Testing the Ability to Apply Mathematical Knowledge

Kai Chung Tam

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2018

ABSTRACT

Testing the Ability to Apply Mathematical Knowledge

Kai Chung Tam

Since the 1960s, the advocacy of *teaching mathematics so as to be useful* is not without hindrance in school curricula, partly due to the lack of appropriate assessment tools. Practical approaches have been accumulating quickly, but researchers showed that they are not satisfactory in testing students' ability to apply mathematical knowledge, be they "word problems" in school textbooks, national tests, or large-scale international assessments. To understand the causes behind the dissatisfaction, there is a need to reveal (1) the theories that are used in the test designs, and (2) what the actual assessments are in various curricula. This motive leads to the purpose of the current study, which is to identify empirically consistent theories about students' ability to apply; the results can be organized as a framework to analyze assessment tools such as PISA, as well as various curricular materials.

Based on the current theories, a framework of assessment analysis is created in order to study the coverage of modeling steps of public assessment items. This study finds that, though many education systems have claims of introducing modeling and application into their curricula, high-stake assessments mostly involve a small fraction of the steps that are required in a full modeling cycle. It furthers an earlier result that certain textbooks, though claiming the importance of modeling, almost ignored the first and last steps of modeling. It is found in this study that public assessments are even more limited: most test items that are supposed to test students' knowledge of application involve only one or two steps of modeling. Furthermore, the tool "modeling spectrum" that is used in the analysis does not only reveal how modeling steps are covered, but can also assists educators to improve or create problems with modeling and application.

Contents

List of Figures

List of Tables

Acknowledgements

When I was getting into the life of being a doctoral student, I could not imagine how demanding the work of a dissertation is, until having been experiencing lots of struggles and challenges. I would like to thank Dr. Bruce Vogeli the first place, as he guides me to grow as a scholar, always sets high expectations, as well as provides a great deal opportunities for academic experiences. Though I came to Teachers College as a non-native English speaker, Dr. Vogeli has believed in me that I can become a successful scholar, not to mention to become fluent in English. While my doctoral study plans were significantly delayed, not only did he allow me to continue the process, but he also urged me in strong words that I must set aside other commitments to get the thesis done. I am deeply thankful about his life teaching, or otherwise I would not survive the doctoral program.

I would also like to thank Dr. Henry Pollak very much. Since 2011, he had been guiding me through the process of writing a dissertation with many ideas and encouragements. There were times when I was dull and slow, not being able to keep up, and even "insulated" myself, but more than once he re-accepted my running back along the track. Thank you for doing so many things for a retarded student, and thank you for always providing helpful advice, and suggestions on my dissertation from biggest ideas to smallest changes. I could not have finished the thesis without your patient help.

I would like to say thank you to Dr. Karp, who is always supportive in answering my questions about the thesis itself and the academia. Thank you Dr. Grabois, Dr. Anderson, and Dr. Gallagher, for your willing to sit the defense committee and provide valuable advises. There are also a lot of people in Teachers College that I owe my thanks to: Dr. Smith, Dr. Nii Nartey, Dr. Hantian Wu, Dr. Ronny Leong, Dr. Mark Causapin, Dr. Janice Robinson, Ms. Juliana, Mr. Gary Ardan, Mr. Russell Gulizia. With all of their support, I was learning and growing in the always welcoming and challenging enviornment of TC and in the Math Education department. I also deeply appreciate Professor Wen, who is my adviser in my Undergraduate and Masters study, since he cares about his students even when they are graduated.

I am indebted a lot to my family, especially my parents, whom I dedicated this work to. I left from them for a higher education since 2002, nearly 16 years ago. They constantly cared about me even though I had not contributing anything to them. I hope that the completion of the thesis could be a token of comfort to them, so as to see that their caring efforts since my childhood are not in vain. My wife Deborah Cui Xin has been worrying for my study since our marriage in 2015, but she tolerates me a lot and takes perfect care of me. I also owe a lot to the brothers and sisters in my church, who constantly offer prayers to God on my well-being and my progress. Rev. Laura Lin, thank you for your guidance in all possible ways. I also thank Elder Lyna for her accommodation when I couldn't find a place to write without distractions.

All these words are not enough in expressing my deep gratitude toward these kind persons during this decade (2008-2018).

All thanks be to God, for He revealed the Truth to my heart and adopted me into His family in 2013, so that I may dedicate my life to serve Him. I hope to use what I learned in my graduate studies to contribute to the eternal good of others.

DEDICATION

I dedicate this dissertation to my honored father and mother

Wai Chong Tam

And

Pui Ching Li

for your substantial support, words of worrying, and persevering encouragement

All glory be to God through the Lord Jesus Christ

Chapter 1

INTRODUCTION

This dissertation presents current theories on the ability to apply mathematical knowledge, and how to use some of these theories to create an analytical framework which helps studying curriculum materials. The current chapter introduces the background and need for the study, states the purpose of the study and the three research questions, and briefly discusses the procedures to study these questions. Chapter 2 reviews the research literature that needs to be considered in order to clarify the relevant concepts and methodology, including how ability to apply mathematical knowledge is conceptualized, how theories are viewed and identified in mathematics education, and how *curriculum* is defined and analyzed. Chapter 3 presents the details of the operations performed in the procedures for each research question. The analyses and the results of each research question are discussed in Chapter 4, for research question 1, and Chapter 5, for research questions 2 and 3. Finally, the conclusion chapter summarizes the findings, points out the limitations and possible contributions of the study, and some directions of further research.

Background and Need for the Study

"Where am I going to use this?"

This anecdotal quotation is popular between school students and teachers, where the pronoun *this* most likely refers to the subject of mathematics[1].

[1]By looking up in an Internet search engine.

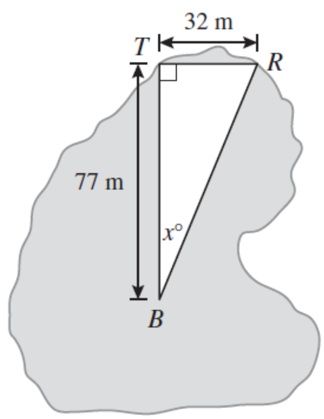
Indeed, much effort has been made in order to show the fact to students and teachers that mathematics is useful, and it helps us to understand more about the "outside world". For instance, mathematical modeling and application has been promoted by educators and mathematicians since the 1960s (Freudenthal, 1968; Klamkin, 1968; Pollak, 1969). A starting point can be chosen as the remarkable colloquium *How to Teach Mathematics so as to be Useful* held in 1968, the proceeding of which was published as the first season of the *Educational Studies in Mathematics*[2]. Through numerous research studies and curriculum projects, many educators have been demonstrating how mathematics can be useful in the real world (e.g., Maddern & Crust, 1989; Garfunkel & Pollak, 1998; Maasz & O'Donoghue, 2011); however, the anecdotal question does not show that mathematics is useless, but one's frustration about the inability to apply the mathematical knowledge he or she has learned. Indeed, despite the awareness of teaching useful mathematics, the advocacy of *teaching mathematics so as to be useful* is not without hindrance. It was observed that school mathematics curricula have not really embraced real-world application (Brown, 2001), and that students still suffer from disconnectedness between the real world and the mathematics that they are taught (Reusser & Stebler, 1997; Inoue, 2005; Depaepe, De Corte, & Verschaffel, 2014). Even though real-world context is often included in assessments, it is not rare that such context has no relation to the solution. For example, the item shown in Figure 1.1 presents a situation where a boat has to turn a certain angle in order to change its destination (Florida Department of Education [FDOE], 2012, p. 76), but the problem can also be solved solely by the right triangle that is drawn clearly.

Another item (Figure 1.2) shows a picture with several men pushing a barrel, and asked student to find the capacity of it (Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005). Since the picture did not provide any numerical value, a student would

[2] *Educational Studies in Mathematics*, Vol. 1, No. 1/2, 1968.

Figure 1.1: An Introductory Example of an Assessment Task Which is Solvable Without Understanding the Context

A tackle shop and restaurant are located on the shore of a lake and are 32 meters (m) apart. A boat on the lake heading toward the tackle shop is a distance of 77 meters from the tackle shop. This situation is shown in the diagram below, where point T represents the location of the tackle shop, point R represents the location of the restaurant, and point B represents the location of the boat.



The driver of the boat wants to change direction to sail toward the restaurant. Which of the following is closest to the value of x?

★ **A.** 23
B. 25
C. 65
D. 67

The item displays a paragraph of description, a picture that represents a map, a question, and four options. The starred option is the correct response. Adapted from *Florida 2012 Geometry End-of-Course Assessment Test Item Specifications Version 2* (p. 76), by Florida Department of Education, 2012, Tallahassee: Florida Department of Education. Copyright 2012 by State of Florida, Department of State. Reprinted with permission.

have to come up with a mathematical model that is suitable for the context, and will also need to be familiar with ways to describe capacity, such as to use appropriate units. From these two examples, we can see that although an assessment task can be posed within some real-world context, it does not necessarily test students' ability to apply mathematical knowledge in that context. However, some analytic framework is needed in order to evaluate these test items in a more consistent and systematic way. Such a framework can be established after a survey of theoretical ideas about the abilty to apply mathematical knowledge.

Given that the use of mathematical knowledge is a topic that concerns many educators and

Figure 1.2: Another Introductory Example of an Assessment Task Whose Solution Requires Thinking Within Context

Quelle: © Bettmann/CORBIS

Wie viel Flüssigkeit passt ungefähr in dieses Fass? Begründe deine Antwort.

This item was written in German. It includes one picture and a line of text below it. The picture (copyright by Bettmann/CORBIS) shows several men handling a large barrel, and supposedly going to open it, as suggested by the handwritten exclaimation, "Zeit mal wieder ein Faß zu oeffnen!" (It's time again to open a barrel!) The text below the picture can be translated as "How much liquid fits into this barrel? Explain your answer." Adapted from *Bildungsstandards im Fach Mathematik für die Allgemeine Hochschulreife (Beschluss der Kultusministerkonferenz vom 18.10.2012)* (p. 16), by Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland Sekretariat and Institut zur Qualitätsentwicklung im Bildungswesen [KMK and IQB], 2015, Köln: Wolters Kluwer. Copyright 2015 by KMK Bonn und Berlin 2012.

researchers, various formulations are available in the literature on how mathematical knowledge should be applied and how students actually apply it (see Chapter 2). In this dissertation, we use *mathematical modeling and application* as the central formulation, while other formulations are interpreted by different parts of the *mathematical modeling steps* (p. 11).

Literature has been discussing the ability, skills, or competencies related to mathematical modeling and application, and the assessment of these competencies (e.g., R. A. Lesh & Lamon, 1992; Lingefjärd & Holmquist, 2005; Jurdak, 2016). It is found difficult to devise a test on mathematical modeling: not only is this difficulty related to the complexity of the subject itself, but it is also entangled with different aspects of the education system, such as the time allocated in the curriculum (Burkhardt, 2006). Peter Frejd (2013) reviewed the literature on the assessment of mathematical modeling, with a conclusion that many of the research studies are empirical, but very few articles focus on the theoretical framework. He also pointed out that very few theoretical studies were grounded on empirical evidences. Thus, this dissertation is an attempt to construct a theoretical framework for the assessment of mathematical modeling.

Moreover, most mathematics educators and curriculum standards call for more attention to modeling and application (e.g. Burkhardt, 2006; Blum, Galbraith, Henn, & Niss, 2007). Given that high-stake assessment is seen as a driving force of the curriculum, as (Burkhardt, 2001) puts it, "*What You Test Is What You Get* in the classroom", the current dissertation also applies the theoretical framework to analyze the gap between the *ideal* way to test the ability to apply and how high-stake assessments are actually designed. Therefore, as stated in the next section, the objective of this dissertation has two parts: a methodological part that *creates* a framework, and an analytical part that *applies* the framework. The purpose of research is thus stated in the following section.

Purpose and Research Questions of the Study

The purpose of this dissertation is to use current theories to establish a theoretical framework of the ability to apply mathematical knowledge, and apply the framework to conduct a comparative analysis of curriculum documents and public assessments of USA, Germany, and Hong Kong, focusing on the role of modeling and application, and how such role changed over time. Note that the first part of the purpose can-be a stand-alone project that provides a collection of available theories, and the second part puts some (but not all) of these theories into a framework so that an analysis can be practically done.

In the following, three research questions are formulated for this purpose.

RQ1. (a) What are current theories on the ability to apply mathematical knowledge, and on the testing of such ability? (b) How do these theories fit into a larger theoretical framework that helps to analyze curriculum documents and assessment materials?

RQ2. (a) How did the 2003 and 2012 Programme of International Student Assessment (PISA) assess the ability to apply mathematics? (b) How is the modeling process assessed by PISA items?

RQ3. (a) Between 1995 and 2003, what was the role of application in the curriculum documents of the United States, Germany, and Hong Kong? (b) In this period, how were their public assessments related to the modeling process? (c) Between 2004 and 2012, what was the role of modeling and application in the curriculum documents of the United States, Germany, and Hong Kong? (d) In this period, how were their public assessments related to the modeling process?

Procedures of the Study

In order to address Research Question 1, a systematic review (Cook, Mulrow, & Haynes, 1997) will be performed on relevant articles in the selected collection, including the ICTMA proceedings (16 volumes), an ICMI study (Blum et al., 2007), several periodicals of mathematics education, and some treatises or book series. The articles can be thought of as two types –

> Type A, *Theoretical*: Study that includes theoretical elements on the ability to apply mathematics, how to define such ability, or how to test it.
>
> Type B, *Empirical*: Study that includes at least one empirical study that directly relates to individuals who performs mathematical modeling or application.

Type-A articles provide existing theories to be surveyed. Type-B articles serve as a "pool" of empirical cases that may support or refute the theories aforementioned. Tsatsaroni, Lerman, and Xu (2003) offered a procedure of extracting theories from existing literature, which were applied to the selected collection. We follow Schoenfeld (2001) on the standards of examining theories, focusing on the falsifiability and the predictability. Number counts of mismatch between a theory and empirical data were recorded, but since the importance of empirical studies are not to be weighted, lower counts does not imply a better theory. A theory is seen as inconsistent with empirical study if there is an unexplained mismatch between this theory and the empirical study.

If there exists one or more theories that are not flawed, they become candidate theories. They will be checked by all the eight criteria described in Schoenfeld (2001), such as the scope of a theory. A refinement step will be done to the theories that are more or less flawed. In the refinement step, amendments of the theories are generated so as to cover the flaws, and a second round of empirical tests is applied, using the rest of them in the pool.

Theories on the processes of mathematical modeling may be quite different. For example, it could be an iteration of modeling cycles, as proposed by Blum and Leiss, or several subprocesses underlying the modeling process, as proposed by Jensen (Frejd, 2013). Nonetheless, theories generally agree that mathematical modeling cannot be assessed by a single dimension of scale (e.g., Højgaard Jensen, 2007)); there are different aspects of modeling that are under concern, and different theories hypothesize different structures that organize these aspects.

A major assumption in Research Question 1 is the existence of an ability to apply mathematical knowledge, and this ability is different from but closely related to the traditionally tested abilities of doing mathematics. This assumption also applies to all Research Questions.

The Research Question 2 mentioned PISA, which is an international assessment program conducted by the Organisation for Economic Co-operation and Development (OECD), investigating 15-year-old students' reading, mathematics, science, and problem-solving competencies in various contexts. PISA assessments have been given once every three years since 2000, while 2003 and 2012 are the years when mathematics was the main focus. PISA tests students' "mathematical literacy" (Organisation for Economic Co-operation and Development [OECD], 2004), which means whether students can "extrapolate" from the mathematics they learned from school to application problems in real contexts. Therefore, one may expect that modeling and application plays an important role in PISA's mathematics assessment. Research Question 2 is to investigate this statement. To answer RQ2(a), official documents of PISA's assessment framework will be reviewed through the framework constructed as a result of RQ1, checking how modeling and application is involved, and in what ways. To answer RQ2(b), some published items of PISA will be analyzed qualitatively by the same framework.

The objective of Research Question 3 is to understand and compare the efforts made by different educational systems concerning the importance of mathematical modeling and application. A survey on the role of modeling will be done on a sample of official documents of USA, Germany, and Hong Kong, within two subsequent periods (1995-2003, and 2004-2012). The choice of the three education systems and the two periods is based on the following: (1) they all participated in PISA 2003 and 2012, allowing for further research relating to their PISA performance; (2) The author will be able to obtain the curriculum materials and understand the languages.

In the curriculum analysis, we take the broader definition of curriculum that includes the ideal, the intended, the implemented, the attained, the assessed, and the expected curriculum (see p. 56). This study emphasizes the intended and the assessed curriculum, with the focus on modeling and application. Each country may have one or more education systems; for each system, curriculum documents and public assessments will be selected from each of the two designated periods, so as to analyze the role of modeling, and how assessment tasks are related to each modeling step. The results of the analysis will be recorded into a database.

Chapter 2

REVIEW OF LITERATURE

Starting from only a few basic principles, one can use a mathematical model to explain a experimental phenomenon, make prediction, or even reorganize a science. The possibility of application has been perceived to be the glory of mathematics (Fisher, 1930, p. 228). One would think that education, bearing a purpose to cultivate future citizens who inherit intellectual assets, should equip students so that they can share this glory – with mathematical knowledge and ways to apply it. In fact, it is natural that application raises interest in many fields. Employers need to know if their employees know sufficient mathematics for the job, and train them purposely if not so. Cognitive psychologists analyze how transfer of knowledge (of mathematics) happens within an individual. Applied mathematicians want to know what mathematical structure one can discover or apply in various technical disciplines. Philosophers of Science ponder why mathematics is so unreasonably effective in the natural world (Wigner, 1960). In the field of mathematics education or education in general, there are also important strands of research that are related to the application of mathematical knowledge, such as mathematical modeling, realistic mathematics, and mathematical literacy. This chapter will review these research strands.

The topic is therefore so broad that a couple of keyword searches may not be able to capture enough significant information from the literature. It becomes even worse when application is also referred to by phrases like *the use of mathematics*, or *mathematical problem solving in context*. Thus, a framework of some sort is needed to serve as a common set of vocabulary to discuss different perspectives and to synthesize the literature on application and other related concepts. We use a framework summarized from the analysis by (Pollak, 2003) with minor modifications, where eight steps of mathematical modeling were cited from (Pollak, 1997). We summarize them with labeling phrases for later use:

Eight Steps of Mathematical Modeling:

(1) *Identification of questions*: identify things in the real world that we want to ask about.
(2) *Conceptual modeling*: identify important objects and important relationship in the real world that are relevant to the posed questions.
(3) *Idealization*: make decisions to ignore some of the objects or their interrelations, and to make assumptions about them.
(4) *Mathematization*, or translation to mathematics: translate the idealized situation into mathematical language; a mathematical model is then obtained, and the idealized questions become mathematical questions.
(5) *Field Choice*: "identify the field or fields of mathematics that are relevant to the model and bring to bear our instincts and knowledge about these fields."
(6) *Mathematical solution*: "We use mathematical methods and insights to get results. Out of this step may come new techniques, interesting examples, solutions, approximations, theorems, algorithms."
(7) *Interpretation*, or translation to real world: translate all results back to the real world.
(8) *Verification and evaluation*: Check if the results sound reasonable in the real world so as to evaluate the quality of the model. If the model is wrong or not satisfactory, find the cause of it and start over.

As (Pollak, 2003) notes, although these are called "steps" of modeling, the actual thinking process is not necessarily in that order. Furthermore, steps 1-8 serve one but not the only definition of mathematical modeling, therefore we would rather say that the eight steps constitute a *framework*, by which we can discuss and analyze other definitions or concepts related to application. As an exemplar, (Pollak, 2003) identified that the common usage of *application of mathematics* as steps 4-7, *word problem* as steps 4 and 6, and *problem formulation* as steps 1-4. A further example shows how this framework serves as a common language in discussing application related concepts: in a typical "application problem" from a pre-calculus or calculus textbook, the idealized situation is usually given, and the kind of mathematics can be inferred from the section title. Therefore, only steps 4 and 6 are involved in the thinking process — in Pollak's definition such "application problem" is indeed equivalent to a word problem (in his definition), but does not cover all steps of an *application*.

In this literature review, we also distinguish between the *application of mathematics* and the *application of mathematical knowledge*. “Application of mathematics” refers to how the discipline of mathematics, including its particular ideas, concepts, facts and analysis are useful in other fields of study or in the outside world. The individual’s knowledge of mathematics is not quite important. On the other hand, “application of mathematical knowledge” concerns how an individual or a group of individuals apply their knowledge of mathematics, which is limited by their own scopes or abilities. It is in the latter sense that we use the word *application* to refer to the use of (individual’s) mathematical knowledge, unless otherwise clarified. To keep the review relevant, we cannot praise the “glory of mathematics” directly, but we try to understand how limited individuals “share”, or “reflect” it.

The central concern of this thesis is the ability to apply mathematical knowledge. In the following, the first section “Difficulty of Solving Word Problems” presents the pivotal works that observe and explain how students applying their mathematical knowledge in right or wrong ways. The second section “Ability to Apply in Terms of Modeling” discusses how mathematical modeling is used as a framework to understand various formulations of applying mathematical knowlege. The eight steps of mathematical modeling will be used as a common framework of depicting the relevant activities to solve real-world problems using mathematics, and the ability to apply mathematical knowledge is conceptualized as *modeling competence*. The third section “Alternative Notions of the Ability to Apply” provides other concepts that are also related to the ability to apply mathematical knowledge. In the last section “Methodological Background: Theories in Mathematics Education” explains some background knowledge that supports the research procedures in Chapter 3.

Difficulty of Solving Word Problems

Word problems play an important but not always positive role in the teaching and learning of mathematics. Sometimes they are simply treated as "applications"; indeed, they are often treated as a way to test students' mastery of mathematical operations and the ability to apply them in context, but it is notoriously known that many students could not solve them. Nonetheless, there are cases where students are better at solving problems expressed in words than those in symbols, while the necessary mathematical contents to solve the problems are similar.

Word problems are also known as verbal problems or story problems. In Gerofsky's dissertation (2001), she observed that word problems have been existing since the ancient time on Babylonian clay tablets and Rhind Papyrus, and the modern ones are usually seen in the ending part of chapter exercises. The literal form of word problems has not changed much till today. It consists of three components: a description of a "story" — usually irrelevant to the solution — the given information, and a question (Gerofsky, 2001, p. 29). In both of the ancient documents and the more modern textbooks, word problems were written in a real-word context but they were mostly identified as "artificial" and without practical relevance. Gerofsky suggested that a possible purpose of these word problems was to deliver the mathematical content while the modern concepts and symbols were not yet available to the ancient writers, or as a pedagogical tool in teaching the mathematical content involved. They are not likely to be directly related to real-life situations, nor was it the problem writers' intent to treat word problems as learning materials about practical applications (Gerofsky, 1997).

Pollak (1969) also points out that school students were mainly educated through word problems as an experience of doing application, and, unfortunately, many word problems pretend to be applications by using words from daily life or other disciplines, while word problems from genuine applications are very rare except for some simple daily applications. Word problems were often created by coating the desired mathematical content by some story (Gerofsky, 2001, p. 167); in other words, they are *mathematical problems looking for words* (Gerofsky, 2001, p. 112).

Despite their impractical nature, many such word problems are used as assessment tool of students' ability to apply. At the same time, it Is well-known that students experience much trouble in solving word problems. The infamous "age-of-the-captain" problem opened up a lot of discussions. When students were asked, "A captain owns 26 sheep and 10 goats. How old is the captain?", many answered "36" or some other definite number that is produced by one of the four basic arithmetic operations of 26 and 10, without casting a doubt about the problem itself (Verschaffel, Greer, & de Corte, 2000, pp. 3-4). In an empirical experiment, Verschaffel, de Corte, and Lasure (1994) composed ten pairs of problems, where each pair consisted of:

> — a standard problem (S problem) asking for the straightforward application of one or more arithmetic operations with the given numbers (e.g., "Steve has bought 5 planks of 2 m each. How many planks of 1 m can he saw out of these planks?"), and
> — a parallel problem (P problem) in which the mathematical modeling assumptions are problematic, at least if one seriously takes into account the realities of the context called up by the problem statement (e.g., "Steve has bought 4 planks of 2.5 m each. How many planks of 1 m can be saw out of these planks?"). (Verschaffel et al., 1994, p. 275)

The "S problem" and "P problem" share similar or same contextual situation, but they were designed so that the S problems can be solved correctly by merely picking a correct operation from the basic four, even without any realistic concern, whereas the P problems require the problem solver to make sense of the problem situation. The main focus of the study is whether students are able to give realistic answers to the P problems. In the example cited, a realistic answer is 8 since 2 planks of 1 m can be saw out of each plank of 2.5 m. Out of the 75 5th graders in Belgium tested, 92% of them responded realistically to no more than 40% of the P problems (Verschaffel et al., 1994). The results have been replicated successfully in Switzerland (Reusser & Stebler, 1997). Apparently, such students are not aware of the reality associated with the word problem. To explain this phenomenon, Verschaffel et al. (2000) describe a good model and a defective model of word problem solving. In the good model, a student reads the problem text and understands it, achieving a situation model, then form a mathematical model using appropriate operations; after the mathematical model is solved, the student will interpret the results in the real situation and see if it makes sense. Only when the results make sense, the student reports it orally or in a written form. In the defective model, the understanding of the text is bypassed, while the student chooses an operation directly by superficial decisions; after some mathematical work, the student reports the results, without making sense of them in the given situation. These two models can also be described in terms of the modeling framework. The good model consists of steps 2 to 7, but the defective model only involves steps 4 to 6, ignoring the real situation. Such apparent ignorance of reality is referred to as *suspension of sense-making* (Verschaffel et al., 2000, p. 6).

