the content of the dimension of variation Representation of Sequences has suggested the type of variation that might exist in response to such an open question.

Despite these differences in the focus of dialogue between student and interviewer in this study and traditional phenomenographic studies, I believe in terms of the analysis of the interview data the methodology is clearly in keeping with phenomenography. The analysis successfully generated a rich set of dimensions of variation, and these dimensions reflect the qualitatively different ways of experiencing sequence generation observed in the interviews. Moreover, the inter-coder validation exercise in Chapter 7 has suggested that the dimensions of variation are both stable and applicable to new datasets.

9.4 Pedagogical implications

The pedagogical implications of the thesis are discussed in two areas. First the section considers whether the example generation task could be useful for students to complete in a different context, and second the section considers how the dimensions of variation Using Definitions and Representation of Sequences might be useful pedagogical content knowledge for a lecturer.

Example generation tasks

Example generation tasks have been suggested by some authors as a good way to introduce new topics to students and to help students explore links in topics they are already familiar with (Dahlberg and Housman, 1997; Hazzan and Zazkis, 1999; Watson and Mason, 2002, 2005). It is tough to argue against Meehan's (2002) claim that students with access to a wider range of examples will have more developed example spaces, and hence authors have suggested it might be beneficial to engage students with more varied collections of examples (Alcock and Simpson, 2009b; Selden and Selden, 1998). I would go even further and propose that students with a rich and varied (evoked) example space are more likely to have concept images that are aligned with formal theory. Perhaps completing this thesis's example generation task in a more pedagogically-focused setting may help students unearth a wider variety of examples with which to work in the future?

However, when the task was given to students in interview conditions the (correct) answers they gave typically did not vary greatly; it was rare for a student to try to provide different or interesting examples across or within questions (a notable exception was Phalgun: see Incident Con-C1, page 168). Watson and Mason (2005) note that by asking students to provide more than one example, or to generate examples that the students believe few others would have thought of, a researcher or teacher can unearth a richer variation of responses. However, in the pilot study interviews, students were struggling to provide any example at all, and probing for different or unusual examples did not seem a positive step forward (and so such requests were not continued in the main study). The conclusions discussed above in response to the first research question were, in part, that some students find example generation difficult, and that they may encourage reasoning based on concept images rather than formal concept definitions. This tends to suggest that, unless students are carefully supported, the use of this type of example generation task as a pedagogical tool may not help those students who are struggling with the role and content of definitions in mathematics.

After most interviews, I asked the student whether they had ever completed a task such as the one they had just finished (none had). Most students seemed receptive to the idea that trying to find examples which satisfy combinations of properties can mean one learns a lot about these properties. Although working with definitions in mathematics relies on understanding what definitions mean and how they work, authors have suggested that being exposed to, and classifying, different collections of mathematical examples can help students understand the definitions when they are formally presented (Alcock and Simpson, 2009b). It is not clear if the example generation task used in the studies reported by this thesis might also be useful for students to complete prior to, during or after the teaching of formal definitions in Analysis.

Dimensions of variation

In terms of the dimensions of variation presented by the thesis, I believe it would be useful for lecturers in mathematics to be aware of the Representation of Sequences dimension. Anecdotally, it is common for lecturers to introduce a new mathematical object to students by either first giving some examples of the object, and deducing or imposing properties that the object must have more generally, or by initially stating some combination of axioms and definitions and then providing examples of the object. However, this thesis has suggested that when some students are asked to provide objects that satisfy combinations of constraints, some do not keep the object invariant. This means that the relationship between definitions and the objects constrained by definitions is a grey area for some students, and so either way of introducing new topics may be confusing. I personally find myself thinking of the different categories of description on this dimension when teaching students about how mathematical objects are constrained by axioms and definitions in mathematics.

Similarly, the dimension of variation Using Definitions also contains incidents that lecturers of mathematics may find informative. The types of deductions inferred by students who make no reference to definitions may be surprising to an experienced mathematician. For instance the absence of a conflict between Joe's spontaneous conception that a constant sequence was neither increasing nor decreasing and what he remembered from lecture about such a sequence being increasing and decreasing (see Incident Jus-B1, page 179).

Implications for my own teaching

The research carried out in this thesis has informed my own teaching. When I teach students about definitions such as *increasing*, I try to emphasise not only the definition's role in clarifying which objects should be classed as increasing, but also that when mathematicians use the term "increasing" they mean only (and precisely) the mathematical definition, forgoing any real-life spontaneous conceptions, even if that means some objects that intrinsically do not seem to be increasing are now classified as such, or vice-versa. It is not straightforward to move from using the everyday meaning of words to using the content of formal definitions to define objects, and I enjoy helping students through this transition.

The thesis has also helped me become more aware of representation-type errors students might make. This is the case not just in Real Analysis, but in all areas of mathematics that I teach and support. For instance, a student studying a basic statistics course recently wrote the following statement:

$$P(A) + P(B) = C$$

where A, B, and C are sets, and $P(X)$ is the probability that any element in the set X occurs. Although this is unconnected with real sequences, it is a typical representation-type error. Rather than just say that the right hand side of the equality needs to also be a probability, we discussed what each object in the equation represents, and the student noticed their own error.

9.5 Implications for research

In this chapter, there have been various ideas which could form the basis of future research. Many of these are briefly recapitulated and discussed here.

It would be very useful to make a systemic comparison of the different ways in which undergraduate students are exposed to definitions. This study has explored the variation between simple definitions involved with sequences, and speculated that there are three types (term-by-term, sequence-wide, and long-term), and that some of the difficulties faced by students are a result of students treating these definition types inappropriately (possibly because of spontaneous conception resulting from the everyday use of words). Perhaps this schema could be extended to include definitions in Analysis and more generally in undergraduate mathematics? For instance, the definition of a sequence converging to a finite limit involves an $\epsilon - \delta$ combination that constrains the long-term behaviour of the sequence, but the way one interacts with the definition may be different when ϵ is treated as a fixed number, or as an arbitrary (but particular) constant. Might such a classification be related to ways one can prove or find counter-examples for statements involving these definitions?

On a similar theme, there is often a discrepancy between similar-looking (or indeed similar-sounding) types of definition, and the way these definitions should be treated in mathematics. For instance "increasing" and "tending to infinity" may both conjure up mental images associated with going up. In a similar way to Kirshner and Awtry's (2004) exposition of mal-rules in algebra based on the visual salience of algebraic transformations (e.g. that rules like $(ab)^2 = a^2b^2$ are fine, but the similar looking

$(a+b)^2 \neq a^2+b^2$ is not), perhaps a similar study can be made between definitions which look (or sound) like they are term-by-term, sequence-wide and long-term. A study that systematically explores spontaneous conceptions related to the everyday meaning of definitions in mathematics would be very useful for both those teaching and studying undergraduate mathematics.

With regards to the object type of answers, the first pilot study suggested that for some students, a variety of objects are considered as sequences (including intervals of the real line, series, and functions defined on all real numbers). The second pilot and main study provided evidence that such answers were not the result of a student trying to get partial credit on their answer; in the interview for that moment the student treated the **INS** response as a sequence. Does this mean that students would classify such objects as sequences before such an example generation task? What about immediately after an example generation task? What if they are first asked to reflect upon the question "what is a sequence?"

From an example generation perspective, there is a question as to the relationship between those students that are successful at generating examples and those that are successful by other measures (e.g. examination results). Authors such as Dahlberg and Housman (1997) suggested that those students who spontaneously generated examples were more successful, but the direction of causation is unknown. Given that so many researchers believe asking students to generate examples is a good pedagogical strategy, it is surprising that there only seems to be empirical grounding for this belief (a recent study by Iannone et al. (in press) attempted to examine links between students asked to generate examples and success in proof production, but they found no clear link).

With respect to the dimensions of variation presented in Chapter 6, is there a relationship between being successful at example generation of sequences, and being predominantly 'located' on sophisticated categories of description within the dimensions? Other phenomenographic studies have devised questionnaires based on the dimensions in which to categorise students ('this answer is typical of a Def-C student'), and then explored the types of answer they give to questions (e.g. Crawford et al., 1998). In particular, is it the case that students who use the more sophisticated sequence construction strategies have a better grasp of definitions, and do such students produce more correct examples?

For this research, a researcher would first need to design a task or survey that would target the different categories of description within dimensions.

Finally, it would be interesting to conduct further exploratory studies that would consider the extent to which the dimensions of variation from the main study data are invariant across similar tasks in other areas of mathematics. Would an example generation involving an object in linear algebra (bases perhaps) result in similar categories of description? If so would there be more incidents for certain categories?

9.6 Final remarks

The purpose of this thesis has been to give a phenomenographic account of how students completed an example generation task focused on sequences. The work reported here has contributed to what is known about how students deal with sequences, formal definitions of sequences properties, and also how students interact with example generation tasks. It has contributed to the wider body of research on Real Analysis, in particular examining how students interact with simple definitions in the context of an example generation task. Research such as this is important if we as a community of mathematics education researchers are to understand how undergraduates think about, and understand, the mathematics they meet during their degrees.

Appendix A

Ethical Documents From the Main Study

A.1 Participant information sheet

PARTICIPANT INFORMATION SHEET
(to be read before Informed Consent Form is completed)

Purpose of study

- To investigate the transition from school-level mathematics to university study from a student's perspective
- To explore the types of answer given to typical problems in introductory analysis

You are not expected to prepare for the study in any way; indeed it is preferable if you do not prepare at all.

The interview will be recorded (to help maintain an accurate record for when I review the interviews), unless you request for this not to happen. The recordings will be treated in strict confidence and any reference to them shall hide the identity of the student in question.

All data obtained from this interview will be destroyed in 6 years, or sooner if the participant wishes.

If you have any further questions about the study please ask!

Now please read and sign the Informed Consent Form.

FIGURE A.1: The participant information sheet presented to students taking part in the main study (based on Loughborough University's 2008 template, the latest version of which may be found at `http://www.lboro.ac.uk/admin/committees/ethical/`).

A.2 Informed consent form

INFORMED CONSENT FORM
(to be completed after Participant Information Sheet has been read)

The purpose and details of this study have been explained to me. I understand that this study is designed to further scientific knowledge and that all procedures have been approved by the Loughborough University Ethical Advisory Committee.

I have read and understood the information sheet and this consent form.

I have had an opportunity to ask questions about my participation.

I understand that I am under no obligation to take part in the study.

I understand that I have the right to withdraw from this study at any stage for any reason, and that I will not be required to explain my reasons for withdrawing.

I understand that all the information I provide will be treated in strict confidence.

I agree to participate in this study.

Your name ______________________________

Your signature ______________________________

Signature of investigator ______________________________

Date ______________________________

Figure A.2: The informed concent form completed by students taking part in the main study (based on Loughborough University's 2008 template, the latest version of which may be found at `http://www.lboro.ac.uk/admin/committees/ethical/`).

Appendix B

Transcripts and Task Sheets Used in the Validation Exercise

B.1 Transcript for Student A

[Task phase begins]

Student A: You just want an example, right?

Interviewer: Yeah any example you like.

[Student completes the first four questions]

Interviewer: So what were you thinking as you did those four?

[Students begins to write]

Interviewer: Feel free to say it out loud, you don't have to write it down.

Student A: Oh, it increases like by a greater amount each time, by a lot more, than just like the same amount.

Student A: And this one, isn't increase that much, it just increasing the sequences, not greatly increasing.

Student A: Increasing/decreasing, like $1, -1, 1, -1$, keeps going up and down.

Student A: And then a sequence that isn't increasing or decreasing, that is just constant.

[Student moves on to the second page]

Student A: I don't know these.

Interviewer: Ok. If you want to have a look at the definition sheet to help you out.

[Student looks at definition sheet]

Interviewer: So what are you thinking as you read that definition.

Student A: I'm thinking what U is. So a_n could be anything less than U. But I don't know. Is it like when you go up to a point, that's up to U? But only all the points before that?

Interviewer: I don't want to say until you feel you're done. I'm happy to go through anything afterwards of course.

Student A: No, I don't know that one.

[Student moves on to question 6]

Student A: And here there's an L. So maybe that's like a_n's in beween U and L. But I don't know if the U counts.

[Student moves on to Question 7, writes answer]

Interviewer: So what did you think about that Question 7 then?

Student A: It's, where is it, it's monotonic if it's increasing or decreasing, but it doesn't have to be constant though does it? I don't think so.

[Student moves on to third sheet, answers question 8 and 9]

Interviewer: So what did you think during that question?

Student A: Because it can become negative infinity as well, can't it?

Interviewer: For this question actually it can't, it has to be positive. Can you not cross it out, draw a line underneath if you want to think about something else.

Interviewer: But for that one you were making it go less and less, were you?

Student A: Yeah.

Interviewer: So what have you got there, $1000, 100, 10, 0.1$, what was the next term going to be?

Student A: 0.01, then 0.001 and so on.

Interviewer: Right, ok. Yeah, sadly it can't tend to minus infinity, it just means positive infinity in this case.

Student A: It can also be a number like, $2, 2$, you can keep going on and on.

Interviewer: What do you mean, you kind of moved you hand there, so are you thinking of a picture?

Student A: Yeah.

Interviewer: Do you want to sketch what you mean?

Student A: Like it keeps going on and on to infinity this way.

Interviewer: So you mean/

Student A: But infinity probably means that way here.

Interviewer: Oh I see what you mean, which one do you think,

Student A: If it's not increasing it can't really go up, to go like that on a graph. So this is the only possible way isn't it?

Interviewer: So if it were like a constant sequence?

Student A: Just like a number, like say 2, it keeps going to infinity.

[Student moves on to question 10]

Student A: I don't understand that one.

Interviewer: Ok.

[Student moves on to question 11, sketches]

Student A: I'm just thinking what numbers that would go with.

Interviewer: Ok, so what have you done there?

Student A: This is not going, does not tend to infinity, so it just levels out and it's/

Interviewer: So you're thinking of some numbers that might fit that?

Interviewer: Have you seen that shape of graph before?

Student A: Yeah.

Interviewer: What sort of things?

Student A: I can't remember.

Student A: Oh, it's kind of like the x cubed graph.

Interviewer: Do you feel you've done them, do you want to look through them at all?

Student A: I think that's the best I can do.

B.2 Task sheet for Student A

Please give an example of each of the following, **or state that this is impossible**

You can write your sequence in any way you choose:
As a list of numbers, as a formula, etc.

You do no need to prove your answers

1. A strictly increasing sequence

1, 4, 9, 16, 25...

2. An increasing sequence that is not strictly increasing

1, 2, 3, 4, 5...

3. A sequence that is both increasing and decreasing

-1, 1, -1, 1, -1, 1, -1...

4. A sequence that is neither increasing nor decreasing

2, 2, 2, 2,...

Please give an example of each of the following, **or state that this is impossible**

You can write your sequence in any way you choose:
As a list of numbers, as a formula, etc.

You do no need to prove your answers

5. A sequence that has no upper bound

6. A sequence that has neither an upper bound nor a lower bound

7. A bounded, monotonic sequence

2, 4, 6, 8, 10...

10, 8, 6, 4, 2.....

Please give an example of each of the following, **or state that this is impossible**

You can write your sequence in any way you choose:
As a list of numbers, as a formula, etc.

You do no need to prove your answers

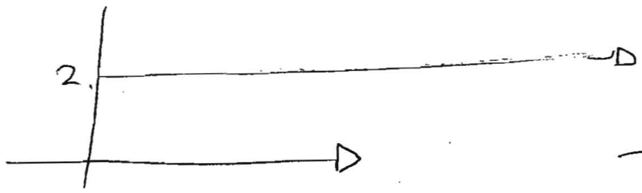
8. A sequence that tends to infinity

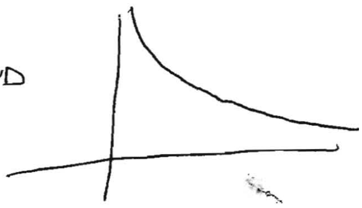
10, 100, 1000 . . . . . .

9. A sequence that tends to infinity that is not increasing

~~1000~~, 100, 10, 0.1, 0.01, 0.001 . . .

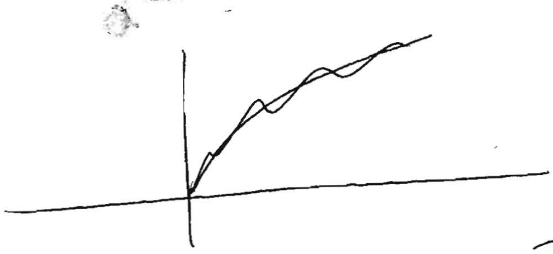
~~2, 2~~,

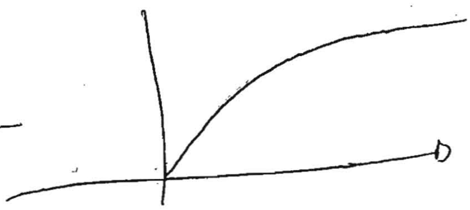




10. A sequence that tends to infinity that is not bounded below

11. A strictly increasing sequence that does not tend to infinity





B.3 Transcript for Student B

[Task phase begins]

[Student starts Question 1]

Student B: Strictly increasing sequence. Alright, strictly increasing sequence, that mean it just increases so you want, strictly increasing sequence, that's if a_{n+1} is more, exactly more than a_n. So this means that, well.

[Student starts Question 2]

Student B: Strictly increasing sequence [whispers definition]. Well, strictly increasing sequence, I suppose, maybe x_{n+1} is equal to $x_n + 1$. If that's maybe, yes that's a nice example, I think. I think that's a nice example.

[Student starts Question 3]

Student B: An increasing sequence that is not strictly increasing. Well, it's not strictly increasing, not strictly increasing. Then it just increases. So, what? So there must be a number where $a_{n+1} = a_n$. So, [begins whispering]: So a sequence of $a_n + 1$ maybe, $n + 1$, a sequence. No., A sequence of numbers, a sequence of numbers, $a_n = 1/n^2$ [back to normal volume] n squared, maybe a_n over n^2, no. [inaudible]

[Student starts Question 4]

Student B: A sequence that is both increasing and decreasing, a sequence which is neither increasing nor decreasing. Ah, right. Well a sequence of. Dirichlet's function I think. Yeah, it's not a sequence actually, but. It's not even a sequence, but, it should be [inaudible].

[Student starts Question 5]

Student B: A sequence that has no upper bound. There is no upper bound. A sequence of real numbers, a sequence of natural numbers, natural numbers.

[Student starts Question 6]

Student B: A sequence that has neither an upper bound nor a lower bound. Nor a lower bound, and no upper bound, one which goes to infinity and negative infinity. What goes to infinity and negative infinity?

[Student starts Question 7]

Student B: Bounded monotonic sequence, monotonic is of course increasing or decreasing. So in this case, well I could give one decreasing and one increasing. I suppose my increasing will be one as $a_n = n$, and $a_n = 1/n$, or whatever.

[Student starts Question 8]

Student B: A sequence that tends to infinity. Oh wow. Maybe. A sequence that tends to infinity. Positive numbers, for n in the natural numbers, even in the real numbers. Even in the real numbers is ok.

[Student starts Question 9]

Student B: A sequence that tends to infinity that is not increasing. Oh. When you state infinity, do you also mean negative infinity as well

Interviewer: For these questions, always positive.

Student B: Always positive, and erm. Not possible I think. I could be wrong, but.

[Student starts Question 10]

Student B: A sequence that tends to infinity that is not bounded below. [Repeats question]. Sequence of real numbers, sequence of rational, or real numbers. Yep.

[Student starts Question 11]

Student B: Strictly increasing that does not tend to infinity. [repeats question]. Tends to a certain, tends to a certain upper bound. Certain upper bound, which means. Maybe, I need to see this. [inaudible]. [Writes] Yes. Ok.

[Student returns to the second sheet, Question 5]

Student B: A sequence that has neither an upper bound nor a lower bound. A sequence that is both increasing and decreasing. A sequence, not possible.

[Student returns to the first sheet, Question 2]

Student B: A sequence that is not strictly increasing. That means that there is something that [reads definition]. Need to find an a_n such that it is the same. It is not strictly increasing so $a_n + 1 = a_n$. What number? Yes. Maybe a_n equals $1 - 1/n$?. That's not strictly going plus one way. $1 + 1/n$, yes, so limit a_n, n tends to infinity is just 1.

[Student turns to the second sheet, Question 6]

Student B: A sequence that has neither an upper bound nor a lower bound. [repeats question twice]. If there is a sequence then it's completely not bounded at all. What sequence is not completely not bounded at all.

Student B: Ok, $1/n$, where n is in the real numbers. Yes.