The ignorance of reality in this case cannot be attributed to cognitive incompetence. Instead, researchers suspected that schooling could be the most responsible factor (Verschaffel et al., 2000; Schoenfeld, 1991, p. 316; Greer, 1997). Indeed, in addition to the replication study, Reusser and Stebler (1997) also find that it varies as a function of schooling type. They articulated some "norms" that students and teachers usually follow when dealing with word problems, and these norms discourages the process of sense making; for example, "do not question the correctness or completeness of problems", "if you do not understand a problem, look at key words, or at previously solved problems, in order to determine a mathematical operation". On the other hand, in an oft-cited treatise, *Street Mathematics and School Mathematics*, Nunes, Schliemann, and Carraher (1993) demonstrate through a series of empirical studies that there exist cases in which student perform much better in solving arithmetic problems in context than solving symbolic problems with exactly the same mathematical operations. It suggests that the contexts presented in the word problem facilitate students' understanding, so that they can see the problem in a more concrete sense, and solve it in more flexible ways (i.e. in "street ways") as opposed to school-taught procedures. Nunes et al. (1993, p. 35)

To summarize, it was found that the difficulty of word problems is mainly related to: (1) the impractical nature of some word problems, which hinders students' realistic sense-making; (2) schooling that binds students with extra "norms" in solving the "standard word problems", but there is no superficial way to tell whether a word problem is standard or not. We see that although word problem tests some aspects of application, the use of it in assessing students' ability to apply may be confounded by (1) the extent to which the problem situation makes sense, and (2) students' belief about solving word problems.

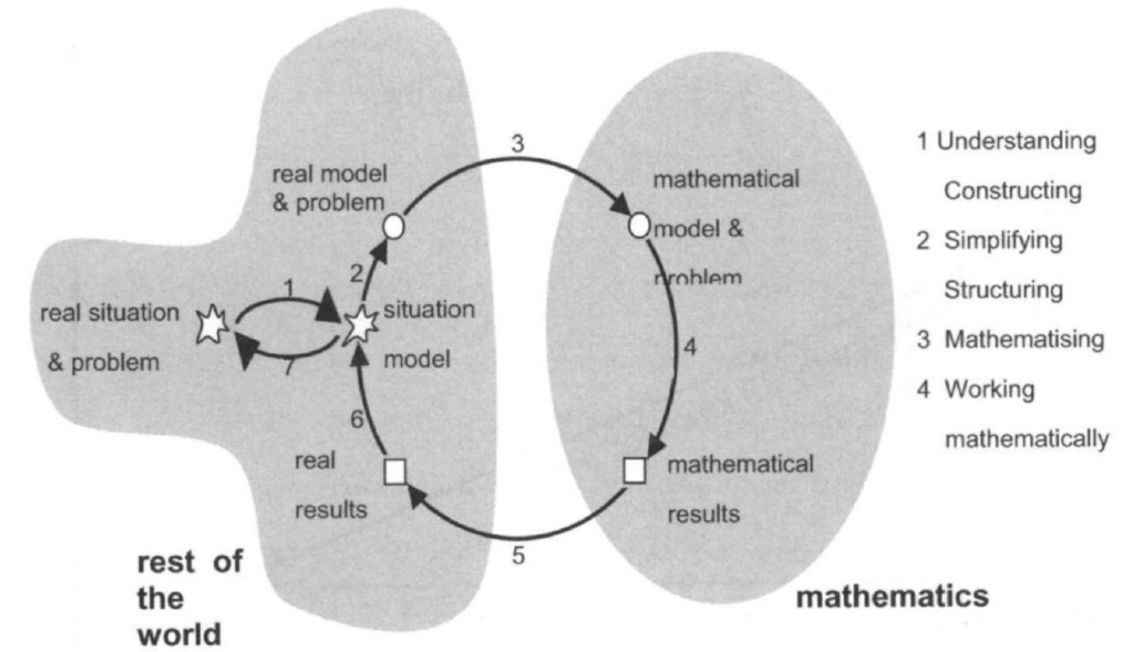
Ability to Apply in Terms of Modeling

Many books and dissertations have reviewed the definitions of mathematical modeling. There exists a great variety of the understanding of mathematical modeling among mathematicians (e.g. Drakes, 2012; Tian, 2014), textbooks (e.g. Germain-Williams, 2014)mathematics teachers (e.g. Gould, 2013), and mathematics education researchers (e.g. Ärlebäck, 2010; Sriraman & Kaiser, 2006). Not surprisingly, not everyone agrees on the details about what mathematical modeling is, but a common understanding is that it involves a "mathematical model", which represents part of the "reality", and also activities associated to the model, the reality, and the connections between them. The variation of understandings lies in the very different descriptions of each element in the previous statement. Questions including the following are often reviewed: (a) What is a mathematical model? (b) What does it mean when we say that a model represents some reality? (c) Which part of reality can be modeled? (d) What activities are associated to the connection between reality and mathematical models? (i.e. modeling processes) A rarely reviewed albeit important question is (e) what does "reality" refer to in modeling? A full review of all the questions is out of scope or unnecessary here. Since the current dissertation focuses on the ability to apply mathematical knowledge, the activities connecting reality and mathematical models are of interest.

The activities of modeling are often organized as a *modeling process*. Though there are many descriptions of such process, it is possible to identify a few unique ones. The first is *modeling cycle*, which mostly assembles the eight modeling steps (Pollak, 2003) described in the beginning of this chapter. In the eighth step, it says "if the model is wrong or not satisfactory, find the cause of it and start over", and therefore it becomes a cycle when the process continues. Blum and Leiß (2007) organized the modeling process as a cycle in a more explicit way, as shown in Figure 2.1, and it is commonly adopted with minor adjustments in the research literature (e.g. C. R. Haines & Crouch, 2010)Jab97. Blum and Leiβ explained that this version of modeling cycle is "more oriented towards the problem solving individual" (Blum & Leiß, 2007, p. 226), as a contrast of earlier views, e.g. in Blum (1995), that focused more on ideal methods of applied problem solving. Especially, the ways in which students construct a real model from the real situation (Step 1 in Figure 2.1) are observed to be idiosyncratic, and therefore do not fit in the ideal descriptions of problem solving.

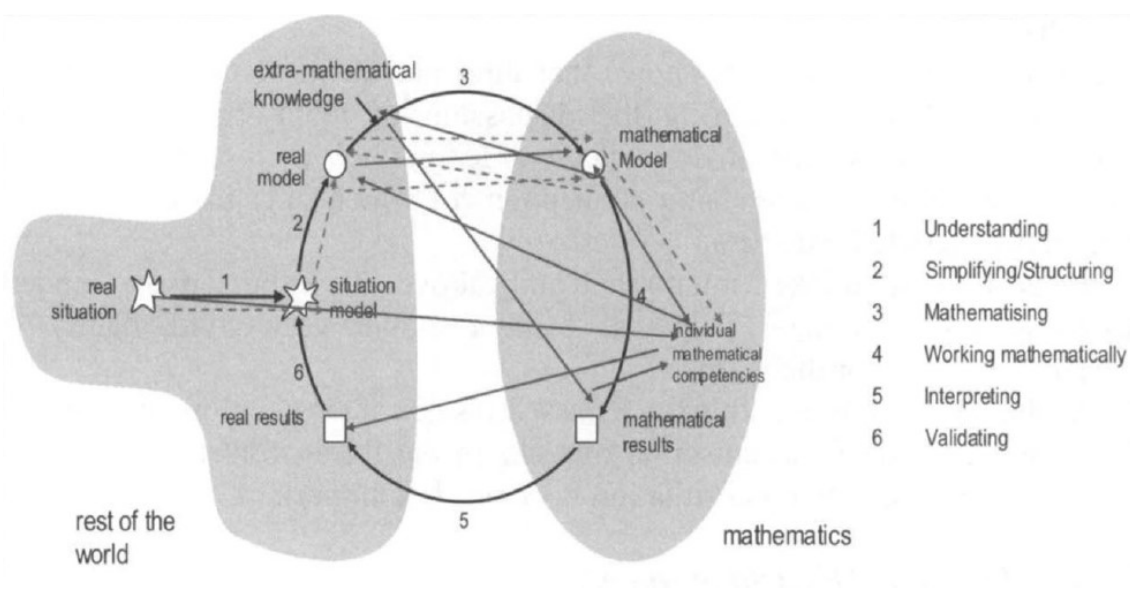
(3) The limitation of the modeling cycle is that it suggests a linear process of modeling, namely, the steps occur one after the previous number. However, (Blum & Leiß, 2007) already noted that this may not be the case. Indeed, as (Frejd, 2013) pointed out, researchers had shown empirically that the "modeling routes" are not likely to resemble the modeling cycle (Borromeo Ferri (2007); also see Figure 2.1). An alternative description to the modeling cycle is the modeling process described in Blomhoj and Højgaard Jensen (2003), which essentially contains all the modeling steps, but these steps are loosely labeled as *sub-processes* of modeling, avoiding a pre-defined order that is not supported by research.

Figure 2.1: Modeling cycle.



This item shows the modeling cycle depicted in Blum and Leiß (2007)

Figure 2.2: Modeling routes.



Modeling cycle (darker solid lines) compared to modeling "routes" of two very different individuals: an "analytical thinker" (solid lines, less dark) and a "visual thinker" (dashed lines). (Borromeo Ferri, 2007)

Modeling competence

In the sense of mathematical modeling, the applying of mathematical knowledge happens when one goes through the modeling process. Thus, the ability to apply mathematical knowledge can be treated as the possession of knowledge and skills necessary for any part of the process. Such ability is often referred to as modeling competence, modeling competency, modeling skills, and so on; the term "modeling competence" will be consistently used hereafter. If the goal is to use mathematical modeling to solve real-world problems, some researchers also include meta-cognitive and emotional / affective capacities as part of modeling competence when they are considered important during the modeling processes. It is thus evident that discussions of modeling competence rely mostly on the modeling process, through which a competent student should be able to solve applied problems mathematically on one hand, and to contextualize their mathematical knowledge on the other. A different description of the modeling process leads to a different understanding of modeling competence and its measurement.

Reviews of the state-of-the-art discussions and works on modeling ability and its assessment are available at different times of the development (Niss, 1993; Blum, Galbraith, Henn, & Niss, 2007; Kaiser & Brand, 2015).

Modeling competence has been discussed extensively, and many theoretical perspectives can be found in two special issues of *ZDM* (38(1) and 38(2)), a chapter of the ICMI 14th Study (Blum et al., 2007), a chapter in ICTMA 12 (C. Haines, Crouch, & Davis, 2000), and another in ICTMA 14 (Kaiser, Blum, Ferri, & Stillman, 2011). In a most recent review, Kaiser and Brand (2015) discussed the development of modeling competence from the debates and discussions in the ICTMA series (from ICTMA 1 to ICTMA 15). In their review, Kaiser and Brand (2015) stated that a "joint understanding" (p. 129) of modeling competence is not yet available, meaning that there are understandings maintained by different scholars, and they may be incompatible. There is a need to synthesize these discussions in order to discern the more accurate than the less useful ones, and more ideally, a coherent and complete view of the matter of modeling competence.

According to Kaiser and Brand (2015), the research studies of modeling competence can be organized into the following general questions: (a) How can modeling competency be conceptualized? (b) How can it be tested? (c) How does a person develop modeling competency? (d) How should such competency be fostered most efficiently? (e) What types of assessment methods are appropriate in practice in order to evaluate modeling competency? Questions (a) and (b) are the most relevant to the current thesis, and will be focused on in this section.

It is not a surprise that the discussions on these questions are often dependent on one another. In particular, answering how modeling competency is conceptualized can inform us about the ways it can be tested, and the demands of measuring such competence necessitate a discussion on the meaning of modeling competence. Nonetheless, for clarity, (a) the nature and (b) the tests of modeling competence are discussed separately.

In a process of construct validation, there are some initial theories about the construct, and some initial designs of measure that are supposed to reflect the current understanding of the construct. Then, the relation between the theories and the designed measures are assessed in order to refine the theories, the measures, or both of them (Westen & Rosenthal, 2003). Therefore, in light of the construct validation process, this section first reviews the theoretical statements about modeling competence (as a construct), and then review the proposed ways to measure it. Finally, the extent to which modeling competence answers the research question 1, i.e. how is modeling competence related the ability to apply mathematical knowledge, will be discussed.

Since the beginning of the discussions on the teaching and learning of modeling, mathematics educators had been concerned about how to help students to become proficient in modeling (see e.g., David Burghes' Prologue in the first ICTMA, 1984), but it was more of an educational demand other than a theme of study until there was a necessity in some large-scale curriculum projects, in which modeling courses were developed, and students' achievement on such courses are defined and assessed as a consequence (Kaiser & Brand, 2015). Thus, the first definitions of modeling competence were equivalent to the proficiency according to specific modeling courses. There were also intensive discussions on how such proficiency can be assessed.

According to Kaiser and Brand, scholars in this early stage followed the trend of psychology and pedagogy in using the construct of "ability" and "skills" to understand modeling competence; however, in an international level, there was no clear definition on this topic of study, and, as concluded by Blum and Niss, the assessment created at that time did not reach "well-defined standards" (p. 131, cited Blum and Niss (1991)). At a national level, nevertheless, some significant work had taken place, such as Treilibs' work at the University of Nottingham in UK, and Kaiser's empirical study in Germany. Treilibs defined the following skill components of modeling: (a) generating pertinent variables; (b) selecting important variables; (c) identifying and posing the questions; (d) generating relationships between the variables; and (e) selecting adequate relationships. In Kaiser's definition of modeling skills, she distinguished it from the ability to apply mathematical knowledge and equated it as "model building skills", which referred to different phases of a modeling process, or "modeling cycle" (p. 131-132, cited Treilibs (1979), Kaiser-Messmer (1986)).

Other scholars were more interested in the cognitive (Matos & Carreira, 1995) and the meta-cognitive (Tanner & Jones, 1995) processes involved in mathematical modeling, and these are still important themes in the more recent studies. Matos and Carreira did not attempt to analyze the nature of modeling competence, but instead constructed cognitive explanations of the students' behavior when they were solving problems in a modeling activity. Thus, students' performance is determined by how well they "activate" the mathematical elements when they are working on the real-world situation (Matos & Carreira, 1995, p. 78). Thus, this point of view implies that modeling competence is strongly dependent on the cognitive process of connecting real-world elements and mathematical elements. Meta-cognitive skills, including the planning, monitoring, and controlling of one's own thinking process, was found to be necessary when Tanner and Jones observed that teaching modeling skills "through strong teacher direction" did not enable students "to transfer to even slightly different situations" (Tanner & Jones, 1995, p. 65).

Danish researchers Blomhøj and Højgaard Jensen made an important theoretical distinction between two types of approaches of understanding modeling competence. They first described the construct competence as the "insightful readiness to act in response to the challenges in a given situation" (Blomhoj & Højgaard Jensen, 2003, p. 126), which makes a contrast to the usual definitions of skills and knowledge because it underscores students' willingness and activeness to respond in the given situation. Then they distinguished approaches to understand modeling competence into "holistic" versus "atomistic". A holistic approach to modeling competence concerns the whole modeling process, whereas an atomistic approach tends to see the modeling process as a combination of sub-processes, such as "mathematizing the situation" and "interpreting the results".

Note: When distinguishing holistic and atomistic approaches, some scholars focus on an understanding of the nature of modeling competence and how to measure it, but more authors refer them to the ways in which modeling competence is developed. While the importance of competence development should be acknowledged, the current chapter only focuses on the former issues.

Examples of holistic approaches: Studies on the cognitive models that govern the whole modeling process (Matos & Carreira, 1995); Examples of atomistic approaches: Treilibs' skill components of modeling; Kaiser's "modeling building skills". In the following, more examples will be described and be classified as holistic or atomistic (cf. Table 2.1)

In the curriculum project Competencies and Mathematical Learning in Denmark (the *KOM Project*, Niss and Højgaard Jensen (2011), modeling competence is embedded in mathematical competence. Continuing the discussions of Blomhoj and Højgaard Jensen (2003), mathematical competence is defined as "someone's insightful readiness to act in response to a certain kind of mathematical challenge of a given situation" (Højgaard Jensen, 2007, p. 142), and modeling competence is defined as "someone's insightful readiness to carry through all parts of a mathematical modelling process in a given situation" (Højgaard Jensen, 2007, p. 143). Niss and Højgaard Jensen (2011) defined modeling competence in a greater detail as follows:

> This competency involves, on the one hand, being able to analyse the foundations and properties of existing models and being able to assess their range and validity. Belonging to this is the ability to "*de-mathematise*" (traits of) existing mathematical models, i.e. being able to decode and interpret model elements and results in terms of the real area or situation which they are supposed to model. On the other hand, the competency involves being able to perform active modelling in given contexts, i.e. mathematising and applying it to situations beyond mathematics itself. (p. 58, Italics original)

In short, Niss and Jensen's definition of modeling competence involves two main parts: (a) *de-mathematization* — focusing on the exposition of an existing mathematical model, and (b) "mathematization" — focusing on model building in a given extra-mathematical context. This definition is seen as holistic according to Kaiser and Brand (2015).

Another holistic approach of modeling competence was devised by GV07. In their attempt of classifying ICMI 14th Study papers on modeling competence, they identified three levels of awareness of mathematical modeling, being implicit, explicit, or critical. Implicit modeling occurs when what students actually do is indeed modeling but they are not aware of the modeling concepts or terminology. Students involve in explicit modeling when they articulate about the steps within their modeling processes. Critical modeling refers to cases when one questions or examines the role of modeling within mathematics, science, and the society.

C. Haines and Izard (1995) applied an item-response theory approach, which can be seen as atomistic. In this approach, a group of experts decided what they considered important when they assess students' modeling work, and then they produced a list of "descriptors" to rate students' modeling achievement. The list is modified and refined in order to fit the psychometric properties required by an item-response theory. Each item of the list correspond to one "sub-competency" of modeling, and thus the whole list describes what the group of experts considered as modeling competence.

As mentioned, Kaiser's work (Kaiser-Messmer, 1986) is another of example of an atomistic approach. She further defined "modeling competencies", each according to a phase within a modeling cycle:

- competencies to understand real-world problems and to construct a reality model;
- competencies to create a mathematical model out of a real-world model;
- competencies to solve mathematical problems within a mathematical model;
- competency to interpret mathematical results in a real-world model or a real situation;
- competency to challenge solutions and, if necessary, to carry out another modeling process. (Kaiser (2007, p. 111), also appeared earlier in Blum and Kaiser (1997).)

It was also realized by researchers using the atomistic approach that there should also be competencies for meta-cognition, as emphasized by Tanner and Jones (1995), and later Maaß (2006). Maaß also included some other abilities that are related to communication, critical reflection, and positive attitudes. She proposed a more comprehensive list of modeling competencies:

1. Sub-competencies to carry out the single steps of the modeling process (adopted from Blum and Kaiser (1997))
2. Meta-cognitive modeling competencies (see also Stillman & Galbraith, 1998)
3. Competencies to structure real world problems and to work with a sense of direction for a solution
4. Competencies to argue in relation to the modeling process and to write down this argumentation
5. Competencies to see the possibilities mathematics offers for the solution of real world problems and to regard these possibilities as positive. (Maaß, 2006, p. 139)

As a result, the atomistic approach tends to include all competencies that would be helpful in performing modeling tasks, though there is a lack of discussion of the hierarchical structure of these "atoms".

Table 2.1: Examples of holistic or atomistic approaches in the study of modeling competence

Holistic approaches	Atomistic approaches
Greer and Verschaffel (2007)	Haines and Izard (1995)
Matos and Carreira (1995)	Kaiser-Meßmer (1986); Kaiser (2007)
Niss and Højgaard (2011)	Maaß (2006)
Treilibs (1979)	

Blomhøj and Højgaard Jensen's distinction is not only helpful in categorizing the existing research, but also offers a richer picture of the understanding of modeling competence. For example, using the *holistic versus atomistic* distinction, Kaiser and Brand (2015) synthesize the existing discussions by contrasting *competences* and *competencies*. When one views modeling competence holistically, there are some competences that one must be proficient on in order to have good overall performance on modeling. The competences are treated as a whole, and criteria can be set to see whether a student achieve a certain level or not. In contrast, under an atomistic view of modeling competence, researchers take an analytic approach to separate the modeling task into subtasks, and assign each task with the competencies necessary, so that the overall competence is determined by the accumulation of these sub-competencies.

The holistic view and the atomistic view are not a pair of dichotomy, as they do not contradict each other. Instead, asKaiser and Brand (2015) suggested, they can be seen as complementary views. More ambitiously,BHJ03 discussed that an integration of both views might be adequate. It is also arguable that since none of these two views is a satisfactory description of modeling competence, nor should a juxtaposition of both be sufficient. Thus this are still open questions: (1) how can modeling competence be described completely and coherently? (2) Does holistic or atomistic view better describe modeling competence? (3) Is an "integrated approach" BHJ03 satisfactory?

To answer these questions, researchers have done empirical studies that are designed for the understanding of modeling competence and its development. Since these studies are based on the measurement of modeling competence, the methods of measurement are reviewed in the following sub-section.

Measurement of Modeling Competence

At least two different aspects of measurement can be discussed. The first aspect is conceptual: what does it mean to measure modeling competence? In particular, what kinds of evidence are to be observed in order to infer a person's modeling competence? The second aspect is more practical: what are particular forms of tests, such as multiple-choice response items or oral presentations, that are available for the measurement of modeling competence? How are the testing instruments evaluated?

This section introduces existing approaches of measuring modeling competence, including Shavelson's model that addresses competence in general, and Niss & Jensen's multidimensional approach that applies to competence in mathematics. Both are applied to mathematical modeling competence, in particular.

Competence measurement in general

In general, researchers treat competence as a construct, for it is not directly measured. Shavelson (2010) proposes a model of measuring competence that is based on the assessment triangle that is composed by Construct, Observation, and Interpretation. Firstly, there must be some initial theory or idea about the construct. Secondly, some tasks are identified that are thought to "evoke" the competence (p. 43), so that a person's behavior can be observed thorough his / her dealing with those tasks. Finally, the observations allow researchers to interpret about the person's performance or the level of performance on such competence. Shavelson mentioned that a *universe of possible tasks* (Shavelson, 2010, p. 43) should be logically implied by the definition of competence, and researchers take a sample of those tasks in order to construct an assessment. He recommended the method of criterion sampling (McClelland, 1973) over the traditional approaches. In Shavelson's view,

> Traditionally, complex tasks are analyzed into their component parts and psychological traits underlying them are identified. Tests are then built to measure each trait, and the sum of the scores on each test is supposed to put the pieces back together again to represent the whole of performance. For McClelland (and me), something is lost in the traditional approach; the whole is greater than the sum of the parts. A person might form a response to a situation by calling on a complex set of skills; those skills might change as the task changes over the course of the assessment. The closer the task reflects real life situations, the more likely the person's responses on the task reflect responses she makes in life. (Shavelson, 2010, p. 47)

Criterion sampling means that researchers construct tests based on what is actually expected or desired to be done in a real situation. For example,

" If you want to know how well a person can drive a car (the criterion), sample his ability to do so by giving him a driver's test. Do not give him a paper-and-pencil test for following directions, a general intelligence test, etc. (McClelland, 1973, p. 7)"

McClelland went quite far to say that paper-and-pencil tests for following directions should not be given in a driver's test; this view can be seen as a reaction to the traditional approaches that rely too much on "word-games" (McClelland, 1973, p. 6)

Under Shavelson and McClelland's point of view, measuring modeling competence necessitates a criterion sample of what a "good modeler" does in a real situation which is to be modeled mathematically. An instrument in measuring modeling competence is considered valid only if it reflects what is desired to happen when the test taker faces a similar, real situation. A person is considered a competent modeler if his / her behavior conforms to how the desired "good modeler" behaves.