Student B: Right, well it's not as easy, I'm sort of finished but, strictly increasing sequence, need to set my parameters, n equals, n is in the natural numbers. And well, some are in the natural numbers, some are not.

[Student goes through each question adding domains]

Student B: So neither increasing nor decreasing, that's a function actually, Dirichlet's function. But then again a function is from n, a sequence as well.

Interviewer: What was that, sorry? I missed that

Student B: I mean a function is merely a sequence as well, yes.

Student B: I think I'm finished, but I know it's probably wrong.

[Discussion phase begins]

Interviewer: So, ok. Of the eleven questions, which are you most sure about and which are you least sure about?

Student B: Well, most sure about. [Turns to Question 6] Actually, natural numbers. Upper bound nor a lower bound. This one is ok, but n is in the real numbers not in the natural numbers. This one, [Question 7] increasing, monotonic, this one, with n in the natural numbers.

Interviewer: So which answer?

Student B: One of the, it could be increasing or decreasing. So I'm just giving one example for increasing and decreasing.

Interviewer: Ok. A sequence that tends to infinity, well. Yeah, basically this one I think. I suppose it tends to infinity, yeah it does. It does because it is a positive power.

Student B: A sequence that tends to infinity that is not increasing. Can't think of anything.

Student B: A sequence that tends to infinity that is not bounded below. Well, not bounded below, I suppose the sequence of rational numbers or real numbers, and then yeah it wouldn't be bounded below.

Student B: A strictly increasing sequence that does not tend to infinity. Ok, so I gave $a_n = (2n + 4)/n$, which means the limit of a_n is 2, basically. Yeah you can work this out, it's 2 as n tends to infinity. And, is it increasing, well, is it increasing? Might not be increasing, well let me see. This is actually decreasing.

Student B: Err, well one thing that actually does not tend to infinity but actually has a strict upper bound. You want a sup, a sup.

Interviewer: What are you imagining, your eyes are close, so are you thinking about something?

Student B: I'm just visualising the numbers. Let me see, if I had n^2, if I had $n = 2$.

Student B: Actually no, this is increasing actually. it is increasing. This is increasing actually. Let me see $2 + 4/n$ becomes smaller, so no it's still decreasing.

Student B: Erm, minus. Minus? Yes, instead $a_n = (2n - 4)/n$. Yes, this one, it means -4, -2 yes, but. It's increasing, but the limit of a_n is still 2. And yes, increasing. Sorry about this.

Interviewer: No that's alright, it's good to see people thinking about things, lovely.

Student B: Strictly increasing sequence for n^2. Strictly increasing. Yes.

Interviewer: So/

Student B: It depends on the parameters of course but, yes. I should write the parameters for this, n in the natural numbers. For the others it's n in the natural numbers.

Interviewer: So, as you've notice yourself, some of them you've chosen n in the natural numbers, and some, well [Question 4] we've got the real overall because we've got rational and irrational so together they're the real numbers, and we've got the reals here. So what influenced which one you chose?

Student B: Well, the negative basically. I needed a negative, and for sequences normally you use the natural numbers: sequence number 1, sequence number 2, by iterations 1,2,3, follow the natal numbers. But, in some cases I want, it says here neither an upper nor a lower bound I know that the natural numbers has a lower bound, which is zero. Well, not exactly zero, but it is,

Interviewer: It depends I guess how you define it

Student B: Yeah, and well, I just needed a negative basically, so it's probably a little bit, maybe it's not exactly right but it, I used the properties that I know to suit the situation. Which is, I suppose, isn't all that unique, because you do it all the time for any decision you make, so.

Interviewer: Yeah, that makes sense. So when you were thinking about these, did you have any pictures in your head, or were you thinking. When I asked you before, you said "the numbers",

Student B: Yes, the numbers

Interviewer: So you were imagining the terms of the sequence?

Student B: Yes

Interviewer: Is this the case with all of them?

Student B: Erm, more or less, yes.

Interviewer: Great, and so, but you didn't really think of any graphs or any pictures?

Student B: Erm

Interviewer: Please don't pretend if you didn't, I'm actually interested in whether you did or not?

Student B: Well, not all of them. But let's see, for the Dirichlet function I did, because I know the graph. And for $1/n$ I did, and yeah. I suppose not all, about 30-40% you do think of a graph, but the other 60-70% probably not. So the majority, to summarise, the majority of time no, but occasionally I think it's nice to have a geometric representation.

Interviewer: And have you ever been asked to do things like this before? To give examples?

Student B: Well not like this, I mean, not all in one go. I suppose informally yeah.

B.4 Task sheet for Student B

Please give an example of each of the following, **or state that this is impossible**

You can write your sequence in any way you choose:
As a list of numbers, as a formula, etc.

You do no need to prove your answers

1. A strictly increasing sequence

~~$x_{n+1} = x_n + 1$~~

$a_n = n^2$, $n \in \mathbb{N}$

2. An increasing sequence that is not strictly increasing

$a_n = 1 - \frac{1}{n}$ $n \in \mathbb{N}$

[illegible]

$\lim_{n \to \infty} a_n = 1$

3. A sequence that is both increasing and decreasing

~~Not possible~~

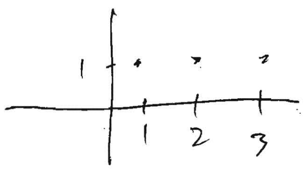
$a_n = (-1)^n(n)$

4. A sequence that is neither increasing nor decreasing

Derichlet's Function

$f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}$

$a_n = 1$



Please give an example of each of the following, **or state that this is impossible**

You can write your sequence in any way you choose:
As a list of numbers, as a formula, etc.

You do no need to prove your answers

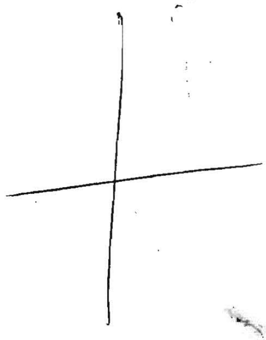
5. A sequence that has no upper bound

sequence of Natural numbers

6. A sequence that has neither an upper bound nor a lower bound

$a_n = \frac{1}{n}$, $n \in \mathbb{R}$

$a_n = (-1)^n (n)$



7. A bounded, monotonic sequence

Increasing or decreasing

$a_n = n$

$a_n = \frac{1}{n}$

$\}$ $n \in \mathbb{N}$

Please give an example of each of the following, **or state that this is impossible**

You can write your sequence in any way you choose:
As a list of numbers, as a formula, etc.

You do no need to prove your answers

8. A sequence that tends to infinity

$a_n = n^4$, $n \in \mathbb{R}$

9. A sequence that tends to infinity that is not increasing

Not ~~possible~~ possible

10. A sequence that tends to infinity that is not bounded below

~~sequence of Q, Q, IR~~

$a_n = (-1)^n (n)$

11. A strictly increasing sequence that does not tend to infinity

$a_n = \frac{2n+4}{n}$ [Decreasing ~~Increasing~~ ~~Decreasing~~ Sequence]

$\lim_{n \to \infty} a_n = 2$

$\to n \in \mathbb{N}$

$a_n = \frac{2n-4}{n}$

$\lim_{n \to \infty} a_n = 2$

[Increasing sequence]

Appendix C

Data from Validation Study

Student	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
WA01	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**C**	**C**
WA02	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WA03	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WA04	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WA05	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WA06	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WA07	**C**	**C**	**C**	**C**	**C**	**IS**	**B**	**C**	**B**	**B**	**B**
WA08	**C**	**IS**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WA09	**C**	**IS**	**C**	**C**	**C**	**IS**	**C**	**C**	**II**	**IS**	**C**
WA10	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WA11	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**
WA12	**C**	**C**	**C**	**C**	**C**	**II**	**IS**	**C**	**B**	**B**	**IS**
WA13	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**
WA14	**C**	**IS**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WA15	**INS**	**IS**	**IS**	**C**	**C**	**C**	**C**	**C**	**IS**	**IS**	**B**
WA16	**C**	**C**	**IS**	**IS**	**C**	**C**	**IS**	**C**	**C**	**C**	**IS**
WA17	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**IS**	**IS**
WA18	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**
WA19	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WA20	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WA21	**C**	**C**	**II**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**
WA22	**C**	**IS**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WA23	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WA24	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WA25	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WA26	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**C**	**C**
WA27	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WA28	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WA29	**INS**	**INS**	**INS**	**C**	**C**	**IS**	**C**	**C**	**B**	**IS**	**IS**
WA30	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WA31	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WA32	**C**	**C**	**C**	**II**	**C**	**IS**	**IS**	**C**	**C**	**IS**	**IS**
WA33	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**

Continued on next page

Student	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
WA34	**C**	**B**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WA35	**C**	**C**	**II**	**C**	**C**	**B**	**B**	**C**	**IS**	**B**	**B**
WA36	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WA37	**C**	**C**	**IS**	**C**	**C**	**IS**	**IS**	**C**	**C**	**IS**	**C**
WA38	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WB01	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WB02	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WB03	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WB04	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**C**	**C**
WB05	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**B**	**C**
WB06	**C**	**IS**	**C**	**C**	**C**	**IS**	**IS**	**C**	**C**	**C**	**C**
WB07	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**II**
WB08	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**IS**	**C**
WB09	**INS**	**INS**	**INS**	**INS**	**INS**	**INS**	**INS**	**INS**	**INS**	**B**	**INS**
WB10	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WB11	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WB12	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**	**II**	**C**	**C**
WB13	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**B**
WB14	**C**	**C**	**IS**	**IS**	**C**	**IS**	**C**	**C**	**IS**	**IS**	**II**
WB15	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**C**	**C**
WB16	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**C**	**C**
WB17	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WB18	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WB19	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WB20	**C**	**C**	**IS**	**IS**	**C**	**C**	**C**	**C**	**II**	**C**	**II**
WB21	**C**	**C**	**IS**	**II**	**C**	**IS**	**C**	**C**	**IS**	**IS**	**IS**
WB22	**C**	**C**	**IS**	**II**	**C**	**C**	**C**	**C**	**IS**	**IS**	**C**
WB23	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WB24	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WB25	**C**	**IS**	**IS**	**C**	**C**	**C**	**C**	**C**	**IS**	**C**	**B**
WB26	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WB27	**C**	**C**	**IS**	**IS**	**C**	**C**	**IS**	**C**	**IS**	**IS**	**C**
WB28	**C**	**C**	**IS**	**C**	**C**	**C**	**C**	**C**	**II**	**IS**	**IS**
WB29	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WB30	**C**	**IS**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**C**	**C**
WC01	**C**	**C**	**C**	**II**	**C**	**C**	**IS**	**C**	**C**	**C**	**II**
WC02	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WC03	**C**	**C**	**C**	**C**	**C**	**IS**	**C**	**C**	**IS**	**C**	**C**
WC04	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WC05	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WC06	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WC07	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**C**	**C**
WC08	**C**	**C**	**IS**	**IS**	**C**	**II**	**C**	**C**	**II**	**C**	**C**
WC09	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**C**	**C**
WC10	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WC11	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**C**	**C**
WC12	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WC13	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WC14	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**IS**	**C**
WC15	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**

Continued on next page

Student	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
WC16	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WC17	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WC18	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**C**	**C**
WC19	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**B**	**C**
WC20	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**C**	**II**	**C**	**C**
WC21	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WC22	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WC23	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WC24	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WC25	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WD01	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**C**	**II**	**C**	**C**
WD02	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WD03	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WD04	**C**	**C**	**INS**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WE01	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WE02	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WE03	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WE04	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WE05	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WE06	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WE07	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**	**C**	**C**	**INS**
WE08	**C**	**C**	**B**	**IS**	**C**	**IS**	**IS**	**C**	**II**	**IS**	**C**
WE09	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**B**	**C**	**B**
WE10	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**IS**	**C**
WE11	**C**	**IS**	**IS**	**IS**	**C**	**IS**	**C**	**C**	**IS**	**IS**	**II**
WE12	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WE13	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**II**
WE14	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**
WE15	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**IS**	**C**
WE16	**C**	**B**	**C**	**IS**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WE17	**C**	**C**	**IS**	**IS**	**C**	**C**	**IS**	**C**	**II**	**C**	**C**
WE18	**C**	**C**	**IS**	**IS**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WE19	**C**	**C**	**C**	**C**	**C**	**IS**	**C**	**C**	**IS**	**IS**	**C**
WE20	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WE21	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**
WE22	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WE23	**C**	**C**	**C**	**C**	**C**	**IS**	**C**	**C**	**II**	**IS**	**C**
WE24	**C**	**C**	**IS**	**IS**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WE25	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**C**	**C**	**C**	**IS**
WE26	**IS**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WE27	**IS**	**C**	**IS**	**IS**	**C**	**INS**	**C**	**C**	**C**	**C**	**C**
WE28	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WE39	**C**	**C**	**C**	**C**	**C**	**IS**	**C**	**C**	**C**	**IS**	**IS**
WE30	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**	**II**	**C**	**C**
WE31	**C**	**C**	**IS**	**IS**	**C**	**C**	**C**	**C**	**II**	**C**	**IS**
WF01	**IS**	**IS**	**IS**	**C**	**INS**	**INS**	**B**	**B**	**B**	**B**	**B**
WF02	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**B**	**B**	**B**
WF03	**C**	**IS**	**II**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**II**
WF04	**C**	**C**	**II**	**II**	**C**	**C**	**C**	**C**	**IS**	**B**	**B**
WF05	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**B**	**C**	**B**

Continued on next page

Student	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11
WF06	**C**	**IS**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WF07	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WF08	**C**	**IS**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WF09	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WF10	**C**	**B**	**C**	**C**	**C**	**C**	**C**	**C**	**II**	**C**	**C**
WF11	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WF12	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WF13	**C**	**INS**	**C**	**IS**	**C**	**C**	**C**	**C**	**IS**	**C**	**C**
WF14	**C**	**IS**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**C**	**C**
WF15	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
WF16	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**IS**	**C**
WF17	**C**	**C**	**IS**	**IS**	**C**	**C**	**C**	**C**	**II**	**B**	**IS**
WF18	**C**	**C**	**IS**	**IS**	**C**	**C**	**C**	**C**	**II**	**C**	**B**
WF19	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**IS**	**C**	**C**

Appendix D

MATLAB Code Used in Subsection 8.3.2

D.1 Code for the bootstrap

```
%Sets up how big the data set, and how many runs
[persons,items]=size(data);
times=1000000;

%Sets up dummy matrices to hold final and working data
AS=ones(times,1);

%Routine to fill up AS
for x=1:times
    R= randi(persons,persons,1);
    A=ones(persons,items);
    j=ones(items,1);
    for i=1:persons,
         A(i,:)=data(R(i),:);
    end
    COV=cov(A);
    AS(x)=(items/(items-1))*(1-(trace(COV))/(j'*COV*j));
end

%Outputting statistics
avg=mean(AS)
var=std(AS)

%Draw a histogram
hist(AS)
```

D.2 Code for Duhachek and Iaobucci's formula

```
% Setting up variables
p=11
n=147
j=ones(p,1)
V

% Entering the formula
Q=((2*p^2)/((p-1)^2*(j'*V*j)*3))*((j'*V*j)*(trace(V^2)+trace(V)^2)-
2*(trace(V))*(j'*V^2*j))
ASE=sqrt(Q/n)
conint=1.96*ASE

% Computing alpha
Alpha=(p/(p-1))*(1-(trace(V))/(j'*V*j))
Alpha_plus=Alpha+conint
Alpha_minus=Alpha-conint
```

Appendix E

Rasch analysis Output of all Person Statistics

Produced Jul 22 15:55 2010

ENTRY					INFIT		OUTFIT	
NUMBER	MEASURE	COUNT	SCORE	S.E.	MNSQ	STD	MNSQ	STD
1	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
2	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
3	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
4	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
5	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
6	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
7	0.46	11.0	6.0	0.75	1.01	0.14	0.71	0.12
8	2.40	11.0	9.0	0.95	0.94	0.10	0.86	0.37
9	1.02	11.0	7.0	0.76	1.12	0.51	0.78	0.30
10	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
11	3.58	11.0	10.0	1.25	1.79	1.04	0.89	0.39
12	0.46	11.0	6.0	0.75	1.01	0.14	0.71	0.12
13	3.58	11.0	10.0	1.25	1.79	1.04	0.89	0.39
14	2.40	11.0	9.0	0.95	0.94	0.10	0.86	0.37
15	−0.12	11.0	5.0	0.78	1.25	0.75	1.04	0.39
16	1.02	11.0	7.0	0.76	1.60	1.92	2.39	1.18
17	1.64	11.0	8.0	0.82	0.63	−0.97	0.39	−0.10
18	3.58	11.0	10.0	1.25	1.79	1.04	0.89	0.39
19	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
20	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
21	2.40	11.0	9.0	0.95	1.76	1.22	1.10	0.54
22	2.40	11.0	9.0	0.95	0.94	0.10	0.86	0.37
23	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
24	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
25	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
26	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
27	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
28	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
29	−0.77	11.0	4.0	0.84	1.27	0.68	0.92	0.32

Continued on next page

ENTRY NUMBER	MEASURE	COUNT	SCORE	S.E.	INFIT MNSQ	INFIT STD	OUTFIT MNSQ	OUTFIT STD
30	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
31	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
32	0.46	11.0	6.0	0.75	1.48	1.61	3.47	1.70
33	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
34	3.58	11.0	10.0	1.25	2.02	1.21	2.48	1.22
35	−0.12	11.0	5.0	0.78	0.86	−0.27	0.60	−0.09
36	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
37	1.02	11.0	7.0	0.76	1.77	2.35	2.52	1.24
38	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
39	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
40	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
41	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
42	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
43	3.58	11.0	10.0	1.25	1.86	1.09	1.07	0.52
44	1.64	11.0	8.0	0.82	2.25	2.50	2.32	1.16
45	2.40	11.0	9.0	0.95	0.62	−0.51	0.34	−0.16
46	2.40	11.0	9.0	0.95	0.70	−0.35	0.41	−0.08
47	−5.58	11.0	0.0	1.99	1.00	0.00	1.00	0.00
48	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
49	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
50	2.40	11.0	9.0	0.95	0.95	0.12	0.91	0.41
51	2.40	11.0	9.0	0.95	0.62	−0.51	0.34	−0.16
52	−0.12	11.0	5.0	0.78	0.63	−1.07	0.44	−0.33
53	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
54	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
55	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
56	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
57	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
58	1.02	11.0	7.0	0.76	0.78	−0.74	0.53	0.07
59	−0.12	11.0	5.0	0.78	0.63	−1.07	0.44	−0.33
60	1.02	11.0	7.0	0.76	0.85	−0.49	0.57	0.11
61	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
62	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
63	1.02	11.0	7.0	0.76	0.88	−0.36	0.60	0.14
64	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
65	0.46	11.0	6.0	0.75	1.03	0.21	0.76	0.18
66	1.02	11.0	7.0	0.76	0.61	−1.51	0.42	−0.06
67	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
68	2.40	11.0	9.0	0.95	0.94	0.10	0.86	0.37
69	1.64	11.0	8.0	0.82	1.82	1.82	1.90	0.98
70	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
71	2.40	11.0	9.0	0.95	0.95	0.12	0.91	0.41
72	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
73	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
74	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
75	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
76	1.02	11.0	7.0	0.76	1.13	0.53	0.77	0.29
77	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
78	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00