As Kaiser and Brand (2015) pointed out, Shavelson (2010) identified two approaches of competence measurement that are also named *holistic* and *analytic*, and have similar meaning as the holistic view and atomistic view of modeling competence that are discussed above. Indeed, Shavelson (2010) also mentioned a *hybrid* approach. Moreover, the classification by Shavelson (2010) does not apply to the definition of competence as a construct, nor the tasks required in a measurement tool, but rather how the responses of a test taker are interpreted and scored. In a holistic scoring, a prescribed rubric guides a rater to assign an overall score from a 6-point rating scale; in an analytical scoring, the test taker performs multiple tasks, and a rater has to assign a score for the performance of each task. Shavelson (2010) briefly mentioned that a holistic approach is short of diagnostic function of a test, whereas an analytical approach can be highly diagnostic, but it also tends to create a "hodgepodge" of scores that are inconclusive. A hybrid approach seeks to combine the holistic and analytic approaches by introducing some universal variables that can be applied in all possible tasks, such as time and accuracy. To Shavelson, however, this hybrid approach can also be inconsistent.

One has to note, however, that there is an important difference in the usage of the words holistic and analytic (atomistic) between Kaiser and Brand (2015) and Shavelson (2010). In Shavelson (2010), the tasks are already sampled, but a test taker's response can be interpreted using holistic or analytic (or hybrid) scoring. In Kaiser and Brand (2015), they described a design to evaluate students' modeling competence by a set of multiple choice items, each of which belongs to exactly one of the four dimensions: (1) Simplifying/Mathematising, (2) Working Mathematically, (3) Interpreting/Validating, and (4) Overall Modelling Competency. The first three dimensions are atomistic components of modeling competency, and so the items that belong to these dimensions can be seen as a sample of tasks under the atomistic view of competency. The fourth dimension represents the holistic view of competency, and therefore the items that belong to it can be seen as a sample of tasks under such view. Therefore, in terms of criterion sampling, the terms holistic and analytic in Kaiser and Brand (2015) indeed refer to the definition of modeling competence, thus it affects how tests items were designed. Moreover, the empirical study described in Kaiser and Brand (2015) also involves a holistic or atomistic approach of teaching, and the tests were to eva-

Table 2.2: Usage of the word holistic and atomistic (analytic)

In the description of modeling competence	Holistic view — competent modelers have the necessary skills to carry out the whole modeling process, and therefore they are able and ready to perform it Atomistic view — competent modelers are able and ready to perform all sub-processes that combine to a whole modeling process.
In the way of scoring (interpretation) of how test takers respond to a designed task	Holistic scoring — an overall rating that reflects the students' overall modeling competence Analytic scoring — one score for each task that the student performs in the test Hybrid scoring — the overall rating is also based on some universal variables that can be objectively measured in each task, such as the time elapsed
In the way of teaching	Holistic approach — whole modeling tasks are assigned to students, with a gradually increasing level of difficulty Atomistic approach — tasks of different components in the modeling process are to be done by students, one component at a time

luate how these approaches are better in improving students' modeling competence. A summary of the usage of the words holistic and atomistic (analytic) is shown in Table 2.2.

Shavelson's approach provides a general viewpoint of competence. As a principle, the measurement of modeling competence depends on the definition, possible observations, and methods of interpretation. Now it should be applied to the case of modeling competence. Definitions of modeling competence are introduced in the last subsection, where it can be described in holistic or atomistic ways, yet they are not contradictory; both of them should be kept as part of the descriptions of modeling competence until a better approach is available. Possible observations refer to certain tasks specifically designed so that it requires modeling competence to accomplish. Interpretations involve the judging of the test takers' performance on the tasks. Thus, a good design of the tasks not only depend on the definition of competence, but also on the ways of inferring how competent a person is on modeling, based on his / her behavior.

Inferring a person's modeling competence

For the purpose of inferring about a person's modeling competence, important formulations were established again by Niss, Blomhøj and Højgaard Jensen. They proposed three dimensions that determine a person's level of competence: (1) degree of coverage, (2) radius of action, and (3) technical level (Højgaard Jensen, 2007). When applied to modeling competence, they mean the following:

> the degree of coverage addresses which part of the modelling process someone can work with and the level of the reflections involved. [...] the radius of action addresses the domain of situations in which someone can perform mathematical modelling activities. [... and] the technical level addresses which kind of mathematics someone can use and how flexible they are in their use of mathematics. (Blomhøj & Jensen, 2007, p. 55)

This formulation goes beyond the modeling cycle, for it implies that the measurement of modeling competence cannot simply be done by looking at how a person is competent at different steps in the cycle. Indeed, the evocation of any modeling step (dimension 1) can be mediated by the context one is working within (dimension 2) and the possible mathematics one is pondering about (dimension 3). There is a high demand of empirical research to understand how these dimensions are related.

Using the three dimensions above, it becomes convenient to classify the kinds of measurement to be done. For the current thesis in particular, the measurement of the ability to apply mathematical knowledge can be seen as focusing on (3) technical level, as such ability can be defined as the extent on the other two dimensions, given that a certain kind of mathematics is mastered.

Many previous studies focused solely on dimension 1, i.e. assessments of how each modeling step is evoked. Houston (2007) reviewed the assessments according to the modeling phases, where he presented four existing assessment schemes, proposed by Hall (1984), Berry and Le Masurier (1984), the UK Assessment Research Group (ARG) (C. Haines & Houston, 2001), and C. Haines et al. (2000). Each scheme specified items of ability to be measured, and Houston (2007) mapped each

Table 2.3: Number of ability items by each phase of modeling

	Hall	Berry & Le Masurier	ARG	Haines, Crouch and Davis
Number of ability items	14*	16	20	8
Modeling phases **				
1 specify the real problem	2	2	2	2
2 create a mathematical model	5	5	6	5
3 specify the math problem	3		2	3
4 solve the math problem	3	4	2	
5 interpret the math solution	2	2	2	1
6 validate the model	2	6	2	1
7a revise	2	1	1	
7b report	4	2	9	

* Each number in this row is not necessarily the sum of the numbers below it, because some ability item was mapped to more than one phase.

** The phases adopted by Houston (2007) are similar, but not the same as the Eight Steps (p. 11).

ability item to one or more modeling phases. Table 2.3 illustrates the variations of these schemes by the number of ability items. The phase "create a mathematical model" is either the first or the second on the list. Some scheme focuses more on the presentation work; other more on the earlier phases.

Henning and Keune (2006) proposed another formulation of competence levels. Their formulation has only one dimension with three levels: "Level 1: Recognize and understand modelling; Level 2: Independent modelling; Level 3: Meta-reflection on modelling" (Henning & Keune, 2006, p. 1668). From the descriptions, it seems that these levels focus more on the meta-knowledge and the attitudes for modeling. The limitation is that it is not specific enough in describing the context, the mathematics actually used, and the complexity of the problem. For example, it described level 2 as follows:

> Level 2 – Independent modelling – is characterized by the ability:
> -to analyze and to structure problems and to abstract quantities,
> -to adopt different perspectives,
> -to set up mathematical models,
> -to work on models,
> -to interpret results and statements of models,
> -to validate models and the whole process. (Henning & Keune, 2006, p. 1668)

It depends strongly on the actual problems and possible models whether one can be treated as a “level 2 modeler”! Nevertheless, this meta-knowledge dimension is not covered by any of the three dimensions described above, and can possibly be treated as the fourth dimension. This will be one of the main concerns in Chapter 4, which analyzes various theoretical proposals and hopefully synthesizes them.

Forms or “modes” of observations / tests

Given the definitions of modeling competence and possible formulations of inference, it is necessary to process a collection of methods to observe students’ actual behaviors that possibly reflect their modeling competence. The observations are usually referred to as “tests”, but the word test also has other connotations. For example, a high-stake test is treated as an important factor of decision making, associating tests to deep pedagogical, administrative, social, and political meanings. These meanings are very important in mathematics education, and in the teaching and learning of modeling in particular. Nevertheless, tests will only be discussed in a measurement sense, i.e. as an observation of students’ behavior through their responses to designed tasks or situations.

Peter Frejd (2013) has reviewed the forms of tests that were studied in published articles, including those that are available in prominent journals and conference proceedings (until 2011). He identified several forms of tests (or *modes* by Frejd), including (a) written tests, (b) projects (including verbal presentations), (c) hands-on tests, (d) portfolio, and (e) contests. The written form of test occurs most frequently, so Frejd further analyzed it into sub-categories such as multiple-choice questions, shorter tasks, extended tasks, unseen / seen tests, final exams, and shorter classroom tests. After his critique of the existing forms of tests, it was apparent that there has not been any satisfactory and reliable mode of modeling assessment. Also, Frejd found that written tests, being the most adopted mode of testing, mostly call on an atomistic view of modeling competence. Therefore he questioned if it is at all possible to construct written modeling tests that reliably assess holistic modeling competence. Finally, Frejd (2013) suggested that though written test is a more prevalent mode of modeling assessment, the modes (c) hands-on tests and (d) portfolio are more promising, whereas (b) and (e) tend to have problems of interpretation, that is, the rating of results tend to be subjective, and the criteria may not be transparent enough.

Frejd (2013) pointed out that these test modes align pretty well with the existing assessment methods in mathematics education in general. An example is Watt's (2005) six assessment methods, including (a) oral tasks, (b) practical tasks, (c) teacher observation, (d) student journals, (e) peer and self-assessment and (f) parental assessment, where parental assessment is an exception, as it has not (yet) been applied to mathematical modeling. However, Frejd's suggestions that some mode is better than the other are not conclusive unless some overarching principles to evaluate the test modes exist. Since such principles are not found in literature that are directly related to modeling, one might fall back temporarily to the McChelland-Shavelson position and ask the following questions: For any given test (or test mode), can it represent the possible modeling tasks or subtasks as a *criterion sample*? (cf. 2) That is, how much of the test requires the test-taker to act like what a "good modeler" is capable to do? In this sense, one may draw some principles from empirical studies of the real 'mathematical modelers', i.e. applied mathematicians. Such studies can be found in some recent studies (Ekol, 2011; Drakes, 2012). These empirical studies will be further analyzed in Chapter 4, seeking to establish some guiding framework to evaluate modeling tests.

Alternative Notions of the Ability to Apply

The modeling process framework provided a way to understand how mathematics is applied to various situations, and the concept of modeling competence can also been seen as a theoretical construct of the ability to apply mathematical knowledge. In the research literature, however, alternative notions can be found that are related to the ability to apply. In the following, a small collection of them are introduced.

Vocational mathematics

Unlike word problems, which are most likely composed for curricular settings, vocational situations provide real-world applications of mathematics. Moreover, it is believed that general education should equip students with competence of mathematics and its application so as to be successful in their future career. Henceforth, it motivates us to review the articles in which application of mathematics plays a role in vocational situations. We use the phrase *vocational mathematics* to refer to any application of this kind. We will see how vocational mathematics is discussed in the recent literature, and how it is related to competencies and assessment of application.

In a narrow sense, vocational mathematics can be seen as a special subject that is taught in vocational schools (either secondary or tertiary), where students are prepared for technical jobs. This purpose is described in a vocational mathematics textbook, published a century ago (Dooley, 1915):

" ... the inability of the regular teachers in mathematics to give the pupils the training in commercial and rule of thumb methods of solving mathematical problems that are so necessary in everyday life. (p. iii)"

Despite of this, perspectives on the development of vocational mathematics did not start to evolve until the 1980s (we refer to, e.g., The Cockcroft Report (Cockcroft, 1982), and the Mathematics in Adult, Technical and Vocational Education Action Group in the 5th International Congress on Mathematics Education (Sträßer & Thiering, 1986)). The discrepancy between school mathematics and how mathematics can actually applied, as mentioned earlier this chapter (see page 15), became a major concern of researchers who are interested in vocational mathematics, and they attempted to theorize in a few different ways. First of all, the traditional view, which tended to describe the discrepancy as a failure of transfer, has been considered problematic (Evans, 2002, p. 75). *Transfer of learning* generally refers to the application of knowledge learned in one context to another, and it was once thought by behaviorists to be natural that students can transfer what they learned as long as similar elements are taught properly. Lave (1988) proposed the perspective of *situated cognition*, that individual's thinking in real contexts is specific to exact practices, and that a classroom environment is for the most part disjoint from the real practice. Therefore, to Lave and other supporters, transfer of learning is not likely to happen through traditional classroom learning. Lave's perspective is based on her anthropological study of individual's arithmetic activity in grocery shopping (Lave, 1988, p. 156). Grocery shoppers are generally performing very well in the arithmetic involved in the shop, but they performed poorly in the academic style problems that were designed as the supposedly same mathematics they did. It is believed that the shoppers were not using the "school ways" to do the arithmetic (Devlin, 1999).

Recent studies show that similar discrepancy is also present between vocational programs and real workplaces. For example, Roth (2014) studied a group of apprentices in a 4-year training program towards a required certificate for licensed electricians in Canada. It is understood by both students and instructors that the mathematical contents they learned "were irrelevant to the worksite", and that it becomes "an integral part of the professional lore" that such irrelevance exists between training program and the workplace, as many of the skilled workers just have to upgrade their certificates during their career life, while these certificates require that they take part in examinations of formal mathematical contents that the workers perceive as irrelevant. Having that said, there is not yet a satisfactory and consistent view on what mathematics is as both an academic discipline and a collection of necessary knowledge that is useful in workplace practice. Indeed, the Topic Study Group 5: *Mathematics In and For Work* in the latest ICME-12 agreed that "mathematics includes the academic discipline as well as the practices in communities and workplaces that involves mathematical thinking." Such a description which juxtaposes the two aspects of vocational mathematics (i.e. for work and in work) suggested a lack of harmony between the two. Nonetheless, many scholars are attempting to harmonize. Another feature of vocational mathematics is that it is often hidden. It is either seen as *common sense* just as described in the quote from the Cockcroft Report (1982), or as *professional knowledge* that workers would not recognize as mathematics. This is known as the invisibility of mathematics in workplaces (e.g. Damlamian & Sträßer, 2009; Wake & Yasukawa, 2015).

Williams and Wake (2007) conceptualized the invisibility as black boxes, and made an assumption that in workplaces, actual mathematics is present within the black boxes. Automation through the usage of instruments and division of labor improve efficiency of facilities. It is therefore beneficial and normal to hide the technical details (e.g. mathematics) so that regular workers do not need to understand or even to be aware of them. As the authors put it, the workers could be "protected from mathematics". Hence, vocation mathematics is the closest to the "real-world", putting less emphasis of mathematics as a subject.

Mathematical literacy

The term *Mathematical literacy* is used in the PISA assessment (cf. 5) that focuses on the "functional use" of the mathematical knowledge rather than the mastery in the mathematics curriculum. (Stacey & Turner, 2015, p. 9). In a recent assessment framework of PISA, mathematical literacy is defined as:

> an individual's capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts, and tools to describe, explain, and predict phenomena. It assists individuals to recognise the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens. (Organisation for Economic Co-operation and Development [OECD], 2013a, p. 25)

It is possible to argue that mathematical literacy is just another synonym for *mathematical competence*, but Niss (2015a) pointed out that this is a misinterpretation of the purpose of the term:

> [Equating mathematical literacy and mathematical competence] is certainly a possible but not really a desirable option. The perspective adopted in PISA, right from the outset, was not to focus on young people's acquisition of a given subject, such as mathematics, but on their ability to navigate successfully as individuals and citizens in a multifaceted society as a result of their compulsory schooling. This zooms in on putting mathematics to use in a variety of mainly extramathematical situations and contexts, in other words the functional aspects of having learnt mathematics. This is what mathematical literacy is all about, being brought about by way of modelling. (Niss, 2015a, p. 53)

Niss defines mathematical competence as a composition of eight sub-competencies, where modeling competence is one of them; mathematical literacy is seen as requiring many of these sub-competencies, but centered most on modeling competence. According to its purpose, mathematical literacy is actually closer in meaning to the term *functional mathematics*, since the primary concern is that of practicality, while mathematics as a subject of knowledge is secondary.

Realistic Mathematics Education

Realistic Mathematics Education (RME) is a theory of mathematics education that is developed by a group of Dutch scholars. RME claimed that realistic situations serve as an environment that triggers learning; at a later stage, students' understanding can become more formal and less dependent on the realistic context. In RME, the meaning of "realistic situation" is broader than things being existent in the real-world, and includes the situations that can be imagined by students, such as fairy tales in a "fantasy world", or mathematical, symbolic situations in a "formal world". Therefore, the concern of the functional use of mathematics in daily life is not emphasized so much as by advocates of mathematical literacy.

Mathematical model is considered important in RME, since it "bridges the gap between the informal, context-related mathematics and the more formal mathematics." (Van den Heuvel-Panhuizen & Drijvers, 2014, p. 523) According to Van den Heuvel-Panhuizen (2003), modeling activities are the means to develop mathematical concepts such as percentage. To RME scholars, the mathematics and the use of it in context cannot be separated. Such perspective traces back to Freudenthal, who viewed mathematics as "not the body of mathematical knowledge, but the activity of solving problems and looking for problems, and, more generally, the activity of organizing matter from reality or mathematical matter — which he called 'mathematization'. " (Freudenthal, 1968)

Finally, there are other phrases that are related to the ability to apply mathematical knowledge. The phrase *applied mathematics* usually refers to a branch of mathematics that studies mathematical models and methods that have usage in other disciplines. According to Stolz (2002), one can trace how mathematical models and modeling entered other subjects by reviewing the development of applied mathematics. Nevertheless, since applied mathematics is often associated with established mathematical topics such as partial differential equations, the mastery of applied mathematics might not indicate that the whole mathematical modeling cycle is mastered, or is only done once in a classical example. To see whether one can really use the knowledge in other situations, there are attempts to measure *transfer* of mathematical knowledge (cf. New, Britton, Sharma, & Brew, 2012; Potgieter, Harding, & Engelbrecht, 2008), a concept studied in psychology.

The various concepts and notions related to the ability to apply mathematical knowledge will be used as keywords in order to collect articles that contain related theories.

Methodological Background: Theories in Mathematics Education

Notions of theory and theoretical framework

One objective of this research is to survey current theories or theoretical frameworks on the ability to apply mathematical knowledge. It is necessary to clarify what are acknowledged as theories before research is done on them, as there is a wide range of understanding of what a theory or a theoretical framework is in mathematics education (e.g. Niss, 2007; Assude, Boero, Herbst, Lerman, & Radford, 2008). For example, *learning theory* is a commonly used term, but as pointed out by Ernest (2010), the theories of learning are often "not specific or testable enough to be called theories" and therefore are better referred to as *learning philosophies*.

Schoenfeld's eight-points (2000; 2001) is a classical set of criteria on evaluating empirical or theoretical work in mathematics education:

1. descriptive power
2. explanatory power
3. Scope
4. predictive power
5. rigor and specificity
6. Falsifiability
7. Replicability
8. multiple sources of evidence

Although Schoenfeld proposed these as "standards for judging *theories, models, and results*", not all theories, models, and results ought to match all of these criteria. For example, in Part I of *Theories of Mathematics Education* (Sriraman & English, 2010b), Jeremy Kilpatrick summarized that a model serves more as a description but is not required to provide as much explanation or prediction to phenomena as a theory should do. To Kilpatrick, theories should satisfy all of the above criteria more fully . "Theories must pass a variety of tests; they require verification across situations and circumstances", and he maintained that calling a piece of work a theory instead of an *approach, theoretical framework*, or a *model* is a very strong assertion (Kilpatrick, 2010). On the contrary, pragmatists see theories as tools that are helpful for understanding data. Grounded Theory is a qualitative methodology for generating theories from empirical data (Teppo, 2015). In constructing grounded theories, the main concern is not their universal validity across cases, but their "usefulness in a specific context" (Bryant, 2009. as cited in Teppo, 2015). We can expect a very different set of criteria being used in testing a grounded theory. For example, Glaser (1998, pp. 18-19. as cited in Teppo, 2015) proposed that a grounded theory should be judged by the following four points: fit, workability, relevance, and modifiability. Modifiability is most contradictory with falsifiability in Schoenfeld's point 6, for it is claimed that a grounded theory is "never right or wrong", but "continually modified" to adapt to new data Teppo (2015). Therefore, we can expect a wide range of works that would be called a "theory", depending on various sets of criteria.

A single, unified definition of theory may not be helpful to address the huge variation of notions (Sriraman & English, 2010a), but a few narrower or operational definitions will be handy in learning different research perspectives and traditions. Niss (2007) defines a theory as:

> ... a system of concepts and claims with certain properties, namely
> - The theory consists of an organised network of concepts [...] and claims about some extensive domain, or a class of domains, of objects, situations and phenomena.
> - the concepts are linked in a connected hierarchy [...], in which a certain set of concepts, taken to be basic, are used as building blocks in the formation of the other concepts in the hierarchy.
> - the claims are either basic hypotheses, assumptions, or axioms, taken as fundamental [...], or statements obtained from the fundamental claims by means of formal or material [...] derivation.

R. Lesh and Sriraman (2010) share a similar view:

" Theories are *cleaned up* bodies of knowledge that are shared by a community. They are the kind of knowledge that gets *embodied in textbooks*. They emphasize *formal/deductive logic*, and they usually try to express ideas *elegantly* using a single language and notation system. R. Lesh and Sriraman (2010, p. 143., italics added)"

By contrast, Bikner-Ahsbahs and Prediger (2010) do not follow this *static* definition as Niss provided. They understand theories as: "[...] constructions in a state of flux. They are more or less consistent systems of concepts and relationships, based on assumptions and norms." Both definitions agree that a theory has to contain a more or less organized system of concepts and claims, some of which being more fundamental to the rest. The consistency of the claims, the hierarchy of the concepts, and the degree of organization are simply properties of a theory.

An alert reader could have noticed that the aforementioned claims in a theory are not only prescriptive statements — they can be generated or derived, too. Theories are functional in the sense that they are able to generate new claims when they are placed within their scope of situations, in order to offer explanations, predictions, or to be falsified. Therefore the following view of Assude is adopted:

> [T]heory in mathematics education deals with teaching and learning mathematics from two points of view. First a structural point of view: theory is an organised and coherent system of concepts and notions in the mathematics education field. Second a functional point of view: a theory is a system of tools that permit a "speculation" about some reality. This "speculation" is an active one because these tools can allow to observe, analyse, interpret a teaching and learning reality (or practices), and can produce new knowledge about this reality (Assude et al., 2008, p. 342)

A helpful way to qualify theories in mathematics education is to look at different *aspects* of theories. <u>Grand vs. Local</u>. A *grand* or *universal* theory of mathematics education is one that could affect most of the research in mathematics education, just as how evolutionary theory affects the study of biology (Silver & Herbst, 2007. as cited in Sriraman & English, 2010a). It describes the whole *mathematics education field* (Assude et al., 2008). However, it was pointed out that such theories hardly exist (Kilpatrick, 2010), and are even not desirable (Sriraman & English, 2010a). Instead, *local* theories address a particular zone of inquiry, with the hope that many local theories can be organized together as *networking theories* (Bikner-Ahsbahs & Prediger, 2010). Networking is one of the important techniques to synthesize existing theories, and will be described in the next section. Foreground vs. Background. Mason and Waywood (1996) compared *foreground* theories developed locally *within* the field of mathematics education, and *background* theories *about* mathematics education, as an application of a pre-existing, supposedly consistent theory from outside fields or disciplines. This pair of modifiers should not be confused with *internal vs. external*, both of which refer to theories created to describe some aspects of mathematics education, constructed with elements within the field (internal, or "*home-grown*", as termed by Steiner (1985)) or from other fields (external, or *interdisciplinary*), such as psychology or sociology. Static vs. Dynamic. Bikner-Ahsbahs and Prediger (2010) distinguishes a *static* and a *dynamic* view of theories. A static theory is one that has relatively stable structure, like the one understood by Niss (2007), whereas a dynamic theory welcomes changes as researchers explore more in the empirical territories. Prediger, Bikner-Ahsbahs, and Arzarello (2008) pointed out that even well-developed theories are still in "a state of flux". These four pairs of modifiers constitute a typological framework for us to understand the kind of a given theory in mathematics education. However, one should notice that each of them is not a dichotomy, but rather represents two directions of a linear continuum. For example, local theories may have different ranges of the field of mathematics education, and a global theory covers the widest imaginable range.

The discussions above have illuminated a variety of theories. Less controversial is the definition of a theoretical framework, and here is one provided by Frank Lester (2005/2010). He first defines a *research framework* in general:

> [...] a research framework is a basic structure of the ideas (i.e., abstractions and relationships) that serve as the basis for a phenomenon that is to be investigated. These abstractions and the (assumed) interrelationships among them represent the relevant features of the phenomenon as determined by the research perspective that has been adopted. (Lester, 2010, p. 69)

In other words, a research framework allows researchers to lay out their perspective and structure of relevant conceptions before the actual research work is done. Lester further explained that there are three types of research frameworks, identified by Eisenhart (1991), including theoretical frameworks, practical frameworks, and conceptual frameworks. In particular, a *theoretical framework* is a research framework that relies on an established, formal theory (Lester, 2010, p. 70). It limits a researcher's perspective, vocabulary, and procedures to a school of thoughts. Using a theoretical framework, researchers who work within the same agenda are enabled to communicate effectively about their progress, but Lester pointed out four shortcomings of this approach to research, including (a) that agreement with theoretical "decrees" is often more important than solid evidence, (b) that the context of data is often ignored, (c) that a theoretical standard may not be relevant at all in practice, (d) that there is no "triangulation" in research, that is, by "embedding one's research in a single theory", the researcher may be totally ignorant about other possible perspectives. A *practical framework* is a research framework that does not subscribe to a formal theory, but is informed by "the accumulated practice knowledge of practitioners and administrators, the findings of previous research, and often the viewpoints offered by public opinion". It is a "quick fix" to problem (c) of theoretical frameworks, but it falls short in its generalizability. Lester (2010) advocated the use of a conceptual framework, which is a research framework that also does not subscribe to a single theory, nor does it maintain a list of best practices, but it is a structure of concepts, relationships between those concepts, and "an argument that the concepts chosen for investigation or interpretation, and any anticipated relationships among them, will be appropriate and useful, given the research problem under investigation" (Eisenhart, 1991, p. 209). Lester showed the usefulness of a conceptual framework over the other two by the example of Models and Modeling Perspective (MMP), which has been briefly mentioned in the introduction. According to Lester (2010), MMP "integrates ideas from a variety of theories" (p. 73), and also uses standards that are not subject to those theories; the main purpose of MMP is not to create a grand theory, but a "system of thinking about problems of mathematics learning".

To Lester, a researcher who uses a conceptual framework is analogical to a *bricoleur*, that is "a handyman who uses whatever tools are available to come up with solutions to everyday problems" (p. 74). Nevertheless, even this analogy has assumed that some good tools are available to the handyman, and the handyman himself also has theories of some sort (perhaps usually called *experience*) to guide him in how to apply those tools. Hence, conceptual framework can be treated as a weaker form of theoretical framework that uses some informal theory, either explicitly or implicitly, to guide the researchers' choice and application of more formal theories and methods.

Construction of theories in mathematics education

Much was discussed in *Theories of Mathematics Education* (Sriraman & English, 2010a) about what a theory or theoretical framework is, what theories were borrowed from the other fields, and the role of theories in various topics of study. However, not much detail was provided about how theories were constructed, developed, or validated, with one exception of a chapter Bikner-Ahsbahs and Prediger (2010) of the networking approach, which provides a way to work on the diversity of existing theories. Thus, this subsection will review materials from outside of the field about the methods of constructing, developing, and validating theories. The networking approach of Bikner-Ahsbahs and Prediger will also be reviewed.

Bikner-Ahsbahs and Prediger (2010) pointed out that the development of theories is a demand in mathematics education, and there is a lack of clear descriptions of how theories are developed. They briefly introduced two kinds of theory development processes: (1) An empirically grounded development process requires a "spiral process" between empirical data and theoretical construction, with a background theory as a base theory. (2) A prescriptive process involves argumentations about the connections among theory elements. During such process, empirical data serve as one of the criteria in evaluating the claims and assumptions of the theory. Note that a prescriptive process needs empirical data, and an empirically grounded process also needs argumentations. Therefore, no clear line can be drawn to distinguish the two kinds of processes.

Another description of theory development can be found in a broader context. In a text on social study methods, Bryman (2012) introduced the concepts of deductive and inductive theory. In developing a deductive theory,

> [the] researcher, on the basis of what is known about in a particular domain and of theoretical considerations in relation to that domain, deduces a hypothesis (or hypotheses) that must then be subjected to empirical scrutiny. [... He] needs to specify how data can be collected in relation to the concepts that make up the hypothesis. (p. 24)

After the data collections, the hypothesis is put to a test, which offers a chance to refute the theory. If empirical evidence shows that the theory was not right, it is to be revised. According to Bryman (2012), the revision a theory based on empirical data is actually an inductive step. Thus, a deductive theory development may involve induction. Similarly, an inductive theory development may also involve deduction. In an inductive development,

“ theory is the outcome of research. In other words, the process of induction involves drawing generalizable inferences out of observations. (p. 26)”

The proposed theory is then put to examination, which can be done by generating hypotheses from the theory, and collecting more data to test the hypothesis. Bryman (2012) maintained that this again requires deduction, since hypotheses have to be deduced from the theory. Moreover, although there may be no initial theory stated, the data collection process must be influenced by some principles in the researchers’ mind.

To compare, a deductive approach tend to theorize before a systematic collection of data, and also to guide how data are collected; empirical data serve to test the theory, whereas an inductive approach theorizes *after* a systematic collection of data, for it is assumed that the empirical data imply some generalizable theory. The distinction between the deductive and the inductive (Bryman, 2012)is parallel to that of the prescriptive and the empirically grounded (Bikner-Ahsbahs and Prediger, 2006, 2010). Both the deductive and the prescriptive base their theories on existing thoughts in the field, whereas both the inductive and the empirically grounded search for theories within empirical data. However, as mentioned above, both Bikner-Ahsbahs and Prediger (2010) and Bryman (2012) admitted that the distinction is not required to be sharp, as one way of development may contain elements from the other. Therefore, in the current thesis, theory development will be viewed as the articulation and execution of three types of elements:

- *Basic assumptions* (or axioms), that are accepted by the researcher, and not supposed to be challenged, at least in the argumentations within the theory.
- *Deductions*, i.e. making claims logically implied by the existing claims and assumptions of the theory. The generation of hypotheses, confirmatory analysis of data, and logical argumentations belong to this type.
- *Inductions*, i.e. making general claims that are consistent with the empirical observations. Exploratory analysis of data, initial observations, and (unconfirmed) generalizations belong to this type.

Indeed, I believe that all these three types of elements can be identified from the development of a theory, though some of them may not be explicitly stated as such by the theory's authors or users. Under this assumption, it is harder to identify theories when they are implicit. Procedures of identifying theories from existing literature will be presented in detail in Chapter III.

The two aforementioned approaches are about establishing a theory "from scratch". Bikner-Ahsbahs and Prediger (2010) points out, however, that many of the theories are developed without being aware of other theoretical works, and therefore hinder communication between researchers. If communication between researchers from different theoretical traditions becomes a challenge, new theories tend to be created much more frequently than the appreciation of existing theories. In his review of mathematics education as a whole, Steen (1999) negatively remarked that the field of mathematics education has been "overwhelmed by complexity and drowned in a sea of competing theories." (p. 236). A more satiric comment reflects an even worse picture of the theories in our field:

" Theories are like toothbrushes [...] everyone has their own and no one wants to use anyone else's. (Campbell, 2006, p. 257)"

A proposed way out of this is the networking approach (Bikner-Ahsbahs & Prediger, 2010), which is a recent attempt to deal with the diversity of theories in mathematics education. Since 2005, the idea has been intensely discussed by the Networking Theories Group, who acknowledged that diversity of theories should be treated as a beneficial resource instead of a trouble, and it can be done only if there are "actively established connections" among the theories (Bikner-Ahsbahs et al., 2014, p. 8). The main idea and examples of networking, and the progress of the Networking Theories Group, are expounded in a special issue of *ZDM* (Prediger, Arzarello, Bosch, & Lenfant, 2008), and later as a book (Bikner-Ahsbahs & Prediger, 2014).

For any two or more theories given, Prediger, Bikner-Ahsbahs, and Arzarello (2008) seek to qualify the connecting strategies among them by defining the degree of integration, which is an ordering depicted in Figure 2.3. On one extreme, there is no connection at all. This "strategy" is to ignore other theories, which is a situation that the group of Networking Theory researchers wish to avoid. Such extreme only serves as a reference point, because it will never be "used". On the other extreme is a strategy of "unifying globally", resulting in a grand theory of everything (about mathematics education). This is again a reference point only, since the networking researchers do not believe that such unification is possible in the near future. However, such hope is desired by some, as discussed in the previous subsection. Between the two extremes are networking strategies, which are to be actually applied when researchers try to connect theories in mathematics education. Prediger and Bikner-Ahsbahs (2014) identified four pairs of networking strategies, as listed below in an increasing degree of integration:

- The first step away from ignoring is *understanding other theories*, and *making* one's own theory *understandable* by others.

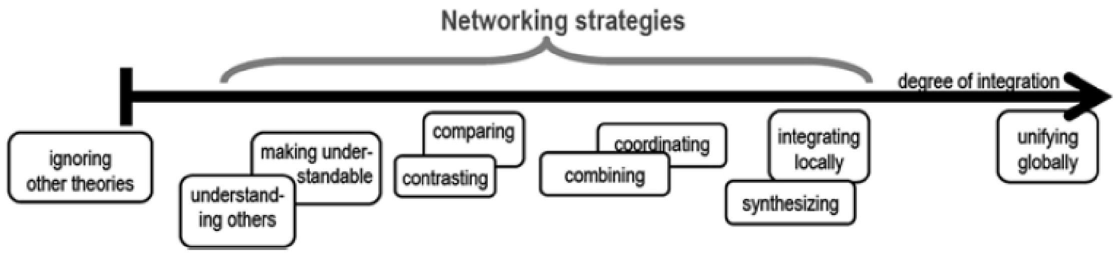


Figure 2.3: Degree of integration for a given connecting strategy (Prediger, Bikner-Ahsbahs, & Arzarello, 2008)

- Based on the understanding of different theories, *comparing* means studying the differences and similarities in general, whereas *contrasting* concentrates more on showing the divergence. Prediger and Bikner-Ahsbahs (2014) maintained that comparing and contrasting help to strengthen the linkage between theories through their similarities, and to illustrate the strengths of individuals theories by the differences. However, a critical reader may argue that the differences could also expose the cavities of one or both theories, depending on the context.
- When both theories are understood deeply enough in the previous stages, it is possible to view the same phenomenon in two theoretical perspectives at the same time. This is called combining if the networking researcher does not seek to make a coherent view, otherwise it is called coordinating. In other words, combining simply produces parallel explanations, whereas coordinating results in a conceptual framework (cf. 2 that seeks to fit common elements of both theories together. Such framework does not mean to be a new theory, since it does not generate claims that are not deductible from one of the two referred theories.
- The three pairs of strategies described above do not really create new theory. However, the next pair of strategies, *synthesizing* and *integrating locally*, mean that a researcher focuses on only a small piece of both theories, creates a new framework out of it, and a new piece of theory emerges as a result. When the two theories are about equally mature in their development, the name synthesizing will be used; otherwise, it is called integrating locally. The word locally was chosen because the synthesis and integration of the theories here should not be mistaken as being done totally.

As Prediger and Bikner-Ahsbahs (2014) admitted, one should not treat these strategies as a model of networking methodology, but rather as an analytical tool to describe the networking process. The networking of theories may involve one or more strategies described above. However, since networking is treated here as a way of constructing theories, only the last two strategies, synthesizing and integrating locally, are taken into consideration.

Validation and falsification of theories in mathematics education

The two previous subsections addressed the issues of what a theory is, and how a theory can be developed in mathematics education. Nonetheless, one needs some criteria to decide whether the theories or the statements of them are valid. Otherwise, as Nesher (2015) put it, "if we have no clear criteria to select among theories, beyond fashion, authority, or 'being nice' to each other, we are doomed to remain with a multiplicity of theories" (p. 144). I agree that the multiplicity of theories does not promote real advance of the field, unless there exist to discern the validness of the theoretical elements so that researchers and practitioners can apply them. In the following, methods to validate theories using empirical data will also be reviewed.

A mathematical proposition can be falsified by a single counterexample, and is validated through a mathematical proof with limited length. However, it is different in science: while a scientific theory is still falsified by one carefully established counterexample (assuming that it is significantly deviated from the theoretically acceptable error), it may never be proved in a mathematical sense. Karl Popper (1972) argued that scientific theories are not possible to be verified, in a sense that it will also work the best in the future. To Popper, theory development is evolutionary (in a Darwinian sense), that one can only eliminate those that have not been performing well, and keep those that work well so far. Theories A and B are said to be competing with each other "in the light of the state of the critical discussion at the time t, and the empirical evidence (test statements) available at the discussion, the theory A is preferable to, or better corroborated than, the theory B." (Popper, 1972, p. 19) In other words, a scientific theory's "survival" (or *verification*) always depends on the current argumentations and the empirical evidences to date.

Traditionally, quantitative research has been associated with theory verification (Punch, 2009) through, for example, hypothesis testing or quantitative meta-analysis. However, our data are sampled from a population of all possible theories and all possible empirical data available from

Table 2.4: 13 tactics for testing or confirming qualitative findings (Miles & Huberman, 1994, p. 263)

Data quality
1. Checking for representativeness
2. checking for researcher effects
3. triangulating
4. weighting the evidence
Unpatterns
5. checking the meaning of outliers
6. using extreme cases
7. following up surprises
8. looking for negative evidence
Test the explanations
9. Making if-then tests
10. ruling out spurious relations
11. Replicating a finding
12. Checking out rival explanations
Consult the origin of the data
13. Getting feedback from informants

a list of journals (cf. Section 3, p. 61), and therefore qualitative methods are consulted. Miles and Huberman (1994) proposed 13 "tactics" for testing or confirming qualitative findings (Table 2.4). The first four deal with the data quality in general. Tactics 5-8 are used when a theory does not match with some data, or when one actively searches for exceptions of a theory; 9-12 are helpful when a theory seems to survive for a while, and thus worthwhile for further testing or falsification. The last tactic is particularly important when the data are collected from human subjects. The tactics 9-13 are more suitable for the verification of a single study or a meta-analysis of similar studies, they do not fit the purpose of the current thesis.

Methodological Background: Comparative Curriculum Studies

Two frameworks for a comparitive analysis of curricula

Bray and Thomas' Cube: comparing education systems in general

The study of curriculum necessitates some degree of comparison, which may be implicit or explicit, but an explicit comparison often provide sharper resolution to discern the topic of interest (Adamson & Morris, 2014). Comparative studies have also been highly relevant ever since the birth of the research field of mathematics education. For example, international comparison of school mathematics curricula was put forth by David Eugene Smith in the beginning of the twentieth century (Kilpatrick, 2008); and, of course, a modern prominent example is the PISA. In order to perform any comparative study of mathematics education, it is important to be aware that the complexity of the entities to be researched is so high, along with their variables and perspectives, that a comparative framework is almost mandatory as a guide for the researcher to obtain any meaningful results.

The framework that is adopted in this study was introduced in *Comparative Education Research: Approaches and Methods* (Mark Bray, Adamson, & Mason, 2014), a state-of-the-art book in this field, published by the Comparative Education Research Centre at the University of Hong Kong (CERC). It is originated from Mark Bray and Thomas (1995), which sought to broaden the field of comparative education. They provided a quite inclusive definition: "comparative education refers to all studies that inspect similarities and/ or differences between two or more phenomena relating to the transmission of knowledge, skills, or attitudes from one person or group to another" (Bray & Thomas, 1995, p. 473). Historically, the typical focus of comparative education was cross-national

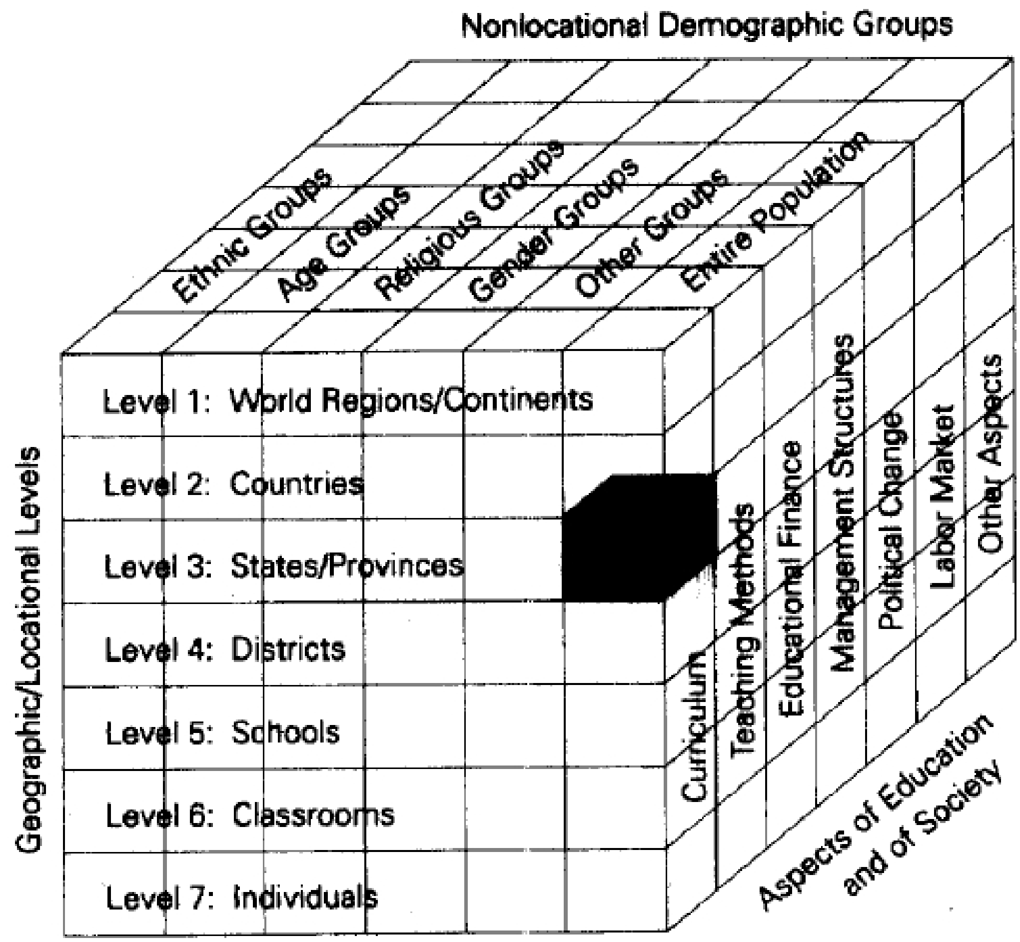


Figure 2.4: The Bray and Thomas Cube

comparison, but that was an unnecessary limitation of scope, according to Bray and Thomas. Thus they built a three-dimensional framework to embrace all kinds of comparative studies that are included in the broad definition stated above. These three dimensions are depicted as a cube as shown in Figure 2.4.

Note that not all three dimensions have the same structure. The first dimension, which was shown in the front of the cube, is *geographic/locational* levels, including Individuals, Classrooms, Schools, Districts, States / Provinces, Countries, and World Regions / Continents. These levels are most commonly seen as hierarchical: classrooms being embedded in schools, schools being embedded in districts, and so forth, but one has to note that some of the hierarchical relationship may not be absolute. For example, individuals could transfer from school to school. In some special cases, there could be integration or disintegration of countries and states (e.g. Germany), and regions in a country that are more autonomous than a state or a province (e.g. Hong Kong). Nevertheless, it is still appropriate to follow Bray and Thomas (1995) to understand this dimension as *multi-level* and to place these levels as horizontal *layers* of the cube. The other two dimensions, *Nonlocational Demographic Groups* and *Aspects of Education and of Society*, are less structural in their nature. Each comparative study has to focus on at least one part of each dimension, and therefore can be located at one of the "mini-cubes". For example, the black mini-cube shown in Figure 2.4 represents studies on the states/provinces level, focusing on the curriculum that is concerned of the whole population. The PISA study described in the previous section is a more comprehensive one that involves a three-level analysis of countries, schools, and individuals, concerning both the curriculum and the teaching methods that have impact on the group of 15-year-olds. Naturally, the current study being described in this section can be located as a sub-location within the range of the PISA study. It is at the countries level, focusing on the curriculum that has impact on the group of 15-year-olds.

The purpose of this cube is not only to locate comparative studies, but also to remind researchers about the limitation of any single study. Bray and Thomas (1995) urged researchers to recognize such limitation, and make use of studies that are from a different level or focus. A further remark regarding this cube is that there is "a further dimension of comparison" across time (ibid, p. 474), and thus a comparative study can be primarily located in these four dimensions.

Adamson & Morris' framework for the comparison of curricula

As a further development of the Bray and Thomas' Cube, the aforementioned book (Bray, Adamson, & Mason, 2014) provided methods and examples for researchers who locate their research in one of the mini-cubes. The most relevant chapter is "Comparing Curricula" (Adamson & Morris, 2014), which provides a framework for the current thesis. In this framework there are three elements: *Purpose and Perspective, Curriculum Focus*, and *Manifestations.* According to their descriptions[1],

> The framework is based on the premise that the inquirer has a purpose, be it utilitarian (e.g. policymaking) or the generation of new understandings. Having a purpose implies the adoption of a perspective. The purpose also informs the question(s) that the inquirer wishes to answer, which in turn would suggest a focal point – an aspect or component of the curriculum – for the inquiry. Data would then be collected from relevant curricular manifestations, which could include documents or behaviours. (ibid, p. 316)

Purpose and Perspective.

The framework allows the researcher to study curriculum with any purpose. The purpose of the current thesis was stated in Section 1.2. Adamson and Morris (2014) defined three different perspectives that loosely categorize any curriculum study. They are the *evaluative perspective*, the *interpretive perspective*, and the *critical perspective.* Researchers take on an evaluative perspective when they "[seek] evidence in order to make informed decisions about the curriculum" (ibid, p. 317). PISA is a prominent example of an evaluative study. A critical perspective is used when one wants to argue or illustrate that some features of a curriculum is desirable or undesirable, usually for the concern of equity, justice, or social reformation. The one that this thesis adopt is the interpretive perspective, in which a researcher "[endeavors] to analyze and explain phenomena" (ibid, p. 319).

Curriculum Focus and Manifestations.

Once the purpose is clear, the researcher needs to select elements or aspects of a curriculum

[1]The description of their framework is applied, but the figure they provided (Adamson & Morris, 2014, p. 270) is not used, for the "interlinks" between each pair of the three elements. Furthermore, in this study, the linkage between *purpose and perspective* and *manifestations* is not analyzed.

that can be compared. Possible elements or aspects proposed by Adamson and Morris include (a) ideologies, societal cultures; (b) curriculum development systems , or "the intended"; (c) curriculum implementation, or "the enacted"), and (d) experiences by the learner, or "the experienced" (ibid, p. 321), and (e) the "null curriculum", that is, important things that are omitted from the curriculum (ibid, p. 322) [2]. These elements or aspects can also be recognized in the field of curriculum studies, and will be described in a more detailed way in the next section, for they are directly related to the analysis. For any of the items (a) to (e), "tangible and intangible manifestations" (ibid, p. 322) are needed through which a researcher may conduct a study. These manifestations are shown in Table 2.5. Tangible manifestations include but not limited to policy and curriculum documents, academic papers, books, teaching materials, lesson plans, assessment materials, minutes of meetings, and notices; intangible manifestations can be teaching and learning experiences reflected from highly subjective and indirect ways such as behavioral responses or interview responses. Tangible manifestations are often easier to obtain; for example, policy and curriculum documents can be found in governmental offices or websites. Intangible manifestations, as they are so named, can only be obtained through indirect and subjective means. Once the three elements of the framework are clear — including the researcher's purpose and perspective, the focus, and the manifestations of this focus — possible research methods must depend on these elements. These methods will be discussed in the next section.

Methods to analyze curricula

This section describes the methods of studying Research Questions 3 in relation to the comparative framework aforementioned. The curriculum focus of the study is discussed, and it is viewed under

[2] A term coined by Elliot Eisner (Quinn & Kridel, 2010; Flinders, Noddings, & Thornton, 1986). Flinders et al. (1986) argued that null curriculum cannot be researched empirically, for one must presuppose a *universe* of curriculum to define what is the *null* in the current curriculum.

Table 2.5: Aspects of Curriculum, adopted from Adamson & Morris (2014, p. 321) and Flinders, Noddings, & Thornton (1986)

Aspect of curriculum	**Typical manifestations**	**Typical research methods**
Ideology	books; academic papers; policy and curriculum documents	discourse analysis
Intended (or Planned)	policy and curriculum documents; prospectuses; teaching materials; lesson plans; assessment materials; minutes of meetings; notices	discourse analysis; interviews
Implemented (or Enacted)	teacher and student action; roles of teachers and students; student interest and involvement; classroom interaction; school interaction; student output	Lesson observations; teacher's log; interviews; Ethnography; activity records
Attained (or Experienced)	change in student attitude and/or behavior; change in teacher attitude and/or behavior; student's cognitive processes	Questionnaires; interviews; Autobiographical narratives; Reflections; psychometric tests
The "null"	(not directly manifested; necessitates a "referent", that is, what the researcher values as educationally significant)	Qualitative methods

a commonly used categorization of curriculum perspectives. Then the possible sources of data are reviewed so as to study the manifestations.