Continued on next page

ENTRY					INFIT		OUTFIT	
NUMBER	MEASURE	COUNT	SCORE	S.E.	MNSQ	STD	MNSQ	STD
79	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
80	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
81	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
82	2.40	11.0	9.0	0.95	0.70	−0.35	0.41	−0.08
83	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
84	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
85	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
86	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
87	3.58	11.0	10.0	1.25	1.86	1.09	1.07	0.52
88	2.40	11.0	9.0	0.95	1.01	0.22	1.30	0.67
89	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
90	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
91	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
92	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
93	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
94	2.40	11.0	9.0	0.95	1.01	0.22	1.30	0.67
95	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
96	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
97	3.58	11.0	10.0	1.25	1.93	1.15	1.44	0.75
98	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
99	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
100	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
101	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
102	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
103	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
104	2.40	11.0	9.0	0.95	1.91	1.38	1.48	0.77
105	−0.12	11.0	5.0	0.78	1.04	0.22	0.88	0.23
106	2.40	11.0	9.0	0.95	0.62	−0.51	0.34	−0.16
107	2.40	11.0	9.0	0.95	0.70	−0.35	0.41	−0.08
108	−0.77	11.0	4.0	0.84	0.51	−1.06	0.34	−0.39
109	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
110	3.58	11.0	10.0	1.25	1.79	1.04	0.89	0.39
111	3.58	11.0	10.0	1.25	1.79	1.04	0.89	0.39
112	2.40	11.0	9.0	0.95	0.70	−0.35	0.41	−0.08
113	2.40	11.0	9.0	0.95	2.14	1.62	1.73	0.90
114	1.02	11.0	7.0	0.76	1.22	0.83	0.88	0.38
115	1.64	11.0	8.0	0.82	0.98	0.08	0.64	0.18
116	1.64	11.0	8.0	0.82	0.97	0.07	0.70	0.23
117	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
118	3.58	11.0	10.0	1.25	1.79	1.04	0.89	0.39
119	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
120	1.64	11.0	8.0	0.82	0.97	0.07	0.70	0.23
121	1.64	11.0	8.0	0.82	0.98	0.08	0.64	0.18
122	2.40	11.0	9.0	0.95	1.97	1.45	1.87	0.97
123	3.58	11.0	10.0	1.25	2.13	1.29	9.90	4.05
124	1.02	11.0	7.0	0.76	2.15	3.22	3.84	1.67
125	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
126	1.64	11.0	8.0	0.82	1.56	1.36	1.58	0.82
127	2.40	11.0	9.0	0.95	0.95	0.12	0.91	0.41

Continued on next page

ENTRY NUMBER	MEASURE	COUNT	SCORE	S.E.	INFIT MNSQ	INFIT STD	OUTFIT MNSQ	OUTFIT STD
128	1.02	11.0	7.0	0.76	0.78	−0.74	0.53	0.07
129	−3.92	11.0	1.0	1.30	2.32	1.58	9.26	2.81
130	1.02	11.0	7.0	0.76	1.14	0.58	4.37	1.82
131	1.02	11.0	7.0	0.76	0.88	−0.36	0.60	0.14
132	0.46	11.0	6.0	0.75	0.60	−1.61	0.42	−0.24
133	2.40	11.0	9.0	0.95	0.62	−0.51	0.34	−0.16
134	3.58	11.0	10.0	1.25	2.02	1.21	2.48	1.22
135	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
136	2.40	11.0	9.0	0.95	0.94	0.10	0.86	0.37
137	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
138	2.40	11.0	9.0	0.95	0.94	0.10	0.86	0.37
139	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
140	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
141	1.64	11.0	8.0	0.82	1.14	0.46	0.81	0.32
142	2.40	11.0	9.0	0.95	0.94	0.10	0.86	0.37
143	5.21	11.0	11.0	1.98	1.00	0.00	1.00	0.00
144	2.40	11.0	9.0	0.95	0.70	−0.35	0.41	−0.08
145	0.46	11.0	6.0	0.75	0.60	−1.61	0.42	−0.24
146	1.02	11.0	7.0	0.76	0.78	−0.74	0.53	0.07
147	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
Anna	1.02	11.0	7.0	0.76	1.11	0.47	0.75	0.28
Ben	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
Valter	2.40	11.0	9.0	0.95	0.70	−0.35	0.41	−0.08
David	1.02	11.0	7.0	0.76	0.78	−0.74	0.53	0.07
Edha	−2.56	11.0	2.0	1.07	1.52	0.90	1.73	0.90
Phalgun	1.64	11.0	8.0	0.82	0.63	−0.97	0.39	−0.10
Guan	1.64	11.0	8.0	0.82	0.80	−0.41	0.53	0.06
Haaroon	−0.77	11.0	4.0	0.84	0.51	−1.06	0.34	−0.39
Ian	−0.12	11.0	5.0	0.78	1.02	0.16	0.86	0.22
Joe	0.46	11.0	6.0	0.75	0.90	−0.28	0.67	0.08
Ken	3.58	11.0	10.0	1.25	0.33	−0.80	0.11	−0.64
Laura	1.64	11.0	8.0	0.82	0.80	−0.41	0.53	0.06
Mike	3.58	11.0	10.0	1.25	1.86	1.09	1.07	0.52
Nicola	1.02	11.0	7.0	0.76	0.89	−0.31	0.67	0.20
Oksana	3.58	11.0	10.0	1.25	1.86	1.09	1.07	0.52

Bibliography

Åkerlind, G. (2005a). Learning about phenomenography: Inerviewing, data analysis and the qualitative research paradigm. In Bowden, J. and Green, P., editors, *Doing developmental phenomenography*, chapter 6, pages 63–73. RMIT University Press, Melbourne, Australia.

Åkerlind, G. (2005b). Phenomenographic methods: A case illustration. In Bowden, J. and Green, P., editors, *Doing developmental phenomenography*, chapter 8, pages 103–127. RMIT University Press, Melbourne, Australia.

Åkerlind, G. (2005c). Variation and commonality in phenomenographic research methods. *Higher Education Research and Development*, 24(4):321–334.

Al-Murani, T. (2006). Teachers' awareness of dimensions of variation: a mathematics intervention project. In Hewitt, D., editor, *Proceedings of the British Socienty for Research into Learning Mathematics*, volume 26, pages 1–6. BSRLM, London.

Alcock, L. (2004). Uses of example objects in proving. In Høines, M. and Fuglestad, A., editors, *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education*, volume 2, pages 17–24. PME, Bergen, Norway.

Alcock, L. and Inglis, M. (2008). Doctoral students' use of examples in evaluating and proving conjectures. *Educational Studies in Mathematics*, 69(2):111–129.

Alcock, L. and Simpson, A. (2001). The Warwick Analysis project: Practice and theory. In Holton, D., editor, *The Teaching and Learning of Mathematics at University Level*, volume 7, pages 99–112. Kluwer Academic Publishers, Netherlands.

Alcock, L. and Simpson, A. (2002). Definitions: Dealing with categories mathematically. *For the Learning of Mathematics*, 22(2):28–34.

Alcock, L. and Simpson, A. (2009a). *Ideas from mathematics education: An introduction for mathematicians.* Higher Education Academy, MSOR Network. Retrieved February 21, 2011 from `http://www.mathstore.ac.uk/repository/AlcockSimpsonBook.pdf`.

Alcock, L. and Simpson, A. (2009b). The role of definitions in example classification. In Tzekaki, M., Kaldrimidou, M., and Sakonidis, C., editors, *Proceedings of the 33rd Conference of the International Group for the Psychology of Mathematics Education*, volume 2, pages 33–40. PME, Thessaloniki, Greece.

Antonini, S. (2001). Negation in mathematics: Obstacles emerging from an exploratory study. In Heuvel-Panhuizen, M., editor, *Proceedings of the 25th Conference of the International Group for the Psychology of Mathematics Education*, volume 2, pages 2–49. PME, Utrecht University, Netherlands.

Antonini, S. (2006). Graduate students' processes in generating examples of mathemtical objects. In Novotná, J., Moraová, H., Krátká, M., and Stehlíková, N., editors, *Proceedings of the 30th Conference of the International Group for the Psychology of Mathematics Education*, volume 2, pages 57–64. PME, Prague.

Artigue, M. (1991). Analysis. In Tall, D., editor, *Advanced Mathematical Thinking*, Mathematics Education Library, chapter 11, pages 167–198. Kluwer Academic Publishers, Netherlands.

Asghari, A. (2004). Organizing with a focus on defining: A phenomenographic approach. In Høines, M. and Fuglestad, A., editors, *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education*, volume 2, pages 63–70. PME, Bergen, Norway.

Asghari, A. (2005). Examples: Generating versus checking. In Hewitt, D. and Noyes, A., editors, *Proceedings of the sixth British Congress of Mathematics Education*, pages 25–32. University of Warwick, Coventry, UK.

Asghari, A. (2007). Examples, a missing link. In Woo, J., Lew, H., Park, K., and Seo, D., editors, *Proceedings of the 31st Conference of the International Group for the Psychology of Mathematics Education*, volume 2, pages 25–32. PME, Seoul.

Baker, F. B. and Kim, S.-H. (2004). *Item Response Theory: Parameter Estimation Techniques.* Marcel Dekker, New York, second edition.

Baker, J., Crampin, M., and Nuttall, J. (1973). A crash course in calculus. *International Journal of Mathematical Education in Science and Technology*, 4(3):335–339.

Bérgé, A. (2006). Convergence of numerical sequences: A commentary on "the vice: Some historically inspired and proof generated steps to limits of sequences" by R. P. Burn. *Educational Studies in Mathematics*, 61(3):395–402.

Bessant, K. (1995). Factors associated with types of mathematics anxiety in college students. *Journal for research in mathematics education*, 26(4):327–345.

Bills, L., Dreyfus, T., Mason, J., Tsamir, P., Watson, A., and Zaslavsky, O. (2006). Exemplification in mathematics education. In Novotná, J., Moraová, H., Krátká, M., and Stehlíková, N., editors, *Proceedings of the 30th Conference of the International Group for the Psychology of Mathematics Education*, volume 1, pages 126–154. PME, Prague.

Bingolbali, E. and Monaghan, J. (2008). Concept image revisited. *Educational Studies in Mathematics*, 68(1):19–35.

Biza, I., Nardi, E., and Zachariades, T. (2008). Persistent images and teacher beliefs about visualisation: The tangent at an inflection point. In Figueras, O., Cortina, J., Alatorre, S., and Sepulveda, A., editors, *Proceedings of the Joint Meeting of PME 32 and PME-NA 30*, volume 2, pages 177–184, Mexico. Cinvestav-UMSNH.

Bond, T. G. and Fox, C. M. (2007). *Applying the Rasch Model.* Lawrence Erlbaum Associates, Mahwah, New Jersey.

Bowden, J. (2005). Reflections of the phenomenographic team research process. In Bowden, J. and Green, P., editors, *Doing developmental phenomenography*, chapter 2, pages 11–31. RMIT University Press, Melbourne, Australia.

Bowden, J. and Green, P. (2005). *Doing developmental phenomenography.* RMIT University Press, Melbourne, Australia.

Bradley, K., Sampson, S., and Royal, K. (2006). Applying the Rasch rating scale model to gain insights into students' conceptualisation of quality mathematics instruction. *Mathematics Education Research Journal*, 18(2):11.

Bryman, A. (2004). *Social Research Methods.* Oxford University Press, Oxford, 2nd edition.

Burn, R. P. (1992). *Numbers and functions: Steps into analysis.* Cambridge University Press, Cambridge, UK.