Foci, Perspectives, and Lenses of a curriculum study

To study different curricula, one has to understand that the word curriculum may not have uniform meaning when it is used by different groups of people, and so a researcher is recommended to view curriculum in different ways (Goodlad & Richter, 1966; Adamson & Morris, 2014; Marsh, 2009). Metaphorically, one can look at what constitute "the curriculum"*The definite article does not mean to acknowledge the existence of an "entity" of curriculum, nor does it mean that there must exist an "ideal" of a curriculum which each person see it differently. Instead, it is simply a label for what people see as it refers to. In other words, a researcher of the curriculum does not assume what curriculum is, but tries to observe what content this label refers to according to particular groups of people.* with different *foci* (or emphases) from a certain *perspective*, and through a particular *lens* (ideology).

Mason and Waywood (1996) identified that there are at least seven broad conceptions of curricula, each of which can be treated as a focus of a comparative study (Adamson & Morris, 2014).

These conceptions are:

1. curriculum as classical heritage — highly-honored subjects in a particular culture;
2. curriculum as established knowledge — academic disciplines as the basis of students' choices;
3. curriculum as social utility — subjects chosen to be the most useful for life in a modern society;
4. curriculum as planned learning — outcomes that are planned, including the subjects taught, and also competence such as critical thinking;
5. curriculum as experienced learning — all that the learner experiences, including the planned or unplanned, and the values that may be implicit to both the teachers and the learners;
6. curriculum as personal transformation — a focus on how both the teachers and the learners are shaped through the involvement of teaching and learning activities;
7. curriculum as life experiences — a part of life that cannot be analyzed alone. Educational institutions are not the sole focus but the whole of life experiences is.

Notice that the first three conceptions are mostly subject-based, and the latter ones are broader. The broadest of these is conception 7, which simply views anything that occurs to students' lives as "curricular". The focus of the current study is best identified as (4) *curriculum as planned learning*, since the intension of the research questions is to analyze what decision makers or action takers did in order to improve students' ability to apply mathematical knowledge. However, a researcher should also seek to understand in each curriculum the kind(s) of conception which the stakeholders take.

Once there is a focus, a researcher needs to consider the perspectives of curriculum (Goodlad & Richter, 1966; Klein, Tye, & Wright, 1979), which usually includes the *intended curriculum*, the *implemented curriculum*, and the *attained curriculum*, and even more. In the scholarship of curriculum studies, the perspectives of curriculum are also known as *perceptions of curriculum* (Klein, Tye, & Goodlad, 1975), *levels of curriculum analysis* (Porter, 2006), *venues of curriculum inquiry* (Schubert, 2008), and *typology of curriculum representations* (van den Akker, 2003), with different variations, which are shown in Table 2.6. The distinction of these different perspectives essentially reflects a fact that different group of people see "the curriculum" quite differently (Goodlad & Richter, 1966). More precisely, what one concerns about the curriculum depends on the role one takes in the system of education, and that is why there are different perspectives. These perspectives have been used as a framework of analyzing a particular curriculum. For instance, *A Study of Schooling* was a project directed by John Goodlad in the 1970s in order to enhance understanding of the school system in the USA, and part of the project was to study the curriculum. Scholars in this project had provided a thorough introduction of five different perspectives (cf. Klein, Tye, & Goodlad, 1975, and Klein, Tye, & Wright, 1979). The following is a brief description of them:

> The ideal — the scholar's view of what ought to be; the formal — the expectations and values about education and curriculum of those interested and concerned people outside the classroom; the instructional — the teacher's beliefs, attitudes, and values about curriculum; the operational — the curriculum being implemented in the classroom as seen by a trained observer; and the experiential — the curriculum which the student experiences.
>
> The experiential curriculum includes two different facets: (1) student perceptions of the curriculum that is offered to them and (2) what is actually learned. (Klein, Tye, & Wright, 1979, p. 244-245)

Following the "typology" provided by Van den Akker (2004), the ideal and the formal (i.e. the written) curricula constitute the intended curriculum, representing goals and standards of teaching and learning, which is usually conceived by scholars, educators, and decision makers. The implemented curriculum includes the instructional and the operational, representing things actually done in the education system in order to deliver the teachings. The attained (or the embraced) curriculum includes the experienced the learned, which directly match the two facets of the experiential curriculum. According to Pollak (2005) [3], the learned curriculum is also associated to the assessed curriculum and the expected curriculum; the former refers to the student outcomes, and the latter refers to the parents', the employers', and the society's expectation of what students should have attained after certain levels of education. Each of these curricula could present a different picture of "the curriculum". Such differences are considered to be major sources of conflict

[3] in an unpublished paper, Henry O. Pollak (2005), "The Beginnings of a Mathematical Model for Mathematics Education", in *Mathematikunterricht im Spannungfeld von Evolution and Evaluation.*

within the education systems, and therefore one of the endeavors of the educators is to minimize them.

A way to explain the discrepancy among these curricula is to introduce concepts like the hidden (or implicit) curriculum and the null curriculum. The hidden curriculum refers to things that schools teach but are not expressed as a subject or parts of the official curriculum. For instance, schools (implicitly) train students to learn compliance, deferring gratification, competition through grades or awards, and so forth (Schubert, 2008). An alert reader should notice that these aspects of the "hidden" curriculum can be absorbed into the attained curriculum, especially the learned curriculum, though it could well be something most students learned but not being explicitly described in the written syllabi. The null curriculum refers to topics that are "minimized or excluded due to priority and budget" (Schubert, 2008, p. 412), or simply "what schools do not teach" (Flinders et al., 1986, p. 33). According to Flinders et al., the concept of null curriculum could only serve as a political tool if the "universal" curriculum is not defined. One cannot define the null curriculum "with operational precision unless we are willing to risk triviality" (ibid, p. 41). Nonetheless, in the context of the current thesis, one may define the universal as any topic, competence, or attitude that is related to mathematical modeling or applications.

The "foci" and the "perspectives" of the study help limiting the scope of the study and clarifying the various understandings of curriculum from different groups of people. However, the researcher also needs to recognize that there is always a "lens" through which different groups of people understand and interpret the curriculum. Such "lenses" are actually ideologies "that are underpinned by normative views and beliefs about the desired role of schooling in society, the nature of knowledge and learning, and the roles of teachers and learners." (Adamson & Morris, p. 312) According to Adamson and Morris, at least six ideologies to understand curriculum are identified. They are (1) *Academic rationalism*, (2) *Social and economic efficiency*, (3) *Social recon-*

Table 2.6: Alternative names and categorizations of perspectives of curriculum

Scholars	Alternative names	Categories
Klein, Tye, & Goodlad (1975)	perceptions of curriculum	The ideal, the formal, the perceived, the operational, and the experiential
Klein, Tye, & Wright (1979)	perspectives of curriculum	The ideal, the formal, the instructional, the operational, and the experiential
Eisner (1979)	The Three Curricula that All Schools Teach	The explicit, the implicit, and the null
Eisner (1994); Uhrmacher (2004) (both cited in Jones, 2012)		The operational, the intended, the received, The hidden, the explicit, the implicit, and the null
van den Akker (2003)	A common, three-level distinction, and its refinement; typology of curriculum representations	The intended, including the ideal and the written; the implemented, including the perceived and the operational; the attained, including the experienced and the learned
Schubert (2008)	Contemporary Venues of Curriculum Inquiry	The intended, the taught, the experienced, the embodied, the hidden, the tested, the null, and the outside

structionism, (4) *Orthodoxy*, (5) *Progressivism*, and (6) *Cognitive pluralism.* Some of these ideologies co-exist in the same group of people; some compete with each other.

Chapter 3

PROCEDURES

In Chapter 1, two sample problems showed that some theoretical framework is needed in order to study how a test item addresses the ability to apply. This chapter will be divided into two sections. The first section presents the details of the procedures for Research Question 1, including the way to identify theoretical elements from existing literature, and to build a framework using these elements. The second section presents the way assessment items and related documents are analyzed, and is used for Research Questions 2 and 3.

Identifying and Synthesizing Theoretical Elements

Method of Tsatsaroni et al. (2003)

In order to "extract" theories from the literature, a systematic literature analysis is applied, under the influence of a method proposed by Tsatsaroni, Lerman, and Xu (2003). A literature analysis can be applied to a larger number of articles than a literature review, and the researcher purposefully creates some framework during the process. An unthoughtful but first approximation to the problem is simply to do a word search of "theory" throughout the published articles, but there are at least two problems. First, there exist articles which actually use some theories, but they did not use the word theory. It is safer to say that we are going to extract theoretical elements from those articles, when they are not identified by the authors as theories. Second, when an article cites some theory, the extent that such theory is connected to the actual study may vary a lot. These issues were discussed in Tsatsaroni et al. (2003).

The project undertaken by Tsatsaroni et al. was to explore and to make sense of the changes of the field of mathematics education research, and how these changes influenced practices in schools. Their basic assumption is Bernstein's (1990/2003) theory of the discourses of education. Based on his analysis of the modes of education in UK, Bernstein distinguishes between two discourses in the field of education: (a) the discourse of "education" in general, that is, a body of knowledge supported by various disciplines (e.g. political science, sociology, psychology); (b) the discourse of professional subjects such as mathematics. Bernstein also depicted six stages of development over time. A complete description of the six stages is out of scope in this chapter, but the interested reader may read Bernstein (2003, pp. 139-141). The vision that these two discourses are undergoing a certain pattern of transformation is the reason why Tsatsaroni et al. (2003) seek to trace the changes of our field. Their method was applied to a sample of articles from journals and conference proceedings, and the steps are summarized as follows (also see Figure 3.1 on p. 63):

Step 0: Take an article. Record the basic citation information.

Step 1: Ask whether the authors used any theories or not. If so, what theory did they use? Did they do it explicitly or implicitly?

Step 2: What is the orientation of the paper, namely, towards the empirical or towards the theoretical? The former intends to develop a theory, even if it draws on some empirical observations or evidences. The latter's main purpose is to describe and/or inform school practice, policy, and so on; theories may be applied but not developed.

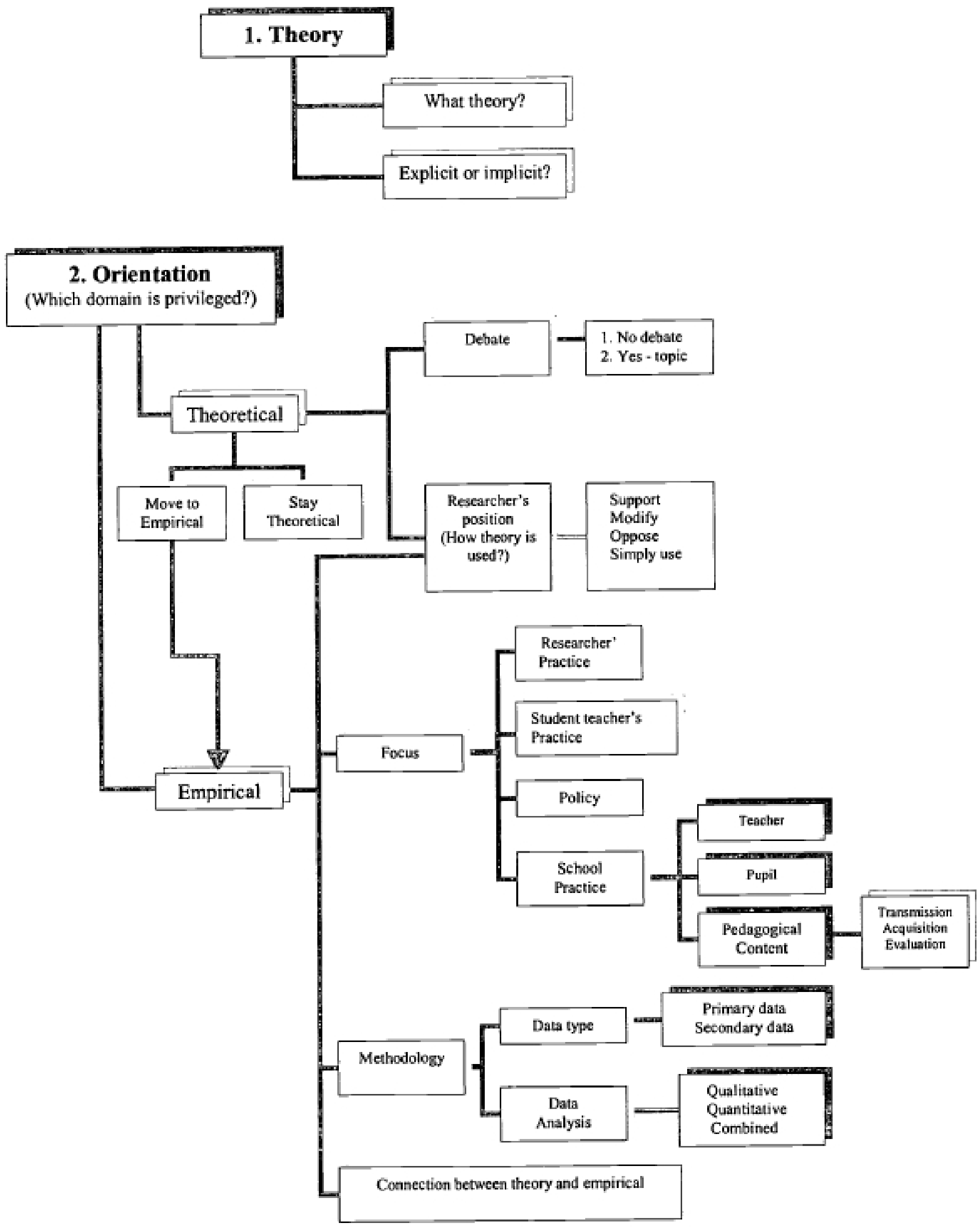
Step 3: How are theories used? Are they modified or supported? Are there any theories borrowed from other fields?

Step 4: If the orientation is towards the empirical, then what is the focus, such as policy, school practice, or researchers' practice?

Step 5: What is the connection between theory and the empirical? Namely, does the theory inform the empirical or supported by the empirical, or are their some conflicts between them?

One has to note again, in step 1, that if theories were used implicitly, it is not straightforward to "record" the hidden theory. In the analysis of Assude et al. (2008), Herbst noticed that some researchers "shy away" (p. 341) from the use of the word *theory*. Some other authors may be less

Figure 3.1: Method of Tsatsaroni et al. (2003)



The figure shows steps one and two of the flow chart that represents an analytical tool to identify theories from literature of mathematics education reserach. Adapted from *A Sociological Description of Changes in the Intellectual Field of Mathematics Education Research: Implications for the Identities of Academics.* (p. 41), by Anna Tsatsaroni, Stephen Lerman, and Guo-Rong Xu. Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 21-25, 2003).

“self-conscious” (p. 40) about the use of theory. As a remedy, Herbst did key word search on the strings “theor,” “framework,” and “construct”. Such method certainly increases the chance (and workload) in identifying hidden theories from academic articles, but its effectiveness is unknown. Another possibility to identify hidden theories is to look for the assumptions made by the researchers, and whether these assumptions are similar to other authors’, and to those of other theories. Again the assumptions are usually implicit, but the hope is based on the belief that any research study must bear some basic assumptions, and that most of them can be made explicit.

Besides the use of theory and its connection with the empirical data, Tsatsaroni et al. (2003) also recorded the research aims, “ideological affiliations”, and “pedagogical models” of the articles they had surveyed. These are not directly applicable to the current research, and therefore will not be detailed here. However, there is one surprising remark: in their analysis of the research aims, they mentioned that “the spread of mathematical topics is very wide”, but modeling and application was not present in their list. Their sample actually contains a number articles on modeling and application, which include Galbraith and Clatworthy (1990), Blum and Niss (1991), Tanner and Jones (1994), and Gravemeijer and Doorman (1999), to name a few. An inspection of the topics of the survey suggests that modeling and application may have been classified into *Problem Solving* or *Mathematics in General.* The reason why modeling and application is not one of the categories is uncertain, thus there is a need to reexamine the sample.

In the following, we adopt the procedures of Tsatsaroni et al. (2003) with some adjustments. These procedures will be used in Chapter IV in order to identify theoretical elements about the ability to apply mathematical knowledge from existing studies.

Procedures for the Extraction of Theoretical Elements

Defining the "Population": Initial lists A and B

First of all, the purpose is to respond to Research Question 1(a), i.e. to collect current theories, and then to compare them with the existing data of empirical cases. *Current theories* are defined as the theories or theoretical elements in the past ten years (2006-2015) that are related to the ability to apply mathematical knowledge. The "population" of the survey is then defined as the publications from 2006 to 2015 that contain theories or theoretical elements about the ability to apply.

These publications can be found in academic journals, conference proceedings, treatises, and dissertations. Possible academic journals to be surveyed are *Educational Studies in Mathematics* (ESM), *Journal for Research in Mathematics Education* (JRME), *Journal für Mathematik-Didaktik* (JMD), *Journal of Mathematical Modelling and Application* (JMMA), *Mathematics Education Research Journal* (MERJ), *Teaching Mathematics and its Applications* (TMA), and *ZDM*. The conferences include ICTMA, International Congress on Mathematics Education (ICME), and PME (Conference of the International Group for the Psychology of Mathematics Education). The treatises are included in *The Advances in Mathematics Education* series, *ICMI Studies* series, and *Mathematics Education Library* series. Dissertations can be obtained from individual databases of the corresponding universities. A preliminary search in the 7 journals yields about more than 2000 entries from 2006 to 2015, but some of them are editorials or book reviews, which are going to be excluded. From 2006 to 2015, 17 conference proceedings and 42 treatise volumes are published. We did not estimate the number of dissertations in this period[1]. It is not straightforward to separate articles that are related to modeling or not, hence all possible publications are collected (conceptually) as a list

[1]Note that dissertations are not included in the sample, due to the potentially large size of the corpus. However, dissertations are important source of theoretical claims and empirical cases, and hence a further research is beneficial. Also note that random sampling is generally not preferred here, because each publication is a creative product that may contain some elements that are very different from all other publications.

known as *Initial List A*.

Another part of the study seeks to compare existing data of empirical research with the current theories. *Existing data of empirical research* is defined as *all* published results of empirical studies on mathematical modeling and/or application. These publications provide a "pool" of empirical cases that may support or falsify any given theoretical element that is about the ability to apply. This hypothetical collection of articles is known as *Initial List B*.

Table 3.1 shows the number of available entries from the Initial List A from each type of publication. These citation entries can be found from academic databases as electronic citations, or from official websites as PDF files. The number of entries in the Initial List B is not empirically counted, but it can be expected to be much more, since it also includes publications before 2006.

Dataset A: A searched list of citations

Further criteria are also needed in order to obtain a more accurate collection of references. For example, it is only necessary to include articles and the chapters that are related to the ability to apply mathematical knowledge. Each article or chapter is to be classified into one of the following types (T stands for Type), but one should note that only T1 and T2 are relevant for further study:

- T1, *Ability to Apply*: directly about the ability to apply mathematical knowledge, including assessments;
- T2, *Teaching and Learning*: not T1, but studies the teaching and learning of application;
- T3, *Application Non-teaching*: not T1 or T2, but still studies modeling and application of mathematics;
- T4, *Non-application*: none of the above.

We expected many articles would be Non-application (T4) and thus decided to perform a keyword search on Initial List A. If an entry responses negatively to the search condition "appl* OR model*" in the titles or in the abstracts, then it is reasonable to classify it as T4, since the terms *modeling* and *application* are widely used by mathematics education researchers since 2006. Note that this may not apply to the earlier articles since there are still other traditions in the study of very similar concerns, such as *word problems*, *vocational mathematics*, *realistic mathematics edu-*

Table 3.1: Summary of Initial List A

Journals	Available publications (2006-15)	specialized on modeling
ESM	Vol. 62-89, 654 articles	Vol 86(2), 10 articles
JRME	Vol. 37-46, 272 articles	no specialized publications
JMD	Vol. 27-36, 184 articles	Vol 31, 14 articles
JMMA	published only 2009-14, 62 articles	all
MERJ	Vol. 18-27, 259 articles	22(2), 11 articles
TMA	Vol. 25-34, 240 articles	all
ZDM	Vol. 38-47, 786 articles	38(2) and (3), 19 articles
Conferences		
ICTMA	4 proceedings (2007, 2010, 2013, 2015), 112 articles	all
ICME-11 (2008)	Official Website and other documents, 723 PDF title pages	TSG 9 and 21
ICME-12 (2012)	One published proceeding, 80 articles	TSG 5 and 6
PME	6 Proceedings, 2,752 PDF title pages	NA
PME-NA	8 Proceedings, 2,647 PDF title pages	NA
Treatises		
AME	14 Books, 401 Chapters	no specialized publications
ICMI Studies	9 Books, 268 Chapters	ICMI-14, 59 chapters
MEL	18 Books, 206 Chapters	no specialized publications
Total	9,646 entries	

cation, *functional mathematics*, *transfer of knowledge*, and *mathematical literacy*. For the sake of completeness, one more step is taken: a search the conjunction of these key terms is done on the non-respondents of "appl* OR model*". If an entry responses negatively to both searches, then it is labeled as T4; otherwise, the citation information of resulting articles are recorded, including the *authors*, *published year*, *title*, and the *volume of the publication*.

Within the 9,646 entries in the Initial List A, only 3,524 of them are available as electronic citations, and the rest of them are PDF title pages. Practically, the treatment of PDF pages are inconvenient comparing to the electronic citations, and thus not fully analyzed (See Appendix A5: Partial treatment of PDF files, p. 219). We divided the 3,524 electronic citations into two parts: One part for articles that were published in journals or conference proceedings under specialized title on modeling or applications ("specialized publication"); another part for the rest of the articles ("non-specialized publication"). Among the 7 selected journals, only *TMA* and *JMAA* bear modeling and application on their titles. Nonetheless, as a topic of rising interest, other journals also had special issues that focused on modeling and application, including *JMD* in 2010 (Volume 31), *MERJ* in 2010 (Volume 22 Issue 2), and *ZDM* in 2006 (Volume 38, Issues 2 and 3). Conference proceedings like ICTMA are of course included. *ICME* has topic study group on modeling and application since its fourth meeting, and *PME* also has a lot of research reports about the topic. As a result, relevant parts of *ICME* and *PME* proceedings are included. [2] One may refer to 3.1 again to see the number of articles available under specialized publications on modeling or applications.

As a result, there are 527 specialized publications and 2,997 non-specialized ones, where 496 (94.1 %) and 711 (25.8%) of them, respectively, are found using keyword searches of "model*", "appl*", and other alternative keywords related to modeling and application. (See Appendix A1, p. 210 for details.) Only these references, about 36% out of all, are kept in Dataset A.

Dataset B: Specialized Publications on Modeling and/or Application

Given the large number of entries in Dataset A, it can be expected that from the Initial List B there would also be a lot of citations, as they are not restricted from 2006-2015. Therefore, to keep the task manageable, we only consider empirical cases from 2006 to 2015. Although many earlier empirical studies were anecdotal as compared to the later, more carefully studied ones, we still need to note that the exclusion of articles before 2006 is a limitation to the study, for empirical cases do not have to be "current" as theories are. In the empirical "pool", we also only consider specialized publications. Database B is thus defined as publications from 2006-2015 that are specialized in modeling and/or application, which is a subset of Dataset A. The Types 1 to 4 also apply to Database B, since the empirical studies are also contained in T1 and T2 by definition.

[2]Unfortunately, electronic citations of the articles in *PME* proceedings are not available to the author.

Classifying the Articles

Given the reference *R* of a targeted article , the procedure of analysis can be broken down as follows:

- *Classifying the Articles:* Each article is classified into one of the four types (T1 - T4). Articles of T3 and T4 are not analyzed further.
- *Examining Theoretical Elements and Empirical Results:* Each article of T1 and T2 is examined, seeing if it contains any theoretical element about the ability to apply. We also check each article to obtain empirical results about modeling and application. Connection between theoretical elements and empirical studies is also analyzed.
- *Evaluating and Selecting the Elements:* "Candidate theories" are selected for the purpose of framework building.

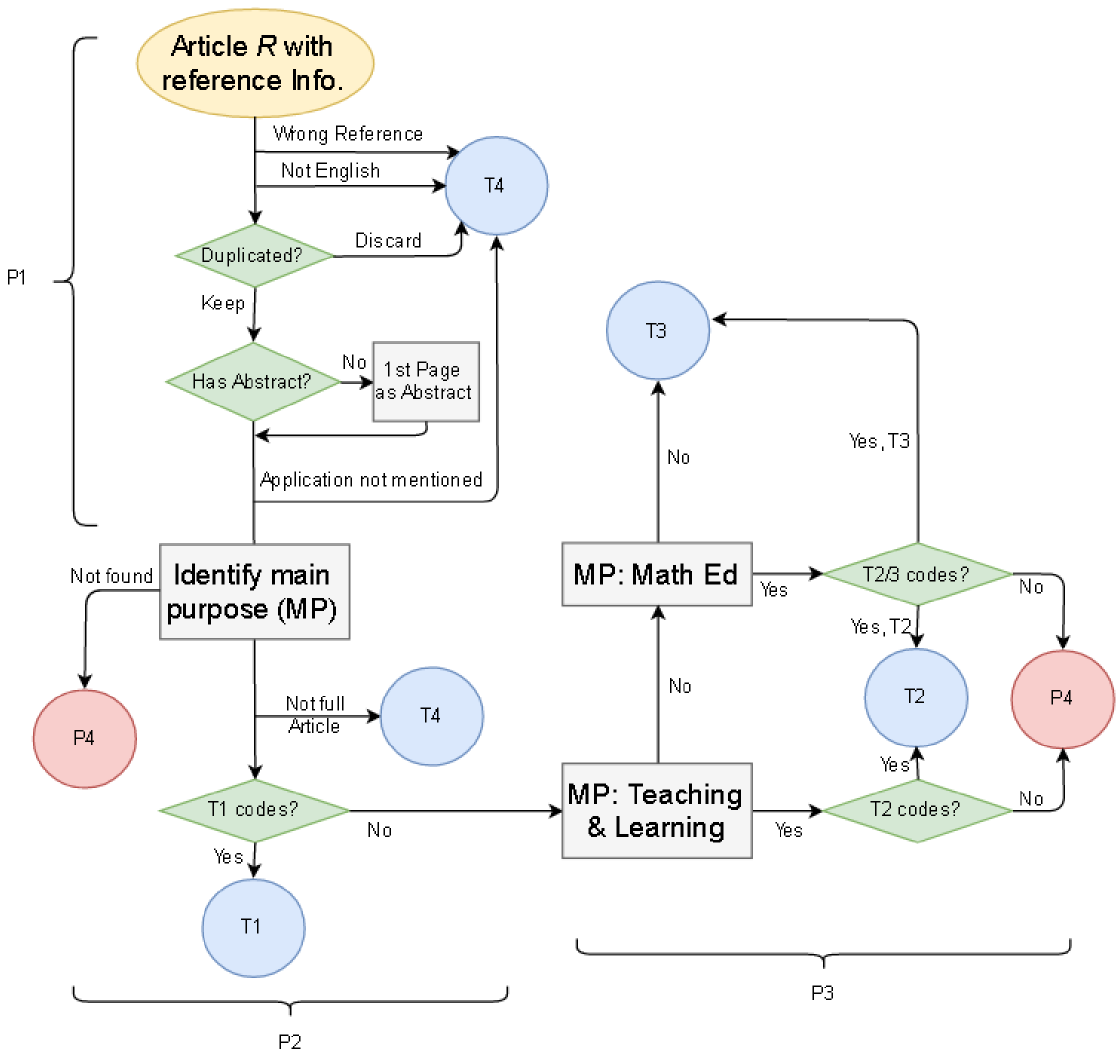
In the method of Tsatsaroni et al. (2003), they asked questions for each article, and encoded them according to the answers. In our case, since there are much more articles, a combination of questioning and keyword search will be used in order to extract the theoretical elements and the empirical cases. We start with classifying the articles.

The descriptions of the four types of articles are mutually exclusive statements; however, operational definitions are necessary to identify an article as one of these types consistently. Each article is assigned for a type and also a code that contains more information. The actual process of classification is based on the operational definitions, along with ways to assist the actual process of identification, such as keyword searches. Finally, some descriptive statistics of the results is presented.

Given the reference of an article *R*, which is one of the 1,273 *EndNote* entries in Dataset A, the operational definition of article types can be illustrated by a flow chart, as shown in Figure 3.2, and is presented under headings P1 to P4 as follows.

P1. Threshold check. Check the reference *R*'s language, citation information, and the abstract in order to check if it may be classified as T4 (Non-application). If the abstract is not available in the publication, the first page of the article will be treated as its working abstract. Also, assign a

Figure 3.2: Article classification procedure



code from the Code List [3] to the reference to indicate the reason why it is classified as T4: it may be a wrong reference, an article not in English; it may also not mention application of mathematical knowledge in both its title and abstract. All of these are classified as T4, for they are not relevant to the study.

P2. Main purpose check. We attempt to identify the main purpose of the article *R* through its title, and phrases in its abstract such as "this study … ". The purpose could also be expressed as a question that is to be answered. If the main purpose or *R* cannot be identified, then go directly to P4. Special Cases. There are cases when *R* is merely an editorial or a note, which only serves as brief introductions to other articles within the same issue of the publication, and is not considered a "full" article. In such cases, it is classified as T3. If it is a full article of which the main purpose can be identified, we check the codes of T1 (Ability to Apply) in order to see if it is to report or analyze ways to measure the ability to apply mathematical knowledge, empirical studies of human subjects' possession of such ability, or to provide theoretical descriptions of such ability. To expedite the process, we search for keywords such as "competenc*", "ability", "assess*", and so forth [4]. If the keywords are found in the title or the abstract of the article, and if they are indeed related to T1 codes (Appendix A2, p. 212), we classify the article as T1. If the article cannot be classfied into T1, then it is not directly related to the ability to apply. Still, the next check is necessary to see if the article implies some assertions about such ability.

P3. Implication Check. Since the main purpose is not the ability to apply, we read the rest of the abstract, including the background, methodology, and significance of the research. The process is guided by the following questions: (a) Does the article address teaching and learning of mathematics in its main purpose? If so, the article is likely to be classified as T2, and one of the

[3] See Appendix A2 for the list of codes.

[4] See Appendix A1, p. 211, for a complete list of keyword searches that has been used in the procedure.

T2 codes in the Code List (Appendix A2, p. 212) can be assigned, or otherwise the article is sent to "P4. Special Cases". Keyword searches such as "teach*", "learn*","develop*" are done to expedite the process (see Appendix A1, p. 211). (b) If the answer to (a) is negative, does the article address at least an issue of mathematics education in general? If so, we first check if the article can be classified as "T2: modeling process", meaning that it contains a study of the modeling process, or "T2: tool", meaning that it presents the usage of a technology or artifact that relates to the application of mathematics. Otherwise, the article is likely to be classified as T3, and one of the T3 codes in the Code List can be assigned, or otherwise the article is sent to "P4. Special Cases".

P4. Special Cases. If the article has not been classified after P1, P2 and P3, then either the main purpose is not clearly identified in P2, or the article is not clearly identified as T2 or T3, though it is related to mathematical modeling or application. These special cases are discussed on an individual basis.

Examining Theoretical Elements and Empirical Results

If an article is classified as T1 or T2, then it is further reviewed through these questions: Does the article use or build a theory explicitly (i.e. the word *theory* is used)? If so, what theory is it? If the article builds a theory, what approaches (i.e. inductions / deductions / networking) does it use? What are the basic assumptions? What inferences, hypotheses, predictions, or explanations were made by this theory? If a theory is not *explicitly* used, look for the inferences, hypotheses, predictions, or explanations that the article made; we record the assumptions, frameworks, and constructs that are employed. In this case, they are counted as *implicit* theoretical elements. If these are still non-existent, then the article contains no theoretical element at all.

We also check if the article contains empirical results on modeling and/or application. If so, record the basic information of the sample studied, a brief description of the instrument used, and summative statements of the findings. Basic information includes the country (or geographical area), number of subjects, age, and level of education. A description of the instrument must record the actual task that is done by the subjects.

How does the article connect the theory and the empirical? Is it *towards the empirical* or *towards the theoretical*? Within the article, is the theory applied to an empirical case (when a theory is present)? Is the empirical study supported by theoretical reasoning, or generalized as some theoretical statement? (when an empirical case is present)? If both theory and empirical case are present, are they consistent or contradictory?

The results of the examination of theoretical elements and empirical results are organized as columns of data. These columns can be attached to Dataset A, but only applied to T1 and T2 articles:

- Type (1-4)
- Presence of theory (use / build / implicit / none)
- Approaches (a list the inductions, deductions, or networking strategies)
- Assumptions
- Functions of theory (a list of inferences, hypotheses, predictions, or explanations)
- Empirical cases (a list of cases with the information of the sample, instrument, and findings)
- Connections (when both theory and empirical cases are present):
    - towards the empirical / towards the theoretical
    - how theory is applied to an empirical case
    - how an empirical case is supported and generalized
    - a list of consistencies and contradictions

Evaluating and Selecting the Elements

We select some "candidate theories", which are to be further tested through the empirical results. The candidates will be selected according to Schoenfeld's criteria (see p. 39). Before the selection process can be started, some preparation of data is needed. First, a Dataset C is created, listing each explicit theory elements, followed by the implicit ones. We also include the article citations (one or more), and basic assumptions. We then add eight columns, which correspond to Schoenfeld's eight points:

(1) *Descriptive power* is evaluated by answering the question, "Is anything (important) missing?" Before using Dataset B, this can only be done when the article uses the theory to describe some realistic case. We record any important feature that the theory does not describe. If no realistic case is used, then record it as "NA".
(2) *Explanatory Power* is shown when a theory provides mechanisms in order to explain how or why some phenomenon happens. We record the phenomena that the theory explained in detail.
(3) *Scope* will be recorded as the type of phenomena the theory covers. For example, a theory about word problem solving may only deal with how students solve story problems with unique answers, but does not deal with open questions.
(4) Initially, *Predictive Power* is simply the list of predictions in Dataset A. We cannot evaluate the predictive power until the theory is put into tests.
(5) *Rigor and Specificity* asks if the theory is well-defined, namely, are the elements of the theory clear enough so that another researcher in the field can identify the same concepts and relations specified by the theory? This can only be evaluated when a survey is also done among scholars or experts in the field, but it is out of the range of the current study. Thus, this criterion will be omitted.
(6) *Falsifiability*, as described in 2, is indeed strongly related to (5). After all, if the theory is not specific enough, it can hardly be falsified. However, it is also related to (4), since a falsifiable theory must be able to formulate clear predictions and testable hypotheses. For each theory, we will record the number of testable predictions and hypotheses that it made. If the number is 0, then the theory is hardly falsifiable.
(7) As Scheonfeld (2001) mentioned, *replicability* is also closely related to (5), and again it cannot be fully evaluated unless another survey is done among scholars or experts in the field. However, if the predictions or hypotheses generated by the theory have been successfully replicated by other researchers, it is a sign of good replicability. Such replications, if exist, will be recorded.
(8) Finally, *multiple sources of evidence* is needed since there may be some variables in the circumstances that the researcher was unaware of. We will record if similar claims are observed in more than one circumstance, and if so, record the variations. For example, do students work in groups or alone? Are they working on the tasks during class time or out of school?

After this, Dataset C will have the following columns:

- Theory index number (starting from 1)
- Citations (a list of Authors and years inherited from Dataset A)
- Assumptions (inherited from Dataset A)
- Descriptive power: Missing features
- Explanatory power (a list of phenomena that the theory explained)
- Scope (a list of different types of phenomena that the theory covers)
- Predictive power (a list of predictions inherited from Dataset A)
- Falsifiability (number of testable predictions and hypotheses)
- Replicability: Existing replications (a list of studies done by other researchers in similar ways)
- Multiple Sources of Evidence (a list of studies done when circumstances are varied)

From Dataset C, we select some (no more than 10) candidate theories to be tested: the five most popular (defined by the number of authors that cites the theory), and five most falsifiable (defined by the largest number of testable predictions and hypotheses). From Dataset B, which is a subset of Dataset A (see p. 68), the empirical cases are indexed separately as Dataset D. Each empirical case should be related to one of the phenomena that is explained by some theory listed in Dataset C (according to the Explanatory Power column). We organize the phenomena in Dataset C into different kinds, then for each empirical case in Dataset D, we record the kinds of phenomena that it is related to, so that we can link it to the candidate theories efficiently. The linkage is done as follows. For each kind of phenomena p, we create a subset of empirical cases c in Dataset D that are related to p. A simple random sample ($n = 5$) is drawn from each subset. For each candidate theory t, if the scope contains K different kinds of phenomena, then the theory is tested throughout all the K subsets. There are three possible results for each triple (theory t, phenomenon p, empirical case c): The empirical case c may be (1) supporting all the theoretical statements on the kind of phenomena, (2) contradictory to some of the theoretical statements, (3) no connection to that theory. The process of analysis and the resulting framework are reported in Chapter IV.

Analyzing curriculum materials of USA, Germany, and Hong Kong

Out of the six PISA assessments done since 2000, PISA 2003 and 2012 are the only two years when mathematical literacy was the main study area. All of the 41 economies that participated in PISA 2003 also did it in 2012, so it is possible to compare the results to investigate trends, and also look into the education systems and their curricula in these participating economies. The author of this thesis is able to read Chinese and English, and also has basic understanding of German with the help of a dictionary. Hence the author could choose from the following 9 economies in order to perform a comparison in their documented curricula:

- English: Australia, UK, USA, Canada
- Chinese: Macau
- English and/or Chinese: Hong Kong
- German: Austria, Germany, Liechtenstein

Out of these 9 places, Germany, USA, and Hong Kong are chosen because the official documents are readily available (convenience sampling), though the others are equally valuable for a comparison. Since there is a great difference in the education systems among these economies, it is not feasible to draw conclusions about the effectiveness of their mathematics curricula in the current thesis. Nonetheless, it is still tractable (and beneficial) to investigate traces of issues that the stakeholders cared about, and the related efforts that were carried out through the curricula. This subsection briefly discusses the data sources that allow us to trace such issues and efforts in each chosen economy from 1991 to 2012. This time frame chosen is 1991 to 2012 because all students who took PISA 2003 were 15 years old, and in 1993 they were 5 years old, no later than the starting age of compulsory education in all three places. Two more years are included before 1993 since some issues and efforts that apparently terminated may still have effect on the systems after 1993.

Hong Kong

As a PISA participant, Hong Kong is a very small place, where the size is between Los Angeles and New York City. It had been one of the British colonies until the 1997 Handover to the People's Republic of China, becoming one of her two *Special Administrative Regions*. Hong Kong's education system is neither dependent on the Great Britain nor the Mainland China, though a lot of cultural similarities can be found from both of them. Since compulsory education had been established in the 1970s, a unified curriculum and examination were available, though the curriculum was implemented differently in the diverse schools. Before 1997, since changes were expected after the historical Handover, there was not many changes in the curriculum until the major renew in 1999. Hong Kong secondary pupils also saw a structural change in 2009, when the 3-2-2-3 scheme (i.e. a British system of 3-year Junior High, 2-year Senior High, 2-year Advanced-Level Programs, and 3-year College) is rearranged to the 3-3-4 scheme (i.e. 3-year Junior High, 3-year Senior High, and 4-year College). Consequently, one more year is added to the compulsory secondary education, and the optional Advanced-Level programs became obsolete in 2012. The more unified Hong Kong Diploma of Secondary Education examinations were first introduced in 2012. Like other East-Asian countries, students and parents in Hong Kong paid most attention to examinations, and therefore it is convenient to treat assessments as a guiding piece of information to study their mathematics curriculum. A sample of curriculum materials attainable is listed in Table 3.2, page 79.

USA

The education system of United States is decentralized, since each state has its own legislation on education, which are implemented in public and private schools. It is constitutional that the Federal government may not control schools, set standards, or determine state budget, but national efforts are possible through promotion of educational goals and funding policies that encourage states to fulfill those goals. For example, at the 1989 Education Summit held at Charlottesville, Virginia, former president George H. W. Bush and 50 state governors agreed on education goals for the whole nation. For example, two of the goals were: "The high school graduation rate will increase to at least 90 percent", and "U.S. students will be first in the world in mathematics and science achievement," by the year 2000 [5]. Although these goals were not fulfilled, it can be seen as a starting point of the *Standards and Accountability Movement* in the 1990s. Though the United

[5] "The National Education Goals Report - 1993, Building a Nation of Learners", September 1993. Document retrieved from https://www2.ed.gov/pubs/goals/report/goalsrpt.txt

States still does not have a true "national curriculum", many states have adopted the Common Core State Standards since 2010.

Within our period of study (1991-2012), it is not likely to see the influence of the Common Core since the full implementation occurred after 2012. Indeed, according to Murphy and Adams (1998), the US education reform enters a "reformation era" in the 1990s, in which four "reform mechanisms" (p. 434) played significant role and focused on particular themes. These mechanisms helped in Chapter 5 as a context when discussing the US mathematics curricula.

- Government: The government promoted the development of national standards and accountability mechanisms. The efforts that can be studied include description of proposed acts, their results, related funding allotments, and new agencies.
- Profession: Organizations such as the National Board for Professional Teaching Standards devoted to strengthen the profession by maintaining a system of certification. Professional development programs also show what was important in educators' views.
- Citizen: Participation of parents in the education system, either in forming new (charter) schools or taking more decisive roles in existing schools, may reflect what citizens expect.
- Market: Privatization movements has become a strategy of stakeholders in order to implement desired ideologies in the education system.

Since the US curriculum was predominantly determined by state education departments, the "population" of documents can be defined as any published curricular materials (from 1991 to 2012) that are related to mathematics. A sample of documents are to be chosen from two states (FL and MA), since these are two of the three states that participated in PISA 2012 individually (but not 2003).

Germany

The year 1991 sees the reunified Germany. The constitutional nature for education is similar to the United States, since it is also the responsibility of the *Länder* (states) responsibility to operate their own Ministry of Education and hire teachers. The basic education system is rather complicated since there is no "mainstream" high schools. Most students first study in *Grundschule* (Primary School), but after so, they enroll into one of the following options: *Gymnasium, Realschule, Hau-*

Table 3.2: Sample of Curriculum Documents (1991-2012)

Germany
National: National educational standards (2003), National assessment (2006) Bavaria: Curriculum before 2003 (Grundschule 2000, 1981; Hauptschule 1997, Realschule 1994; Gymnasium 1991) and after 2003 (Grundschule 2014; Mittelschule 2004; Realschule, 2007; Gymnasium 2013, 2004), state-wide assessments (2003, 2012) North Rhine-Westphalia: Curriculum before 2003 (Gymnasium 1996, 1999) and after 2003 (Core Curriculum 2004; Grundschule 2008; Hauptschule 2011; Gesamtschule 2004; Realschule 2004; Gymnasium 2007)
Hong Kong
Mathematics Curriculum (1983, 1985, 1999), Hong Kong Attainment Tests (2003), Certificate of Education Examination (2003), Advanced Level Examination (2003), and Diploma of Secondary Education Examination (2012).
USA
Professional: The NCTM Standards (1989, 2000) Connecticut: Mathematics Curriculum Framework (2005), Connecticut Academic Performance Test items (2012) and technical reports (2003, 2012) Florida: Sunshine State Standards (1996), Next Generation Sunshine State Standards (2007), Florida Comprehensive Assessment Test sample items (2004, 2005, 2012) and specifications (2004, 2005, 2010) Massachusetts: The Education Reform Act (1993); Mathematics Curricula (1996, 2000, 2004, 2011) available at and Massachusetts Department of Elementary and Secondary Education; Massachusetts Comprehensive Assessment System (MCAS) mathematics tests items and technical reports (2003, 2012)

ptschule, or Gesamtschule (i.e. Grammar School, Intermediate School, Secondary general school, or Comprehensive School). However, they have developed a standard that was adapted by all states in 2004, quickly after the "shock" of PISA 2003 results, and also embraced national assessment in 2006. Germany is supposedly the most reactive country to the PISA shock, and thus an ideal case to study for the effectiveness of education policies. Indeed, the OECD official research acknowledged Germany's rapid improvements in the subsequent PISA tests to their reforms since 2004 (Organisation for Economic Co-operation and Development [OECD], 2011), and the national standards also emphasizes modeling as a central competence in mathematics. Therefore, it is natural to compare Germany's mathematics curriculum before and after 2004. Since the reunified Germany has 16 federal states, the two most populous ones (North Rhine-Westphalia and Bavaria) are chosen in order to see the differences of their mathematics curricula before a national standard was born.

Chapter 4

ANALYSIS OF CURRENT THEORIES

In this chapter, the actual steps and results of the study outlined in Chapter III are presented. The main purpose is to answer research question RQ1(a) (p. 6), namely, to collect and evaluate theories on the ability to apply mathematical knowledge. The methodology was discussed in Chapter II, the first Methodological Background section (p. 39). As mentioned in Chapter III, p. 65, articles from academic journals, treatises, and conference proceedings are included (See Table 3.1, p. 67). All the targeted journals and treatises are available as online database entries, some of which can be imported to EndNote X7, a reference management software (Reuters, 2013). As a result, there are 3,524 of them that can be imported as EndNote references. The articles that are only available as PDF pages are excluded; all of them are proceedings of conferences such as ICME and PME. The procedures described in p. 69 are treated as an outline of this chapter, including Classifying the Articles, Examining Theoretical Elements and Empirical Results, and Evaluating and Selecting the Elements.

Classifying the Articles

In order to classify all the articles into one of the four Types (see p. 66), a procedure was described in Chapter III, p. 69, guided by a flowchart (Figure 3.2, p. 70). In the actual process, codes were associated to each type of articles, indicating the reason why an article is classified in such type. For example, T1 has three codes, “T1 test ~”, “T1 study ~”, and “T1 describe ~”; each article is assigned with exactly one code. If an article is assigned with "T1 test ~", for example, it means that the article is classified as T1, because it proposes or studies a test on the ability to apply mathematical knowledge. A complete description of the codes is collected in Appendix A2 (p. 212). The list of codes was stablized (i.e. no new codes was created) after about 300 articles were classified.

Keyword searches were also employed to accelerate the classification process, and it worked as follows. The first round of keyword search used "model*", "appl*", and other related keywords to sort out articles that do not contain *any* word about modeling and/or education (See Appendix A1, p. 210). These articles are immediately classified as "T4". After that, a list of keywords related to T1 codes are created, such as "assess", "competen*", etc., and were searched subsequently. For each article, we check if any of these keywords is present in the Abstract or Title of the article, and if it bears the intended meaning. If so, the article can be immediately classified according to such meaning; otherwise, if all of those keywords do not bear the intended meaning, the article is still classified as "T4".

For example, if the word "model" in an Abstract of the article is not related to mathematical modeling at all, and the article does not include any other related keywords, then it is still classified as "T4". Another example: when we searched for "assess", we intended to find words that refer to an assessment of the ability to apply, or a test of students' competence of mathematical modeling. If an article's title or abstract include the keyword, and it has the intended meaning, then it is immediately assigned with "T1 test ~". Once an article is classified, it will not be changed. Therefore, when we searched for "competen*" later, some articles are already assigned with a code and they are not reviewed again. For articles that do not have a code yet, we check if the keyword has the intended meaning: a study of students' modeling competence ("T1 study ~"), or a description of such competence ("T1 describe ~"). If an article is classified as T1 in one code, then another code will not apply; therefore, the order of analysis matters, and the code list has to be prioritized: a code has a higher priority to be assigned than another if it is listed earlier. Both lists of keywords (Appendix A1) and the associated codes (Appendix A2) are presented in the order of decreasing priority. If an article could be classified as T1 for more than one reasons, hence more than one code could be assigned, then, it is more likely that "T1 test ~" is assigned than is "T1 study ~" or "T1 describe ~". It is assumed that, using the accelerated process, the articles will be classified in roughly the same "type" as is classified through the process defined by the flow chart. This is a point to be further investigated, but will not be addressed here, since the classification procedure is, after all, not based on a perfectly deterministic algorithm.

The results of the classification are as follows. Since only T1 and T2 articles will be useful, we did not keep track of all the T3 and T4 articles. Out of all 3,524 EndNote references to the targeted articles, 527 of them (15%) belongs to specialized publications, i.e. publications that serve exclusively for the topic of modeling and/or application. There are 47 and 197 of them that are classified as T1 and T2, respectively, that is, 46% of the specialized publications are useful for further analysis. On the contrary, out of the non-specialized publications, only 10 and 79 articles are classified as T1 and T2, respectively, making up about 3.0% of the non-specialized publications. It shows that the specialized publications are quite sufficient in covering our research topic, which is not a surprise. After the classification, we no longer need to distinguish whether the article comes from a specialized publication or not. Thus, from this point on, we are working on the 333 (=47+197+10+79) articles as a new subset of our data.

It is illustrative to show one article of each Type as examples. An article in T1 is *Assessment of Applied Mathematics and Modelling: Using a Laboratory-Like Environment* by P. Vos (2007), published in *Modelling and Applications in Mathematics Education* (ICMI Study 14). The main purpose of this article is to study the trend of Dutch secondary school students' ability to apply mathematics in practical situations, which fits well in T1 since it focuses on the ability to apply mathematical knowledge. In T1, there are three possible codes, including "test ~", "study ~", and "describe ~". Vos (2007) is coded as "test ~", for the study of trends relies heavily on an instrument which is able to measure students' ability. In our analysis, a total of 57 articles were classified as T1, 20 of which were coded as "test ~". 23 articles were coded as "study ~", meaning that the article's main purpose is to report or analyze empirical study of human subjects' possession of the ability to apply. The other 14 articles were coded as "describe ~", since they provide or discuss theoretical descriptions of individual's ability to apply. One may realize with common sense that the three codes should overlap, but "test ~" was prioritized over "study ~", which in turn is prioritized over "describe ~", being consistent to the current thesis.