Callingham, R. and Bond, T. (2006). Editorial: Research in mathematics education and Rasch measurement. *Mathematics Education Research Journal*, 18(2):1–10.

Çetin, N. (2009). The performance of undergraduate students in the limit concept. *International journal of mathematical education in science and technology*, 40(3):323.

Chae, S. and Tall, D. (2001). Construction of conceptual knowledge: The case of computer-aided exploration of period doubling. *Research in Mathematics Education*, 3(1):199–209.

Chazan, D. (1993). High school geometry students' justification for their views of empirical evidence and mathematical proof. *Educational Studies in Mathematics*, 24(4):359–387.

Chen, C.-M., Lee, H.-M., and Chen, Y.-H. (2005). Personalized e-learning system using item response theory. *Computers & Education*, 44(3):237–255.

Chi, M., Bassok, M., Lewis, M., Reimann, P., and Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science: A Multidisciplinary Journal*, 13(2):145–182.

Chick, H. L. (2009). Choice and use of examples as a window on mathematics knowledge for teaching. *For the Learning of Mathematics*, 29(3):26–30.

Clark, M. and Lovric, M. (2008). Suggestion for a theoretical model for secondary-tertiary transition in mathematics. *Mathematics Education Research Journal*, 20(2):25–37.

Clark, M. and Lovric, M. (2009). Understanding secondary–tertiary transition in mathematics. *International Journal of Mathematical Education in Science and Technology*, 40(6):755–776.

Clark, R., Nguyen, F., and Sweller, J. (2006). *Efficiency in learning: Evidence-based guidelines to manage cognitive load.* Pfeiffer, San Francisco.

Coe, R. (2008). Comparability of GCSE examinations in different subjects: An application of the rasch model. *Oxford Review of Education*, 34(5):609–636.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Lawrence Erlbaum, New Jersey, USA.

Cohen, L., Manion, L., and Morrision, K. (2007). *Research methods in education.* Routledge, Abingdon, Oxford, 6th edition.

Copes, L. (1982). The perry development scheme: A metaphor for leaning and teaching mathematics. *For the Learning of Mathematics*, 3(1):38–44.

Cornu, B. (1991). Limits. In Tall, D., editor, *Advanced Mathematical Thinking*, Mathematics Education Libraryh, chapter 10, pages 153–166. Kluwer Academic Publishers, Netherlands.

Cortina, J. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, 78(1):98–104.

Cottrill, J., Dubinsky, E., Nichols, D., Schwingendorf, K., Thomas, K., and Vidakovic, D. (1996). Understanding the limit concept: Beginning with a coordinated process scheme. *Journal of Mathematical Behavior*, 15(2):167–192.

Crawford, K., Gordon, S., Nicholas, J., and Prosser, M. (1994). Conceptions of mathematics and how it is learned: The perspectives of students entering university. *Learning and Instruction*, 4(4):331–345.

Crawford, K., Gordon, S., Nicholas, J., and Prosser, M. (1998). University mathematics students' conceptions of mathematics. *Studies in Higher Education*, 23(1):87–94.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3):297–334.

Crouch, C. and Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, 69:970.

Dahlberg, R. P. and Housman, D. L. (1997). Facilitating learning events through example generation. *Educational Studies in Mathematics*, 33(3):283–299.

Dobby, J. and Duckworth, D. (1979). *Objective Assessment by Means of Item Banking.* Evans Bros for the Schools Council/Methuen, London.

Draper, S. (2009). Catalytic assessment: Understanding how MCQs and EVS can foster deep learning. *British Journal of Educational Technology*, 40(2):285–293.

Dubinsky, E. (1991). Reflective abstraction. In Tall, D., editor, *Advanced Mathematical Thinking*, Mathematics Education Library, chapter 7, pages 95–123. Kluwer Academic Publishers, Netherlands.

Dubinsky, E., Elterman, F., and Gong, C. (1988). The student's construction of quantification. *For the learning of mathematics*, 8(2):44–51.

Dubinsky, E. and Yiparaki, O. (2000). On student understanding of AE and EA quantification. *Research in collegiate mathematics education IV*, pages 239–286.

Duhachek, A. and Iacobucci, D. (2004). Alpha's standard error (ASE): An accurate and precise confidence interval estimate. *Journal of Applied Psychology*, 89(5):792–808.

Edwards, A. (2010). Using the rasch analysis fit-statistic to identify uncharacteristic responses to undergraduate tasks. In Green, D., editor, *Proceedings of the 2009 CETL-MSOR Conference.*, pages 31–36. The Maths, Stats & OR Network, Milton Keynes.

Edwards, A. and Alcock, L. (2008). Using example generation to expore students' concept images of sequence properties. In Figueras, O., Cortina, J., Alatorre, S., and Sepulveda, A., editors, *Proceedings of the Joint Meeting of PME 32 and PME-NA 30*, volume 1, page 253, Mexico. Cinvestav-UMSNH.

Edwards, A. and Alcock, L. (2010a). How do undergraduate students navigate their example spaces? In *Proceedings of the 13th Annual Conference on Research in Undergraduate Mathematics Education*, Raleigh, NC. Retrieved February 21, 2011 from `http://sigmaa.maa.org/rume/crume2010/Archive/Edwards.pdf`.

Edwards, A. and Alcock, L. (2010b). Using rasch analysis to identify uncharacteristic responses to undergraduate assessments. *Teaching Mathematics and its Applications*, 29(4):165–175.

Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical science*, 1(1):54–75.

Emons, W., Sijtsma, K., and Meijer, R. (2005). Global, local, and graphical person-fit analysis using person-response functions. *Psychological Methods*, 10:101–119.

Entwistle, N. (1997). Introduction: Phenomenography in higher education. *Higher Education Research and Development*, 16(2):127–134.

Epp, S. (2003). The role of logic in teaching proof. *American Mathematical Monthly*, 110(10):886–899.

Ericsson, K. A. and Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3):215–51.

Ernest, P. (1993). *The Philosophy of Mathematics Education.* Studies in Mathematics Education Series. Falmer Press, Taylor & Francis Inc., Bristol, PA.

Evans, J., Barston, J., and Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory and Cognition*, 11(3):295–306.

Evans, J. and Handley, S. (1999). The role of negation in conditional inference. *The Quarterly Journal of Experimental Psychology Section A*, 52(3):739–769.

Feynman, R., Leighton, R., and Hutchings, E. (1997). *"Surely you're joking, Mr. Feynman!": Adventures of a curious character.* WW Norton & Company, London.

Field, A. (2009). *Discovering statistics using SPSS.* Sage publications, London, third edition.

Fischbein, E. (1982). Intuition and proof. *For the Learning of Mathematics*, 3(2):9–24.

Fischer, G. (1981). On the existence and uniqueness of maximum-likelihood estimates in the rasch model. *Psychometrika*, 46(1):59–77.

Fischer, G. and Molenaar, I. (1995). *Rasch models: Foundations, recent developments and applications.* Springer-Verlag, New York, NY.

Fleming, W. G. (1986). The interview: A neglected issue in research on student learning. *Higher Education*, 15(5):547–63.

Frank, M. (1988). Problem Solving and Mathematical Beliefs. *Arithmetic Teacher*, 35(5):32–34.

Garofalo, J. (1989). Beliefs and their influence on mathematical performance. *Mathematics Teacher*, 82(7):502–505.

Gelbaum, B. and Olmsted, J. (2003). *Counterexamples in analysis.* Courier Dover Publications, Mineola, NY.

Ginsburg, H. (1981). The clinical interview in psychological research on mathematical thinking: Aims, rationales, techniques. *For the Learning of Mathematics*, 1(3):4–11.

Glas, C. and Verhelst, N. (1995). Testing the rasch model. In Fischer, G. and Molenaar, I., editors, *Rasch models: Foundations, recent developments and applications*, chapter 5, pages 69–95. Springer-Verlag, New York.

Glaser, B. (1992). *Basics of grounded theory.* Sociology Press, California.

Glaser, B. and Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Aldine de Gruyter, Hawthorne, New York 10532.

Goldenberg, P. and Mason, J. (2008). Shedding light on and with example spaces. *Educational Studies in Mathematics*, 69(2):183–194.

Goldin, G. (1997). Observing mathematical problem solving through task-based interviews. *Journal for Research in Mathematics Education. Monograph*, 9:40–62.

Green, K. (1996). Dimensional analyses of complex data. *Structural Equation Modeling: A Multidisciplinary Journal*, 3(1):50–61.

Green, P. (2005). A rigorous journey into phenomenography: From a naturalistic inquirer standpoint. In Bowden, J. and Green, P., editors, *Doing developmental phenomenography*, pages 32–46. RMIT University Press, Melbourne, Australia.

Grimbeek, P. and Nisbet, S. (2006). Surveying primary teachers about compulsory numeracy testing: Combining factor analysis with rasch analysis. *Mathematics Education Research Journal*, 18(2):27.

Gueudet, G. (2008). Investigating the secondary–tertiary transition. *Educational Studies in Mathematics*, 67(3):237–254.

Haines, C. and Crouch, R. (2001). Recognizing constructs within mathematical modelling. *Teaching Mathematics and its Applications*, 20(3):129–138.

Halmos, P. (1985). *I want to be a mathematician.* Springer-Verlag, New York, NY.

Hardy, G. (1967). *A mathematician's apology.* Cambridge Univ Press, Cambridge, UK.

Harel, G. and Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. In Kaput, J., Schoenfeld, A., and Dubinsky, E., editors, *Research in collegiate mathematics education III*, pages 234–283. American Mathematical Society, Providence, RI.

Harel, G. and Sowder, L. (2005). Advanced mathematical-thinking at any age: Its nature and its development. *Mathematical Thinking and Learning*, 7(1):27–50.

Hasselgren, B. and Beach, D. (1997). Phenomenography: A "good-for-nothing brother" of phenomenology? Outline of an analysis. *Higher Education Research and Development*, 16(2):191–202.

Hazzan, O. and Zazkis, R. (1999). A perspective on "give an example" tasks as opportunities to construct links among mathematical concepts. *Focus on Learning Problems in Mathematics*, 21(4):1–14.

Holton, D., editor (2001). *The Teaching and Learning of Mathematics at University Level: An ICMI study.* Kluwer Academic Publishers, Netherlands.

Hong, Y. Y., Kerr, S., Klymchuk, S., McHardy, J., Murphy, P., Spencer, S., Thomas, M. O. J., and Watson, P. (2009). A comparison of teacher and lecturer perspectives on the transition from secondary to tertiary mathematics education. *International Journal of Mathematical Education in Science and Technology*, 40(7):877–889.