A typical example of a T2 article is Van Hecke (2011), *Release the prisoners game*, published in the journal *Teaching Mathematics and its Applications*. This is coded as "example" since it has a quite straightforward presentation of a mathematical model (about the optimal strategy to win a game), and then a lesson plan about this model. T2 articles cannot be classified as any of the three codes of T1, but their main purpose is to study themes of teaching and learning that are also related to application. Among T2 articles, educational artifacts such as lesson plans, modeling tasks, teaching strategies, and so forth, still provide implications about the ability to apply. For example, a lesson might include a way to test whether students have obtained certain modeling skills after a period of learning. In Van Hecke (2011), the authors did not include a test in the lesson, but their lesson plan implied that statistical concepts and programming techniques are important for successful modeling. Also, the same task can be used in classes of various levels of statistical knowledge. Thus, this article provided some indirect information of modeling ability, and the openness of modeling tasks. There are 276 articles classified as T2, with 17 distinct codes. The three most popular codes are "example" (71), "modeling process" (39), and "develop ~" (36).

One T3 article is Niss (2015b), which deals with two kinds of purposes of mathematical modeling, and is coded as "T3: general". Although this article proposes an important idea of prescriptive modeling, it does not provide statements about the ability to apply, as T1 requires, nor does it deal with specific topics of teaching and learning applications, therefore it is too general to be included in our study.

The articles that are classified as T4 are either incomplete, such as an editorial note, or not in English, such as the German articles published in the *JMD*, or simply do not contain anything about modeling or application. Note that the Title or Abstract of all these articles is supposed to be hit by the keyword searches on the words about modeling or application, but they are the "false positives", i.e. those words in the Title or Abstract of the articles do not bear the intended meaning. For example, in a T4 article under the title "Does students confidence in their ability in mathematics matter?", the Abstract states, "A regression model was produced which predicted a 12% increase in mathematics marks per increase in GCSE mathematics grade…" Evidently, the word "model" here does not make the article relevant to the education of modeling and/or education.

At this stage, there are 54 T1 articles and 276 T2 articles. They are not treated in the same manner, since the T1 articles are more relevant to our research question RQ1(a), as they are directly related to the ability to apply. T2 articles are only potentially helpful, since they contain elements that *may* concern the ability to apply. All T1 articles are selected, whereas the selection of T2 articles are subject to an additional screening process: given a T2 article, its abstract is to be read again, with its T2 code as a clue of whether this article would imply anything worthy of further investigation. For example, if an article is coded as "T2: assessment analysis", it means that the article analyzes an existing assessment of mathematics that may contain modeling and/or application, or offers some principles of such analysis. We can check if the assessment has any part that tests the ability to apply; if the assessment contains modeling and/or application but the intent is not to test the ability – for example, only to show a variety of problems, or only to lead students to appreciate application more – then the article is not included. Another example is as follows. If an article is coded as “T2: course development”, it describes the development of a modeling course, or the principles related to such development. We can check if assessment of the ability to apply is an integral part of the course (or not). Examples are shown as follows for a T2 article that is selected after the screening and another that is not.

A T2 article that is not selected: Kaiser & Schwarz (2010)

```
T2 code: assessment analysis
Reason of non-selection: Its main purpose is to present good
examples of authentic modeling problems, but these problems are
not for testing students' modeling ability, but for motivating their
learning.
Title: Authentic Modelling Problems in Mathematics Education --
Examples and Experiences
Abstract: In this paper, we describe mathematical modelling
activities, which deal with authentic problems. These kinds
of problems have been tackled in various modelling activities,
amongst others in a modelling week. After a description of the
theoretical approach used, one of these authentic modelling problems
is described in detail showing students' solutions. Based on the
evaluation of a modelling week with several hundred students,
it is argued that these kinds of authentic problems are feasible
with students from upper secondary level. Furthermore, it became
apparent that most students would appreciate these kinds of examples
included in school mathematics in order to promote their skills to
use mathematics in their real life.
```

A T2 article that is selected: Schukajlow, Krug, & Rakoczy (2015)

```
T2 code: pedagogy
Reason of selection: The abstract mentions tests of solving modeling
problems and there is a measurement of performance.
Title: Effects of prompting multiple solutions for modelling
problems on students’ performance
Abstract: [...] we conducted an experimental study with 144 ninth
graders from six German classes from middle track schools. We had
two experimental groups: In one experimental group, students were
required to provide two solutions for modelling problems related
to the topic of Pythagoras’ theorem; in the other group, they were
asked to find one solution for each problem. Students’ performance
in solving tasks with and without a connection to the real world was
assessed before and after a five-lesson teaching unit. [...] .
```

In total, 21 out of 276 T2 references were selected. Along with the 54 T1 articles, there are 75 articles to be analyzed.

Examining Theoretical Elements and Empirical Results

Given each of the 75 articles, the whole article is skimmed through, and some parts of it is quoted or paraphrased so that relevant theoretical elements and empirical studies are "extracted". The theoretical and empirical contents, if both exists, will also be analyzed for their connection to each other.

Note that the theoretical elements in an article are not necessarily stated as a *theory*, but a self-contained group of theoretical elements has to include theoretical constructs, assumptions, and hypotheses or inferential statements. In the descriptions below, such a group is simply referred to as a "theory", which is known as "explicit" if the author described it as a theory or a theoretical framework, and "implicit" otherwise. Each article may contain more than one theories, so they are numbered as Th 1, Th 2, etc. For example, in Vos (2007), two implicit theories are identified and recorded as follows:

Th 1: Validity and Reliability for modeling assessments

```
1.1 Constructs: validity, reliability
1.2 Assumption: validity can be tested by (1) expert appraisal with
respect to an intended curriculum, (2) assessment grid
1.3 Inference: modeling assessment can be tested by expert appraisal
with respect to the Dutch RME-based intended mathematics curriculum
1.4 Inference: modeling assessment can be tested by the grid
designed for assessment of modelling and applying mathematics
1.5 Constructs: mathematizing, rewriting (generalizing and
simplifying), interpreting, and reflecting.
1.6 Assumption: each item could be fitted into one category, or two
categories with weights
1.7 Assumption: Cronbach's alpha > 0.5 is acceptable for
reliability; p value of chi square < 0.05 considered as indicator of
unequal testing circumstances
```

Th 2: Alternative assessment

```
2.1 Assumption: paper-and-pencil tests cannot evaluate all practical
competencies from an intended mathematics curriculum
2.2 Construct: Niss' criteria, (a) testing through open questions
and for higher order skills, (b) being open to a range of methods or
approaches, (c) making students disclose their own understanding,
(d) allowing students to undertake practical work, (e) asking for
performances and products, (f) being as an activity worthwhile for
students' learning, and (g) integrating real-life situations and
several subjects (Niss, 1993).
2.3 Inference: As observation, interviews, portfolio are expensive
and may not be reliable, TIMSS Performance Assessment provides a
good example of alternative assessment using hands-on tasks.
```

The author cited and used the two theories, but did not construct new elements in them, therefore they are also recorded as "Use" instead of "Build". The numbering of the constructs, assumptions, and inferences is necessary for later reference. Among the 20 sampled articles, 5 of them do not present a theory, and the rest present one ore more theories, totaling 38 of them. Only two articles have explicit theory (Niss, 2013; Wake & Hardy, 2007), while seven articles have attempts to build theories (Roorda, Vos, & Goedhart, 2007; Niss, 2013; Maaß, 2007; C. Haines & Crouch, 2007; Ferri, 2011; English, 2010; Antonius, 2007).

Empirical studies, if exist in an article, are extracted with their sample descriptions, instruments used, methods of analysis, and findings. These elements are not required to be all available at the same time. The studies are numbered as "Emp i, Emp j", etc., where i and j are numbers distinct from the theory indices. From Vos (2007), there is only one study extracted:

Emp 3: Trend study of Dutch students' ability to apply

```
3.1 Sample: 1995, n=437 grade 8 students from 49 secondary schools;
2000, n=234 from 27 schools
3.2 Instrument: TIMSS Performance study items, with trend
calculation
3.3 Finding: the reliability coefficients are satisfactory, but the
trend comparability depends on circumstances, and varied much for
different items.
3.4 Finding: students' performance has no significant change from
1995 to 2000.
```

After the theoretical elements and empirical studies are recorded, we check the connection between them, including the tendency (towards empirical or theoretical), consistencies, and contraditions. In Vos (2007), the tendency of the author is to report the usage of TIMSS Performance Assessment and how it was applied along with the Realistic Mathematics Education programs in the Netherlands. Therefore this paper is "towards empirical". There are no contradictions recorded, but one consistency is that the inter-rater reliability score was high, showing that the assessment had good validity.

After these steps, sufficient information is obtained through the reading an articles, and recorded as Dataset B, so that the original article is not read again. Such information include theoretical elements and empirical studies that are related to application of mathematical knowledge, but only within individual articles.

Evaluating and Selecting the Elements

Finally, we analyze possible "networking" (cf. p. 44) among different articles. When a collection of articles and the extractions of theoretical elements and empirical studies are available (i.e. Dataset A and B are ready), we proceed to critique the theories systematically, using Schoenfeld's eight criteria. For practical convenience, the criteria are considered in the following order: Scope, Predictive power, Explanatory power, Descriptive power, Falsifiability, Replicability, Multiple sources of evidence.

For each article, read the list of the theories that it presented. Each theory must address at least one topic or phenomenon related to application, and this information is recorded under the Scope column. For example, Th 1 in Vos (2007) addresses one topic: (1) psychometric properties of modeling assessment, and Th 2 also addresses one topic: (1) non-traditional forms of assessment. These are treated as the scope of the theories. Then, form predictions from the assumptions, hypotheses, or inference statements that belong to the theories. Note that the predictions depend

Table 4.1: Vos (2007) with Schoenfeld's criteria

Scope:	Th 1: Validity and Reliability for modeling assessments (1) psychometric properties of modeling assessment Th 2: Alternative assessment (1) non-traditional forms of assessment
Predictive Power:	Th 1 (none) Th 2 (2.3) If other hands-on assessment items similar to the TIMSS Performance Assessment are available, they are likely to be more valid psychometrically, compared to traditional written tests, and more reliable, compared to portfolio or project assessments.
Explanatory Power:	(no explanations available)
Descriptive Power:	(1) (3.3) The item characteristics and assessment circumstances that might affect the reliability and comparability were not described (2) (3.4) The characteristics of the 1995 and 2000 samples were not described
Rigor and Specificity:	NA
Falsifiability:	Th 1: 0 Th 2: 1
Replicability:	(empty)
Multiple sources of evidence:	Only TIMSS hands-on items are available.

much on the researcher who forms them, due to the possible subjectivity when those assumptions, hypotheses, and inferences are interpreted.

The list of empirical studies are then read in order to check if all the findings are explained by the theories. The explained findings are listed under the "explanatory power" along with the theory, and the unexplained ones are listed under the "descriptive power", mentioning the reasons why they are not reached by the theories.

Thus far, four columns are done with an article, and the ones associated with Vos (2007) are shown in Table 4.1.

The other four columns are treated as follows: *Rigor and Specificity* is always "NA" because it is out of the scope of this research (see p. 65). The *Falsifiability* column simply records the number of predictions made for each theory. Although falsifiability is not the same as predictive power, we treat available predictions as possible way to falsify a theory. *Replicability* is recorded only when a replication study is done. Finally, under the *Multiple sources of evidence* column, a

Table 4.2: Theory: Modeling Competence

Number of papers on this theory: 3
Prediction(s): (1) If one of the three dimensions is not well established (being able to activate the modeling steps, being able to understand the context, being able to proceed the mathematical techniques), then student is not able to carry out the modeling process well. (2) If other hands-on assessment items similar to the TIMSS Performance Assessment are available, they are likely to be more valid psychometrically, compared to traditional written tests, and more reliable, compared to portfolio or project assessments.

comment is written about the available sources. In the first analysis that was done, there was not any replication study, and all empirical studies do not provide multiple sources of evidence.

If the same theory is used in two different articles, then they are merged as one theory, meaning that the set of explanations, predictions, etc., of the theory are combined. Also, if one theory extends another, as is often the case, they are merged too. After the merging, the most popular theory and the most falsifiable theory are identified. In the first analysis, the most popular theory is *Modeling Competence* (Højgaard Jensen, 2007; Kaiser, 2007; Maaß, 2007); note that all of these papers were presented on the same conference (ICTMA 12). The theories that have the greatest number of falsifiability are *Project Examinations* (Antonius, 2007), *Early experiences of data modeling* (Lehrer & Lesh, 2003, cited in English, 2010), *Two reflections on meta-cognition* (Ferri, 2011), and *Cultural historical activity theory* (Wake & Hardy, 2007). We treat *Modeling Competence* and *Project Examination* as "candidate theories" and put them into tests, namely, to try to line them up with empirical cases.

The list of empirical cases is simply all the empirical cases extracted from the articles, and they were classified according to their scope also. Then, a search into the empirical case was done to see if the predictions of the theories would be supported or not. The two "candidate theories", *Modeling Competence* and *Project Examination*, are presented in Tables 4.2 and 4.3.

The first prediction of *Theory: Modeling Competence*, if not to be considered a tautology, must

Table 4.3: Theory: Project Examination, Antonius (2007)

Number of papers on this theory: 3
In this paper, Antonius discussed project examination as a tool to assess modeling competence, as opposed to traditional written tests. The theoretical elements were implicit, but the article generally assumes that project examination can be an adequate assessment tool for the purpose.
Prediction(s): (1) Effort explains much variations of project performance (2) A 30-min presentation is enough to determine whether a student cheated; a 10-min presentation is not sufficient (3) A students' modeling ability is well measured by project examination

be interpreted in the sense that two kinds of measurement of modeling are available: the holistic and the atomistic. The holistic assessment treat modeling and application as a whole task, to check if a student is able to perform the overall task and answer related questions well. In an atomistic assessment each dimension is tested separately. Thus, the statement (1) predicts that if one does not do well on one dimension of the atomistic assessment, he must not be able to do well in a holistic task. Having this in mind, we found that there was not any empirical case that measures both holistically and atomistically, and therefore the predictions are neither supported nor refuted so far. It is worth further research on students who do both holistic and atomistic assessments.

The predictions of the *Theory: Project Examination* are very specific, but since no other articles studied project examination, the three predictions are neither supported nor refuted. Further research is again needed.

Results: a Collection of Theoretical Elements

After the analyzing the articles, the predictions extracted from the articles are not supported or refuted by other articles. So far, it is possible to adjoin all the theoretical elements together to build a larger framework without obvious contradictions. This is done by gathering all the entries of the Scope column, so that a few categories emerge out of the researcher's prior knowledge:

Framework resulting from the analysis

A. Nature of Modeling and Application (4 theories)

1. What activities and tasks are considered modeling and/or application? Niss (2013).
2. Role of modeling in mathematics education: Ludwig & Xu (2010).
3. Role of modeling in another or a more general perspective: Wake & Hardy (2007).
4. Existing Framework about teaching and learning of modeling: Gallardo (2009).

B. Modeling Process (5 theories)

1. The process as a whole: Ekol (2011); Lesh & Yoon, (2007); Mischo & Maass (2012); Haines & Crouch (2013); Galbraith & Stillman (2006); Haines & Crouch (2007).
2. The process as many sub-processes: Caron & Bélair (2007)
3. On the sub-processes about the *real-world situation*: Leiss, Schukajlow, Blum, Messner, and Pekrun (2010)
4. On the sub-processes about the mathematical model: Roorda, Vos, & Goedhart (2007)
5. How modeling is actually done (usually by students): Niss (2013); Haines & Crouch (2007); Haines & Crouch (2013).

C. Nature of Modeling competence (7 theories)

1. Basic principles about competence descriptions: Coben & Weeks (2014).
2. Definition or description of modeling competence: Blomhøj & Jensen (2007); Henning & Keune (2007); Jensen (2007); Kaiser (2007); Kaiser & Brand (2015); Kaiser & Maaß (2007); Maaß (2006); Maaß (2007).
3. "Levels" model: Blomhøj & Jensen (2007); Ludwig & Xu (2010).
4. Sub-competencies model: Caron & Bélair (2007); Niss (2013).
5. Professional demand model: Coben& Weeks (2014).
6. Factors or aspects that affect modeling performance: Maaß (2007); Mischo & Maaß (2012); Palm & Nyström (2009).
7. Observations or descriptions of modeling difficulty: Bock, Dooren & Janssens (2007); Crouch & Haines (2007); Galbraith & Stillman (2006); Galbraith, Stillman, Brown and Edwards (2007); Klymchuk, Zverkova, Gruenwald and Sauerbier (2010); Wake & Hardy (2007); Palm & Nyström (2009).

D. Measurement of Modeling competence (5 theories)

1. Basic principles about measurement: Antonius (2007); Izard (2007); Jensen (2007); Mischo & Maaß (2012); Palm (2007); Reit & Ludwig (2015); Vos (2007); Coben & Weeks (2014).
2. Assessment Frameworks: Greer & Verschaffel (2007); Haines & Crouch (2013); Houston (2007); Ludwig & Xu (2010); Mischo & Maaß (2012).
3. Assessment Tools (in general): Haines & Crouch (2007); Kaiser (2007); Turner (2007); Alpers (2013).
4. Written assessment methods: Crouch & Haines (2007).
5. Non-written assessment methods: Antonius (2007); Vos (2007).

E. Development of Modeling competence (5 theories)

1. Descriptions of developmental stages: Greer & Verschaffel (2007); Henning & Keune (2007); Makar & Confrey (2007).
2. The development process of modeling competence: Crouch & Haines (2007); Wake & Hardy (2007).
3. Pedagogical tools for the development of modeling competence: Bock, Dooren and Janssens (2007); Ekol (2011); Iversen & Larson (2006); Lesh & Yoon (2007).
4. Beginner experience and development in modeling: English (2010).
5. Workplace experience and development in modeling: Wake & Hardy (2007); Coben & Weeks (2014).

F. More general attributes and their development (3 theories)

1. Mathematics Competence that includes modeling competence: Antonius (2007).

2. General abilities that are related to modeling competence: Bock et al. (2007); Ferri (2011); Leiss et al. (2010); Maaß (2007); Stillman (2011).
3. Beliefs: Kaiser & Maaß (2007).

This is an open framework, since if another article provides a new theoretical element, it can "challenge" some of the theoretical elements within the corresponding category. The theories within this framework are also to be re-organized in the next chapter to fit the particular usage: the framework will be organized into a form that judges assessment items, to see how much they are capable of assessing modeling abilities, and then the same framework will be used to analyze the role of modeling in different curricula.

During the process, seven theories are selected, for they either have more testable claims, or they are proposed by more scholars (see Appendix A3, p. 215). It seems that the predictions of these theories can hardly be verified or disproved using existing research. This shows that although there are many articles that discussed the ability to apply mathematical knowledge, not many of them attempted to make theoretical, falsifiable claims. This observation is consistent to that of Kilpatrick (2010), Niss (2007), and Frejd (2013). Furthermore, some testable claims do exist, but they opened up more questions rather than participated in debates on specific details. Ideally, if there are more articles that deal with an identical or similar set of research questions, then a meta-analysis will be possible so as to contribute more significant results to the field. Note that this process is ongoing, and the dataset can be updated as long as more research related to the topic is published.

In this thesis, the best that can be done is to consolidate all the theoretical elements, namely, to make a larger framework out of them, treating it as if it is one theory, though the number of assumptions will be too many to fit the parsimony criteria, suggested by the classical *Occam's razor*. Therefore, the resulting framework should be viewed *dynamically*, so that it is subject to future needs of data analysis. The validness criteria of a dynamic theory is, therefore, how the statements of the theory can provide insights and self-consistent understanding of an issue.

Chapter 5

ANALYSIS OF CURRICULUM AND ASSESSMENTS

The theories synthesized from Chapter IV are sufficient to create a framework to analyze the modeling aspects of assessment items. In this Chapter, the main work is to build the analytical framework, and to apply it to curriculum materials.

Principle to build such framework was discussed in Chapter II, based on the *assessment triad* of *construct*, *observation*, and *interpretation*, originally described in Shavelson (2010) as the *assessment triangle*. Clearly, the construct to be measured is modeling competence, which is mainly seen as the ability to work in the modeling cycles. The means of observation is the modeling task, which can be any kind of assessment items ranging from answering a multiple-choice question, doing a modeling project, or solving a real-world problem. Interpretation of the observed results can also vary, due to the complexity of modeling competence.

In the narrative literature review of Chapter II, we already identified some of the thoughts on each of these three elements, but we also found some inconsistencies. In Chapter IV, we attempted to identify a bigger picture of the literature about modeling competence assessment, and synthesized the theoretical elements to create a larger theoretical framework. This framework can now serve as a basis to construct assessment frameworks, as shown in Tables 5.1, 5.2, and 5.3 (pp. 96, 97, and 98). There are multiple assessment frameworks because for each element of the assessment triad, there exist multiple views or descriptions.

As a parallel to the building of assessment frameworks, the same thing can be said about the

Table 5.1: Assessment Framework 1: Atomistic description using Modeling Cycle

Assessment Framework 1: Atomistic description using Modeling Cycle (based on Theories 2, 3, and 5)
Construct being assessed: Modeling Competence - First, a modeling process can be described ideally as a cycle of steps, which was depicted in Figure 2.1. A person is working on a real-world problem can invoke this cycle, starting with identifying key variables and relationships, making appropriate assumptions, and setting reasonable goals, then coming up with a mathematical model using mathematical objects. This is called *mathematization*. - After that, the modeler has to provide a mathematical solution, which is to be translated back to the *real world* (*de-mathematization*). A successful modeler not only has to work on (a) each modeling steps, but also (b) transition from one step to another. He also need other general competencies such as that of (c) reading, (d) using tools, (e) argumentation, and (f) working as a team. Finally, he will need to have (g) meta-knowledge about the modeling process, (h) general knowledge about the context the problem lies in, and (i) correct beliefs and (j) positive attitudes toward using mathematics to solve real-world problems. - Modeling Competence thus consists of all of the above elements (a) - (j), or more, if identified by any further research.
Modeling Tasks: - Ideally, a test of modeling contains at least one item which evokes each of the sub-competency separately, and there may be problems that require more than one sub-competency, which are supposed to be harder.
Inference of students' modeling competences: - Quantitative models such as Item Response Theory are most likely used, in order to analyze the multiple choice items, or open constructed-response items coded by two or more raters.

building of curriculum frameworks, that the theoretical framework resulted from Chapter IV (p. 91) can be applied to curriculum analysis. The current work applies the "Foci, Perspectives, and Lenses" approach, which is based on the works of Goodlad & Richter (1966), Adamson & Morris (2014), and Marsh (2009) (Chapter II, p. 56). It is possible to create three curriculum frameworks in the same way that the three assessment frameworks were created. The *focus* of a curriculum has to include mathematical modeling here is again modeling competence, which is the same as the *construct* in the assessment frameworks. The full range of perspectives of curriculum, including the intended curriculum, the implemented curriculum, the attained curriculum, the hidden curriculum, and the null curriculum, is not possible to be covered; only the intended, the attained, and the null curriculum are studied through the analysis of policy and curriculum documents, some teaching materials and change in PISA performance. The null curriculum is especially of interest

Table 5.2: Assessment Framework 2: “Holistic” descriptions based on a modeling cycle

Assessment Framework 2: “Holistic” descriptions based on a modeling cycle (based on Theories 1 and 7)
Construct being assessed: Modeling Competence - The modeling cycle is also defined similarly as in the “atomistic” framework, but the focus is on the modeling behavior that is desired. Thus, modeling competence is simply defined as whatever a successful modeler is able to do while a novice is not.
Modeling Tasks - Again, the focus is not to examine individual modeling steps, but to create a series of modeling problems from the easiest to the hardest; even the easiest kind of modeling problem should contain a whole modeling cycle, though it may not even be noticed by the problem solver.
Inference of students’ modeling competences - A modeling “level” can be defined relatively, according to what expert modelers can do. One set of descriptors is as follows. (a) Ideal behavior: consciously going through each step of the modeling cycle, revise as necessary. The ideal modeler reports not only the results but also the significant ideas throughout the whole thinking process. (b) An adequate Modeler takes into account the relationship between the real world and the mathematical world, and knows how to translate back and forth between the two worlds. (c) An Inadequate modeler thinks about the relationship between the real world and the mathematical world but is lack of knowledge to have the whole modeling process done. (d) An unable modeler does not take into account the relationship, but keeps thinking from only one side of the modeling problem, either in mathematical terms or realistic terms, but not both. - There are other similar ways to define these descriptors. Moreover, one can also define modeling levels simply by the extent a student is able to work through the modeling cycle. A modeler of the highest level is able to go through the whole process and also revise when necessary; a lower level modeler is either one who can just complete one cycle but not good at reporting and revising, or one who is not able to finish the whole cycle. The higher is a modeler’s level, the further he can proceed within a cycle or beyond. - Finally, the level description is always dependent on the modeler’s education background and the complexity of the problem. A one-dimensional continuum is assumed on which a modeling problem is located according to its complexity level, and a modeler’s typical level can also be located, according the average performance over students with equivalent education background. Both quantitative and qualitative methods can be used in determining ones’ modeling level.