Housman, D. and Porter, M. (2003). Proof schemes and learning strategies of above-average mathematics students. *Educational Studies in Mathematics*, 53(2):139–158.

Iannone, P., Inglis, M., Mejia-Ramos, J., Siemons, J., and Weber, K. (2009). How do undergraduate students generate examples of mathematical concepts? In Tzekaki, M., Kaldrimidou, M., and Sakonidis, C., editors, *Proceedings of the 33rd Conference of the International Group for the Psychology of Mathematics Education*, volume 3, pages 217–24. PME, Thessaloniki, Greece.

Iannone, P., Inglis, M., Mejía-Ramos, J., Simpson, A., and Weber, K. (in press). Does generating examples aid proof production? *Educational Studies in Mathematics.*

Jahnke, H. (2008). Theorems that admit exceptions, including a remark on Toulmin. *ZDM*, 40(3):363–371.

Kaiser, H. (1970). A second generation little jiffy. *Psychometrika*, 35(4):401–415.

Kieren, T. (2000). Dichotomies or binoculars: Reflections on the papers by Steffe and Thompson and by Lerman. *Journal for Research in Mathematics Education*, 31(2):228–233.

Kirshner, D. and Awtry, T. (2004). Visual salience of algebraic transformations. *Journal for Research in Mathematics Education*, 35(4):224–257.

Kline, T. (2005). *Psychological Testing: A practical approach to design and evaluation.* Sage publications, London.

Koichu, B. and Harel, G. (2007). Triadic interaction in clinical task-based interviews with mathematics teachers. *Educational Studies in Mathematics*, 65(3):349–365.

Larsson, J. and Holström, I. (2007). Phenomenographic or phenomenological analysis: Does it matter? Examples from a study on anaesthesiologists' work. *International Journal of Qualitative Studies on Health and Well-Being*, 2(1):55–64.

Lawson, D. M. (2006). Applying the item response theory to classroom examinations. *Journal of Manipulative and Physiological Therapeutics*, 29(5):393–397.

Leow, R. and Morgan-Short, K. (2004). To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. *Studies in second language acquisition*, 26(01):35–57.

Lerman, S. (1996). Intersubjectivity in mathematics learning: A challenge to the radical constructivist paradigm? *Journal for Research in Mathematics Education*, 27(2):133–50.

Lerman, S. (2000). A case of interpretations of "social": A response to Steffe and Thompson. *Journal for Research in Mathematics Education*, 31(2):210–27.

Lester, F. (2005). On the theoretical, conceptual, and philosophical foundations for research in mathematics education. *ZDM*, 37(6):457–467.

Lewins, A. and Silver, C. (2007). *Using software in qualitative research: A step-by-step guide*. Sage Publications Ltd, London.

Linacre, J. (2009). *Winsteps (Version 3.68.2)[Computer Software]*. Winsteps.com, Beaverton, Oregon.

Lithner, J. (2008). A research framework for creative and imitative reasoning. *Educational Studies in Mathematics*, 67(3):255–276.

Lybeck, L. (1981). *Archimedes in the classroom. A subject-didactic narrative [Abstract]*. PhD thesis, Acta Universitatis Gothoburgensis, Göteborg.

Manly, B. (2005). *Multivariate statistical methods: A primer*. Chapman & Hall/CRC, Boca Raton, FL.

Marton, F. (1975). On non-verbatim learning: 1. Level of processing and level of outcome. *Scandinavian Journal of Psychology*, 16(1):273–279.

Marton, F. (1981). Phenomenography: Describing conceptions of the world around us. *Instructional Science*, 10(2):177–200.

Marton, F. and Booth, S. (1997). *Learning and Awareness*. Lawrence Erlbaum Associates, Mahwah, NJ.

Marton, F. and Säljö, R. (1976). On qualitative differences in learning: I. outcome and process. *British Journal of Educational Psychology*, 46(1):4–11.

Marton, F. and Säljö, R. (2005). Approaches to learning. In Marton, F., Hounsell, D., and Entwistle, N., editors, *The Experience of Learning: Implications for teaching and studying in higher education*, chapter 3, pages 39–58. Edinburgh, University of Edinburgh, Centre for Teaching, Learning and Assessment, 3rd (internet) edition.

Marton, F. and Svensson, L. (1979). Conceptions of research in student learning. *Higher Education*, 8(4):471–486.

Marton, F. and Tsui, A. B. M. (2004). *Classroom discourse and the space of learning*. Lawrence Erlbaum Associates, Mahwah, NJ.

Mason, J. (in press). Phenomenology of example construction. *ZDM*, pages 1–10.

Mason, J. and Klymchuk, S. (2009). *Using counter-examples in calculus*. Imperial College Press, London.

Mason, J. and Pimm, D. (1984). Generic examples: Seeing the general in the particular. *Educational Studies in Mathematics*, 15(3):277–289.

Maxwell, J. (1992). Understanding and validity in qualitative research. *Harvard Educational Review*, 62(3):279–300.

Meehan, M. (2002). Students meeting advanced mathematics for the first time: Can mathematics education research help? *Irish Mathematical Society Bulletin*, 49:71–82.

Meehan, M. (2007). Student generated examples and the transition to advanced mathematical thinking. In Pitta, D., Philippou, P., and Philippou, G., editors, *Proceedings of CERME 5*, pages 2349–2358, Larnaca, Cyprus.

Michener, E. (1978). Understanding understanding mathematics. *Cognitive Science: A Multidisciplinary Journal*, 2(4):361–383.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2):81–97.

Misailidou, C. and Williams, J. (2003). Diagnostic assessment of children's proportional reasoning. *The Journal of Mathematical Behavior*, 22(3):335–368.

Monaghan, J. (1991). Problems with the language of limits. *For the Learning of Mathematics*, 11(3):20–24.

Moore, R. (1994). Making the transition to formal proof. *Educational Studies in Mathematics*, 27(3):249–266.

Moreno, R. (2006). When worked examples don't work: Is cognitive load theory at an impasse? *Learning and Instruction*, 16(2):170.

Morrow, M. (2004). Calculus students' views of justification and proof in mathematics. *PRIMUS*, 14(2):104–126.

Muhr, T. (1991). ATLAS.ti: A prototype for the support of text interpretation. *Qualitative Sociology*, 14(4):349–371.

Muhr, T. (2010). *ATLAS.ti (Version 6.1.16)[Computer Software]*. Scientific Software Development, Berlin.

Nardi, E. and Iannone, P. (2006). *How to prove it: a brief guide for teaching proof to year 1 mathematics undergraduate*. University of East Anglia, Centre for Applied Research in Education, Norwich, UK.

National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics*. NCTM, Reston, VA.

Neuman, D. (1997). Chapter 5: Phenomenography: Exploring the roots of numeracy. *Journal for Research in Mathematics Education. Monograph*, 9:63–177.

Nisbett, R. and Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3):231–259.

Ochs, E. (1979). Transcription as theory. In Ochs, E. and Schieffelin, B., editors, *Developmental pragmatics*, chapter 3, pages 43–72. Academic Press.

Oliver, D. G., Serovich, J. M., and Mason, T. L. (2005). Constraints and opportunities with interview transcription: Towards reflection in qualitative research. *Social Forces*, 84(2):1273–1289.

Patton, M. (2002). *Qualitative research and evaluation methods*. Sage Publications, Inc, London.

Perry, W. (1988). Different worlds in the same classroom. In Ramsden, P., editor, *Improving Learning: New Perspectives.*, pages 145–161. Kogan Page, London.

Piaget, J. (1950). *The psychology of intelligence.* Routledge, London, NY.

Poland, B. (2001). Transcription quality. In Gubrium, J. and Holstein, J., editors, *Handbook of interview research: Context and method*, chapter 30, pages 629–649. Sage publications.

Pring, R. (2000). *Philosophy of educational research.* Continuum, London.

Ramsden, P. (1984). The context of learning. In Marton, F., Hounsell, D., and Entwistle, N., editors, *The experience of learning*, pages 144–164. Scottish Academic Press Ltd, Edinburgh.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests.* The University of Chicago Press, Chicago, IL, expanded edition.

Recio, A. M. and Godino, J. D. (2001). Institutional and personal meanings of mathematical proof. *Educational Studies in Mathematics*, 48(1):83–99.

Reise, S. (2000). Using multilevel logistic regression to evaluate person-fit in IRT models. *Multivariate Behavioral Research*, 35(4):543–568.

Rencher, A. (2002). *Methods of multivariate analysis.* Wiley-Interscience, New York, NY.

Renström, L., Andersson, B., and Marton, F. (1990). Students conceptions of matter. *Journal of Educational Psychology*, 82(3):555–569.

Richardson, J. T. E. (1999). The concepts and methods of phenomenographic research. *Review of Educational Research*, 69(1):53–82.

Robert, A. and Speer, N. (2001). Research on the teaching and learning of calculus/elementary analysis. In Holton, D., editor, *The Teaching and Learning of Mathematics at University Level*, volume 7, pages 283–299. Kluwer Academic Publishers, Netherlands.

Rowlands, S., Graham, T., and Berry, J. (2001). An objectivist critique of relativism in mathematics education. *Science & Education*, 10(3):215–241.

Runesson, U. (1999). Teaching as constituting a space of variation. In *Paper presented at the 8th EARLI conference*, Göteborg, Sweden.

Runesson, U. (2006). What is it possible to learn? On variation as a necessary condition for learning. *Scandinavian Journal of Educational Research*, 50(4):397–410.

Ryan, J. and McCrae, B. (2006). Assessing pre-service teachers' mathematics subject knowledge. *Mathematics teacher education and development*, 7:72–89.

Ryan, J. and Williams, J. (2007). Mathsmaps for diagnostic assessment with pre-service teachers: Stories of mathematical knowledge. *Research in Mathematics Education*, 9(1):95–109.

Sá, W., West, R., and Stanovich, K. (1999). The domain specificity and generality of belief bias in reasoning and judgment. *Journal of Educational Psychology*, 91(3):497–510.

Säljö, R. (1997). Talk as data and practice: A critical look at phenomenographic inquiry and the appeal to experience. *Higher Education Research and Development*, 16(2):173–190.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4):350–353.

Schoenfeld, A. (2004). The math wars. *Educational Policy*, 18(1):253–286.

Schwarzenberger, R. and Tall, D. (1978). Conflicts in the learning of real numbers and limits. *Mathematics Teaching*, 82:44–49.

Selden, A. and Selden, J. (1998). *The role of examples in learning mathematics.* MAA online. Retrieved February 21, 2011 from `http://www.maa.org/t_and_l/sampler/rs_5.html`.

Selden, A. and Selden, J. (1999). The role of logic in the validation of mathematical proofs. *Tennessee Technological University Mathematics Department Technical Report*, 1:1–13.

Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics*, 22(1):1–36.

Sierpinska, A. (1987). Humanities students and epistemological obstacles related to limits. *Educational Studies in Mathematics*, 18(4):371–397.

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching*, 77:20–26.

Smith, A. (2004). *Making mathematics count: The report of professor Adrian Smith's inquiry into post-14 mathematics education.* The Stationery Office, London.

Smith, R. (1996). A comparison of methods for determining dimensionality in rasch measurement. *Structural Equation Modeling: A Multidisciplinary Journal*, 3(1):25–40.

Sørensen, H. (2005). Exceptions and counterexamples: Understanding Abel's comment on Cauchy's theorem. *Historia Mathematica*, 32(4):453–480.

SPSS (2010). *SPSS for Windows (Version 17.0)[Computer Software].* SPSS Inc., Chicago.

Spyrou, P. and Zagorianakos, A. (2009). An exploration of mathematics students' distinguishing between function and arbitrary relation. *Proceedings of the British Society for Research into Learning Mathematics*, 29(3):85–90.

Spyrou, P. and Zagorianakos, A. (2010). Greek students' understandings of the distinction between function and relation. *Research in Mathematics Education*, 12(2):163–164.

Stacey, K. and Steinle, V. (2006). A case of the inapplicability of the rasch model: Mapping conceptual learning. *Mathematics Education Research Journal*, 18(2):77–92.

Steffe, L. P. and Thompson, P. W. (2000). Interaction or intersubjectivity? A reply to Lerman. *Journal for Research in Mathematics Education*, 31(2):191–209.

Stewart, S. and Thomas, M. (2010). Student learning of basis, span and linear independence in linear algebra. *International Journal of Mathematical Education in Science and Technology*, 41(2):173–188.

Strauss, A. L. and Corbin, J. (1988). *Basics of qualitative research: Techniques and procedures for developing grounded theory.* Sage publications, London.

Strauss, A. L. and Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory.* Sage publications, London.

Stylianides, A. and Al-Murani, T. (2010). Can a proof and a counterexample coexist? Students' conceptions about the relationship between proof and refutation. *Research in Mathematics Education*, 12(1):21–36.

Svensson, L. (1977). On qualitative differences in learning: III. Study skill and learning. *British Journal of Educational Psychology*, 47(3):233–243.

Svensson, L. (1997). Theoretical foundations of phenomenography. *Higher Education Research and Development*, 16(2):159–171.

Sweller, J. and Cooper, G. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, 2(1):59–89.

Tall, D. (1986). Constructing the concept image of a tangent. In *Proceedings of the 11th Conference of the International Group for the Psychology of Mathematics Education*, volume 3, pages 69–75, Montréal, Canada.

Tall, D. (1988). Concept image and concept definition. In de Lange, J. and Doorman, M., editors, *Senior Secondary Mathematics Education*, 37–41, Utrecht. OW&OC.

Tall, D., editor (1991a). *Advanced Mathematical Thinking.* Kluwer Academic Publishers, Netherlands.

Tall, D. (1991b). The psychology of advanced mathematical thinking. In Tall, D., editor, *Advanced Mathematical Thinking*, Mathematics Education Library, chapter 1, pages 3–21. Kluwer Academic Publishers, Netherlands.

Tall, D. (2003). Using technology to support an embodied approach to learning concepts in mathematics. *Historia e tecnologia no Ensino da Matemática*, 1:1–28.

Tall, D. (2004). Thinking through three worlds of mathematics. In Høines, M. and Fuglestad, A., editors, *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education*, volume 4, pages 281–288. PME, Bergen, Norway.

Tall, D. and Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, 12(2):151–69.

Tennant, A. and Pallant, J. (2006). Unidimensionality matters. *Rasch Measurement Transactions*, 20(1):1048–1051.

Tsamir, P., Tirosh, D., and Levenson, E. (2008). Intuitive nonexamples: The case of triangles. *Educational Studies in Mathematics*, 69(2):81–95.

Uhlig, F. (2002). The role of proof in comprehending and teaching elementary linear algebra. *Educational Studies in Mathematics*, 50(3):335–346.

van Zyl, J., Neudecker, H., and Nel, D. (2000). On the distribution of the maximum likelihood estimator of Cronbach's alpha. *Psychometrika*, 65(3):271–280.

Vermunt, J. (1996). Metacognitive, cognitive and affective aspects of learning styles and strategies: A phenomenographic analysis. *Higher education*, 31(1):25–50.

Vinner, S. (1976). The naive concept of definition in mathematics. *Educational Studies in Mathematics*, 7(4):413–429.

Vinner, S. (1983). Concept definition, concept image and the notion of function. *International Journal of Mathematical Education in Science and Technology*, 14(3):293–305.

Vinner, S. (1991). The role of definitions in the teaching and learning of mathematics. In Tall, D., editor, *Advanced Mathematical Thinking*, Mathematics Education Library, chapter 5, pages 65–81. Kluwer Academic Publishers, Netherlands.

Vinner, S. (1997). The pseudo-conceptual and the pseudo-analytical thought processes in mathematics learning. *Educational Studies in Mathematics*, 34(2):97–129.

von Glasersfeld, E. (1984). An introduction to radical constructivism. In Watzlawick, P., editor, *The invented reality: How do we know what we believe we know?*, pages 17–40. Norton, New York, NY.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press, Cambridge, MA.

Wang, W. and Chen, C. (2005). Item parameter recovery, standard error estimates, and fit statistics of the winsteps program for the family of rasch models. *Educational and Psychological Measurement*, 65(3):376–404.

Watson, A. and Chick, H. (in press). Qualities of examples in learning and teaching. *ZDM*, pages 1–12.

Watson, A. and Mason, J. (2002). Extending example spaces as a learning/teaching strategy in mathematics. In Cockburn, A. and Nardi, E., editors, *Proceedings of the 26th Conference of the International Group for the Psychology of Mathematics Education*, volume 4, pages 377–385.

Watson, A. and Mason, J. (2004). The exercise as mathematical object: Dimensions of possible variation in practice. *Research into Learning Mathematics*, 24(2):107.

Watson, A. and Mason, J. (2005). *Mathematics as a constructive activity: Learners generating examples*. Lawrence Erlbaum Associates, Mahwah, NJ.

Watson, A. and Mason, J. (2006). Seeing an exercise as a single mathematical object: Using variation to structure sense-making. *Mathematical Thinking and Learning*, 8(2):91–111.

Watson, A. and Shipman, S. (2008). Using learner generated examples to introduce new concepts. *Educational Studies in Mathematics*, 69(2):97–109.

Watson, J., Kelly, B., and Izard, J. (2006). A longitudinal study of student a longitudinal study of student understanding of chance and data. *Mathematics Education Research Journal*, 18(2):40–55.

Waugh, R. F. (1998). The course experience questionnaire: A Rasch measurement model analysis. *Higher Education Research and Development*, 17(1):45–64.

Weber, K. (2002). Beyond proving and explaining: Proofs that justify the use of definitions and axiomatic structures and proofs that illustrate technique. *For the Learning of Mathematics*, 22(3):14–17.

Weber, K. (2008). The role of affect in learning real analysis: A case study. *Research in Mathematics Education*, 10(1):71–85.

Weber, K. (2009). How syntactic reasoners can develop understanding, evaluate conjectures, and generate counterexamples in advanced mathematics. *Journal of Mathematical Behavior*, 28(2-3):200–208.

Weber, K. (2010). Mathematics majors' perceptions of conviction, validity, and proof. *Mathematical Thinking and Learning*, 12(4):306–336.

Weber, K., Brophy, A., and Lin, K. (2008). Learning advanced mathematical concepts by reading text. In *Proceedings of the 11th conference for research in undergraduate mathematics education*, San Diego, CA.

Weber, K. and Mejia-Ramos, J. (in press). Why and how mathematicians read proofs: an exploratory study. *Educational Studies in Mathematics*.

Williams, S. (1991). Models of limit held by college calculus students. *Journal for Research in Mathematics Education*, 22(2):219–236.

Wilson, E. and Hilferty, M. (1931). The distribution of chi-square. *Proceedings of the National Academy of Sciences of the United States of America*, 17(12):684.

Wood, L. (2001). The secondary-tertiary interface. In Holton, D., editor, *The Teaching and Learning of Mathematics at University Level*, pages 87–98. Kluwer Academic Publishers, Netherlands.

Wright, B. D. and Masters, G. N. (1982). *Rating Scale Analysis*. Mesa Press, Chicago, IL.

Yorke, M. (2006). Student engagement: Deep, surface or strategic. In *Keynote address to the 9th Pacific Rim conference on the first year in higher education*, pages 12–14, Griffith University, Australia.

Zaslavsky, O. and Peled, I. (1996). Inhibiting factors in generating examples by mathematics teachers and student teachers: The case of binary operation. *Journal for Research in Mathematics Education*, 27(1):67–78.

Zaslavsky, O. and Shir, K. (2005). Students' conceptions of a mathematical definition. *Journal for Research in Mathematics Education*, 36(4):317–346.

Zazkis, R. and Chernoff, E. J. (2008). What makes a counterexample exemplary? *Educational Studies in Mathematics*, 68(3):195–208.

Zazkis, R. and Leikin, R. (2007). Generating examples: From pedagogical tool to a research tool. *For the Learning of Mathematics*, 27(2):15–21.

Zazkis, R. and Leikin, R. (2008). Exemplifying definitions: A case of a square. *Educational Studies in Mathematics*, 69(2):131–148.

Typesetting

This thesis has been typeset in LaTeX 2_ε using TeXShop and BibDesk. I am very grateful to those that have given their time to produce this open-source software.

The thesis template used is one modified from a freely-released version by Steve R. Gunn and Sunil Patel, available at `http://www.sunilpatel.co.uk/thesis-template/`.

Contact details

Antony Edwards
Mathematics Education Centre
Loughborough University
Loughborough
Leicestershire
LE11 3TU

a.w.edwards@lboro.ac.uk

Compiled on 30th May 2011.