Table 5.3: Assessment Framework 3: "Models and Modeling Perspective" approach

Assessment Framework 3: "Models and Modeling Perspective" approach (based on Theory 4)
Construct being assessed: Modeling Competence - In this approach, modeling is not pursued as an extra thing to be learned, but rather, students learn mathematics through activities , known as "Modeling Eliciting Activities" (MEA), that are essentially mathematical modeling. Modeling Competence is then equivocal to students' capacity and willingness to work on these MEAs, which are supposed to be integrated in students' daily mathematical learning experience.
Modeling Tasks: MEA - Principles in designing MEAs are: "(a) Students must be engaged in problem solving activities in which they clearly recognize the need to revise or refine their current ways of thinking about the situation, (b) Students must be challenged to express their current understandings in forms that they themselves can test and revise multiple times, (c) The conceptual tools that students develop should be expected to be shareable (with others) and re-useable (beyond the immediate situation) beyond the specific situations in which they were developed." R. Lesh and Yoon, 2007 - Standardized test items do not likely fit all these principles, but the more an item follows these principle the better.
Inference: - Ways to infer students' modeling competence in this approach are not explicitly described. One may use methods that are dependent on the specific form of a task.

because it means what parts or aspects of modeling a curriculum neglects or underemphasizes. Finally, the lenses (or ideologies) of the curricula can be discussed through the curricula's view of modeling.

Six ideologies were mentioned in Chapter II (p. 59), which are academic rationalism, social and economic efficiency, social reconstructionism, orthodoxy, progressivism, and cognitive pluralism. They are also summarized in Table 5.4 (p. 100) for convenience. It can be argued that these ideologies correspond to different ways of looking at mathematical modeling and its role in education. To an academic rationalist, mathematical modeling is a subject to be pursued. The best resource of knowledge about mathematical modeling is from the applied mathematicians, both in the academia or the industry. These applied mathematicians provide the foundation of the discipline of mathematical modeling, so that the education system can adopt it, either as a part of mathematics, or a separated subject. To a person who cares more about social and economic efficiency, mathematical modeling provide ways to incorporate an abundant collection of concepts and techniques from mathematics in order to solve realistic problems, and eventually contribute to social and economic growth. A social reconstructionist would use mathematical modeling as an educating tool, for example, to show how one could break the "walls" between different subjects, as well as break the walls among the (bad) rigid structure of the society. Orthodoxy means that the educators who write the curriculum intend to promote a certain political or religious view by and through the introduction of modeling, such as employing politically relevant situations for the modeling tasks. Similar arguments are available for progressivism and cognitive pluralism. The ideology of the curriculum materials can be recorded as shown in a later section, Table 5.5 (p. 102). Curriculum excerpts from different materials will also be presented in each education system.

In the rest of this chapter, a selection of PISA items is to be analyzed, and parts of the USA, German, and Hong Kong's curriculum will also be compared under these frameworks.

Analysis of curriculum materials

Using Table 5.4 as a basic tool, one may analyze the modeling focus of a curriculum from any given perspective, and for any *education system* of interest. An education system is one that is self-contained, with a stable population of teachers and students, and maintains its own curriculum materials and assessment tools. It is not to say that such a system is not influenced by agents outside of the system, but rather the system has its own decision makers who are able and responsible to define how the system should work. Under this definition, an education system can be a private school, a school district, a city, a country, or even a group of countries that share an educational vision. For example, in our analysis of PISA materials, all countries that participated the PISA program is supposed to share the same vision of the PISA committee, and therefore can

Table 5.4: A tool of analysis: Assessment and Curriculum Frameworks

	Modeling Competence
CF1/2	Is modeling one of the goals or important part in the curriculum?
CF1/2	Is modeling introduced as a modeling cycle?
AF1/2	Is modeling cycle mentioned in the rationale of the assessments?
AF1,CF1	Are the individual modeling steps mentioned?
AF1,CF1	What sub-competencies of modeling are mentioned?
AF2,CF2	What behaviors are expected of competent modelers?
AF2,CF2	What behaviors are expected of less competent modelers?
CF3	Does the curriculum attempt to shape teaching and learning of mathematics through modeling?
	Modeling Tasks
AF1	What sub-competencies of modeling are assessed in the item?
AF1	How difficult is the item with respect to the assessed competencies?
AF2	Does the item involve the whole process of modeling?
AF2	If the whole modeling process is required, how is it compared to the expected behavior of competent modelers or less competent modelers?
AF3	Does the task follow the three MEA principles? (1. Students should clearly recognize the need to revise or refine their current way of thinking about the situation; 2. Students are challenged to express their own understanding so that they can test or revise it; 3. In the task, students develop some conceptual tools, which are shareable and re-usable.)
CF1	Are curriculum materials being selected so that each sub-competency is addressed?
CF2	Is the whole modeling process taught in each grade level?
CF1/2	Is modeling introduced through an atomistic or a holistic approach?
CF3	Is the mathematics curriculum organized according to the MEA principles?
	Inference Methods
AF1/2	How are students' response to the items coded?
AF1/2	What quantitative analysis methods are used to report the performance?
AF1	How are different sub-competencies of modeling being analyzed?
AF2	How are different "levels" of modelers being identified?
AF2	What are some qualitative methods used?
	Ideologies (reflected through the role of modeling)
CF1/2	(*Academic rationalism*) mathematical modeling is a subject established by applied mathematicians that is to be pursued in mathematics education.
CF	(*Social and economic efficiency*) Learning to solve realistic problems in the society using mathematics is the purpose of mathematical modeling.
CF	(*Social reconstructionist*) Mathematical modeling can be treated as an educating tool to promote social reformation.
CF	(*Orthodoxy*) To employ modeling tasks that are relevant to political or religious views, so that a student is influenced toward such view through mathematics learning.
CF3	(*Progressivism*) Using modeling to create a holistic environment in which students are provided with opportunities for enhancing their personal and intellectual development
CF3	(*Cognitive pluralism*) Modeling provides room for diverse inputs and outcomes and can be adopted by students of various levels.
	Note: Three of the Ideology categories were not assigned to anyone of the three curriculum frameworks. Nonetheless, there exists other theories that correspond to some of these categories; see Discussion section.

be treated as one system (*PISA participants*). For USA, we will study the curriculum of two states — Florida and Massachusetts; for Germany, only one federal state Bavaria but two school types — *Hauptschule* and *Gymnasium*; for Hong Kong, the whole city is treated as one system.

In the following analysis, the intended curriculum and the assessed curriculum are studied. The first step is therefore to understand the materials to be analyzed, including the curriculum specifications and some past public examination papers. More current documents are often available at the websites of the governments' ministry or department of education. Older documents, especially those before 2000, may be obtained on the Internet also, but it depends on the availability of electronic archives.

Given a collection of available documents, Table 5.4 is applied to analyze intended curriculum frameworks and some assessment tasks that are related to them. We create a spreadsheet data file in order to store the results of the analysis. Each data entry (called a *fragment*) is a piece of chosen material which can be a sentence, a paragraph, a figure, or a combination of them. Each data entry is also indexed by a unique pair (Material Code, Fragment number). For example, since FL96SSS is the material code for the Florida's Sunchine State Standards 1996, (FL96SSS, 4) means the 4th chosen piece of material of this curriculum document.

For each fragment, the original wording and the page number is also recorded; for German documents, a translation is also kept in the spreadsheet. Furthermore, other data columns for analysis are described in Table 5.5. In the following, we will illustrate how to work with these data columns for analysis, along with some examples.

For each education system, we first try to figure out its curriculum ideology(-ies), which is not explicitly stated but has to be inferred from, for example, statements about the purpose of the curriculum, descriptions of the role of mathematics, and how the contents of mathematics are presented. An example (in the gray box) can illustrate this point:

Table 5.5: Data Columns of a Fragment of curriculum material

Material code (for a list of material codes and number of fragments, check Appendix.)
Fragment Number
Page Number
Material Content
Ideologies — one or more of the six ideologies that can be inferred from the fragment, including *Academic rationalism*, *Social and economic efficiency*, *Social reconstructionist*, *Orthodoxy*, *Progressivism*, and *Cognitive pluralism*.
Modeling Competence — specified requirement or expectation of application or modeling ability
Modeling Tasks — For a fragment that is considered an application task of school mathematics, specify the kind of real-world situation and the mathematics that the task is related to. If necessary, also include the kind of inference that would be done to a possible response of the task.
Modeling Sub-competencies: Does the task involve one of the following sub-competencies? (This part is further explained in Table 5.6.)

Example: HK07KLA (Key Learning Area, Hong Kong, 2007).

Fragment 1 & 3 (page 1):

```
Mathematics is a powerful means in a technology-oriented and
information-rich society to help students acquire the ability
to communicate, explore, conjecture, reason logically and solve
problems using a variety of methods.
Many of the developments, plans and decisions made in modern society
rely, to some extent, on the use of measures, structures, patterns,
shapes and the analysis of quantitative information. Therefore,
mathematical experiences acquired at the senior secondary level
enable students to become mathematically literate citizens who are
more able to cope with the demands of the workplace.
```

These statements reflects a view of *Social and Economic Efficiency*, since mathematics is seen as an important tool for citizens in the modern society. It is also an important subject to learn in order to get ready for the mathematical demand of the workplaces.

Fragment 2 (page 1):

```
Mathematics provides a means to acquire, organise and apply
information, and plays an important role in communicating ideas
through pictorial, graphical, symbolic, descriptive and analytical
representations. Hence, mathematics at the senior secondary level
helps to lay a strong foundation for students’ lifelong learning,
and provides a platform for the acquisition of new knowledge in this
rapidly changing world.
```

The focus this fragment is more about how a person may develop in various thinking skills when studying mathematics. *Progressivism* fits the most of this view, for the role of mathematics learning is seen as an opportunity for personal and intellectual development.

Secondly, the role of modeling and/or application is to be described. Some documents treat modeling and application as a key part of the curriculum, which is not the case in other ones, but all documents have mentioned application of mathematics. Therefore, we will compare the importance of application relative to each curriculum document.

Thirdly, the most important part in this analysis is to study the test items or tasks that were used in each school system in a specific period of time; it is assumed that the chosen tests are related to the curriculum framework that was implemented in the same period. Of course, not all tasks are supposed to be related to modeling and application, and not all word problems properly test the application ability. To analyze how a mathematical task is related to modeling and application, we will take the *atomistic* point of view of modeling competence (cf. Chapter II, p. 19), that is, each task is considered to be related to some (or none) of the modeling steps. In addition, we will identify the context that the task lies in, and the mathematical knowledge that is used. For example, the following material involves considerations of the real model and the mathematical model:

Example: KMK12Abi, Fragments 6 & 7 (Pages 41, 42). Translation is available in the next gray box.

Aufgabe

```
Auf einer bestimmten Strecke verwendet eine Fluggesellschaft Flugzeuge mit jeweils 100
Plätzen, die vor Flugantritt gebucht und bezahlt werden.

Vereinfachend soll davon ausgegangen werden, dass

- jeder Platz einzeln gebucht wird,
- auf dieser Strecke stets alle Plätze gebucht werden,
- ein Platz, dessen Buchung storniert wird, nicht erneut gebucht werden kann.
```

Im Mittel werden 10% der gebuchten Plätze kurzfristig storniert (d. h. von den Leuten, die gebucht haben, doch wieder abgesagt).

Von einer Person, die tatsächlich fliegt, nimmt die Fluggesellschaft 200€ ein, bei einer Stornierung wegen teilweiser Erstattung nur 100€.

a) Nennen Sie eine Annahme, sodass die mögliche Anzahl der Stornierungen bei einem Flug als binomialverteilt modelliert werden kann. Beschreiben Sie eine reale Situation, in der diese Annahme **nicht** zutrifft.

Im Folgenden wird nun angenommen, dass die mögliche Anzahl der Stornierungen bei einem Flug tatsächlich binomialverteilt ist.
b) Gegeben sind die folgenden drei Ereignisse:

E1: „Bei einem Flug werden genau 90 Plätze tatsächlich genutzt. "
E2: „Bei einem Flug werden höchstens 90 Plätze tatsächlich genutzt. "
E3: „Bei einem Flug werden mindestens 90 Plätze tatsächlich genutzt. "

Stellen Sie jeweils einen Term auf, mit welchem die Wahrscheinlichkeit des entsprechenden Ereignisses berechnet werden kann.

c) Berechnen Sie, welche Einnahmen die Fluggesellschaft auf lange Sicht im Mittel pro Flug erwarten kann.

Um die Flugzeuge besser auszulasten, bietet die Fluggesellschaft für jeden Flug 108 statt 100 Plätze an, also 8 mehr als verfügbar. Stets werden alle angebotenen Plätze gebucht. Auch wenn im Mittel 10% aller Buchungen storniert werden, geht die Fluggesellschaft damit das Risiko ein, dass eine Person, die einen Platz gebucht hat, ihren Flug nicht antreten kann, da der gebuchte Platz nicht zur Verfügung steht.

d) Berechnen Sie die zu erwartenden Einnahmen der Fluggesellschaft pro Flug im Mittel unter Berücksichtigung der Stornierungen aus dem Ticketverkauf.

e) Berechnen Sie die Wahrscheinlichkeit, dass mindestens ein Fluggast, der nicht storniert hat, seinen gebuchten Flug wegen Überbuchung nicht antreten kann.

Ob sich die Geschäftspraxis, Überbuchungen zuzulassen, für die Fluggesellschaft rein ökonomisch lohnt, soll im Weiteren untersucht werden.

Bei jedem Fluggast, der seinen gebuchten Flug antreten will, dies aber wegen Überbuchung nicht kann, zahlt die Fluggesellschaft diesem für Hotelkosten bzw. entstandenen Ärger eine Ausgleichszahlung von 1500€, erstattet dem Fluggast aber den Kaufpreis nicht zurück.

Der Fluggast erhält jedoch einen Flug zum gebuchten Ziel mit einer anderen Maschine. Der Fluggesellschaft entstehen hierdurch Kosten von 200€.

Die folgende Teilaufgabe bezieht sich auf diese Praxis der Fluggesellschaft, Überbuchungen zuzulassen und Ausgleichszahlungen zu leisten.

f) Begründen Sie, dass die im Mittel pro Flug zu erwartende Anzahl von Fluggästen, die wegen Überbuchung abgewiesen werden müssen, mit dem folgenden Term berechnet werden kann:

$$\sum_{k=1}^{8} \binom{108}{100+k} \cdot 0,9^{100+k} \cdot 0,1^{8-k}$$

Berechnen Sie diesen Term und beurteilen Sie begründet, ob sich die Praxis der Überbuchung mit 8 zusätzlichen Plätzen aus Sicht der Fluggesellschaft lohnt.

Example: KMK12Abi, Fragments 6 & 7 (Pages 41, 42), English Translation:

Task

On a particular route, an airline uses airplanes with 100 seats each, which are booked and paid for before the flight.

As a simplification, it should be assumed that

- each place is booked individually,
- on this route, all places are booked,
- a place whose posting is canceled can not be posted again.

On average, 10% of the booked seats are canceled at short notice (that is, canceled from the people who have booked).

From a person who actually flies, the airline takes 200€. In case of cancellation, a partial reimbursement to the person is 100€ only.

(a) Name an assumption so that the possible number of cancellations in a flight can be modeled as a binomial distribution. Describe a real situation where this assumption is not true.

In the following it is now assumed that the possible number of cancellations during a flight is actually binomial distributed.

(b) Given are the following three events:

E1: "Exactly 90 seats are actually used during a flight."
E2: "A maximum of 90 seats are actually used during a flight."
E3: "At least 90 seats are actually used during a flight."

Set up a respective term, with which the probability of the corresponding event can be calculated.

(c) Calculate the revenue the airline can expect in the long run on average per flight.

In order to make the aircraft more efficient, the airline offers 108 seats instead of 100 for each flight, which is more than 8 available. All offered places are always booked. Even if an average of 10% of all bookings are canceled, the airline takes the risk that a person who has booked a place can not make their flight, since the booked space is not available.

(d) Calculate the expected revenue of the airline per flight, taking into account the cancellations from the ticket sale.

(e) Calculate the likelihood that at least one passenger who has not canceled can not make his booked flight due to overbooking.

Whether the business practice of allowing overbooking is always more economic for the airline is to be investigated further.

For every passenger who wants to take his booked flight, but is unable to do so due to overbooking, the airline pays a compensation of 1500€ for hotel costs and / or resulting anger, but does not refund the purchase price to the passenger.

However, the passenger receives a flight to the booked destination with another machine. The airline will incur a cost of 200€.

```
The following sub-task refers to this airline's practice of allowing overbooking and
making compensation payments.

(f) Explain that the number of passengers per flight that is expected to be rejected on
account of overbooking can be calculated with the following term:
```

$$\sum_{k=1}^{8} \binom{108}{100+k} \cdot 0.9^{100+k} \cdot 0.1^{8-k}$$

```
Calculate this term and assess your reasons for the fact that the practice of
overbooking with 8 additional seats from the perspective of the airline is worthwhile.
```

This task is an exemplar offered in the Bavarian college entrance exam standards (mathematics, 2012), which is designed to be related to modeling competence. A high school student may not be very familiar with booking flight tickets, but the context is a very common business activity that can be understood. Here are the modeling steps involved:

[1] *Pose a real-world question.* The real-world situation is a certain airline that provides reservation of flight tickets. A possible question is the revenue of a certain flight, which is posed in question (f) by the task description.

[2] *Model conceptually.* It can be identified that some important concepts are the cost of a flight, number of seats, ticket price, how likely a person cancels the trip, and refund policy due to cancellations. The task assumed and presented these concepts. There is no process of conceptual modeling.

[3] *Idealize.* There are some hidden assumptions that necessitate the practice of flight booking, i.e. the cost of each flight is high, and depends little to the actual number of passengers. The problem also provided three simplifications that needs to be understood by the problem solver. A real model of the problem is 100 independent bookers who have 10 % of cancellation rate. However, since the independence was not explicitly stated, it is treated as a point of assessment, i.e. question (a).

[4] *Mathematize.* A binomially distributed variable is a suitable mathematical model of this

problem. The problem solver needs to understand it, but not necessary to create or derive a model. Fluency of the particular model is needed in order solve question (b), since statements of random events have to be translated into mathematical relations in order to solve correctly.

[5] *Solve mathematically.* Usual calculations of probability and binomial distributions are needed from questions (b) to (f). No mathematical proof is required.

[6] *Interpret.* Question (f) requires an interpretation of a numerical result, concluding whether it is financially worthwhile for the described practice.

[7] *Verify / Criticize.* Question (a) asked the problem solver to challenge the use of the model by making an alternative assumption.

[8] *Re-iterate the modeling process.* No re-iteration is required, even alternative assumptions exist, that there may be groups of passengers instead of all independent individuals. Modeling under the alternative assumption is out of the scope of the task.

Note that eight modeling steps were also mentioned in the beginning of Chapter 2, but the step *Field Choice* is not included here, since it can be absorbed into [4] and [5] — In [4], the mathematization may need to be done within a certain field of mathematics; in [5], the mathematical methods can also be drawn from related fields. Furthermore, one more step "Re-iterate" is added to the end for a possibility that the modeling cycle has to be done again due to an unsatisfying result of [7]. The eight modeling steps are ordered by the modeling cycle, but such order is not necessarily followed by a student who actual performs the task.

Table 5.6 serves as an overview of how the modeling steps are covered in this task. It is called a *modeling spectrum*[1] of the task, which is a record with 8×3 entries. For each modeling step from [1] to [8], the task description may show that they have done such a step, and if so, the student

[1]Since the way we analyze assessment tasks can be seen as an *atomistic* approach, the term *spectrum* is borrowed, though it is not the best analogically: a spectrum in optics shows intensity over a range of wavelengths, whereas the modeling spectrum shows how each modeling step is presented in the task. It may fit well to use a term like *modeling profile*.

Table 5.6: A modeling "spectrum": modeling steps involved in a sample task

Task Name: Flight Booking 1
Context: Business — Flight ticketing
Mathematical knowledge: Probability — Binomial distribution

Modeling Steps	Done by the task description:	Student needs to understand:	Student needs to perform:
[1] Pose Real-world question	Yes (f)	Yes	No
[2] Conceptual modeling	Yes	No	No
[3] Idealization	Yes	Yes	Yes (a)
[4] Mathematize	Yes (a, f)	Yes	Yes (b-e)
[5] Solve Mathematically	No	NA	Yes (c-f)
[6] Interpret	Yes (f)	Yes	Yes (f)
[7] Verify and/or Criticize	Yes (a)	Yes	Yes (a)
[8] Re-iterate	No	NA	No

may need to understand how it was done in order to solve the task; the task may also require the student to perform the modeling step[2]. Therefore there are 8 rows for each modeling step, and 3 columns as *Done by the task description*, *Student needs to understand*, and *Student needs to perform*.

The column, *Done by the task description*, is recorded "Yes" if the task description has shown how to do a certain step, or it can be inferred that the step must have been done before the task can be solved. The column, *Student needs to understand*, is only meaningful if the step is *done by the task description*; otherwise it is recorded as *not applicable* ("NA"). If a modeling step is already done by the task author (or has been shown how to do), does the student need to understand it in order to finish the task? If so, both columns are recorded "Yes"; otherwise, *Student needs to understand* is recorded as "No". For these two columns, let us take the first two steps in Table 5.6 as an example. For step [1], the task description effectively poses a real-world relevant question in (f): "Is the practice of flight overbooking profitable?" Such a question is vital in a modeling task, as it serves as the purpose of doing the hard work. Understanding this question is also necessary to perform the task, hence the "Yes, Yes" in the first two columns. The "(f)" means that the step is done in sub-question (f) in the task, since the task has a lot of sub-questions. As for

[2]The task might require the student to perform a modeling step completely by himself, or partially, with some help from the task descriptions, but this is not recorded in the modeling spectrum.

step [2], conceptual modeling is supposed to be done by the task, since the "important objects and important relationship" are already identified, including the *ticket price*, *cancellation policies*, and so on. However, it is possible for a student to simply work on the task without being aware of the conceptual modeling process. Thus "Yes, No" are recorded in the first two columns for this step.

The column, *Student needs to perform*, records if a student needs to work on a corresponding step in order to finish the task. To completely solve the *Flight Booking 1* task, a student needs to work on five modeling steps[3]: to do some idealization as required by question (a) (e.g. to assume that flight bookers behave independently), some mathematization of the verbal statements from (b) to (e), much mathematical work that involve calculations of probabilities and expectations (c-f), to explain why the numerical result in (f) shows that the practice is worthwhile, and, finally, to criticize the independence assumption by providing counterexamples. Many of the steps are not expected to be done by the student independently. Four of the five steps are also *Done by the task description*, which can be understood as "scaffoldings" provided by the task. The step that does not have scaffoldings is [5] *Solve mathematically*, for this step is recorded as "(No, N/A, Yes)". This is an example that shows how a task involves many modeling steps and how the modeling spectrum is recorded, but most tasks that we see in the following analyses does not involve this many steps. Indeed, a task that involves four steps or more is not typical, and the most popular tasks cover only one or two steps.

In the following sections, results of the analysis will be presented for each system. We will also highlight some findings that may contribute to the understanding of the thesis. The "modeling

[3]Besides working on modeling steps, the task also allows students to be creative, pose more real-world questions, and extend the model. For example, suppose a passenger knows that a flight he booked is so popular that overbooking is very likely, then he might "plan" to be solicited to take a later flight, as well as getting a compensation. This imagined real-world scenario may or may not occur, but does create more modeling opportunities as require more mathematics, e.g. game theory between the flight company and the customer. If a task likely allows such extensions, we may say that the task is *extensible*, and it may be beneficial to analyze the *extensibility* of a set of tasks. This is not be covered in the current work, but a further study can be done.

spectrum” depicted in Table 5.6 is used as a tool to classify items according to their coverage of the modeling steps.

Choice of curricular documents

PISA

The Programme of International Student Assessment (PISA) is a project commenced by the OECD in 2000, with the purpose of evaluating the overall performance of 15-year old students within a certain country (or region). Up to now, there had been six international assessments conducted in three-year cycles. According to PISA's official documents (OECD, 2003, 2012), the assessments are not based on individual subjects, but rather on different kinds of *literacy*. For example, a mathematical literate student should be able to recognize and make use of basic mathematical ideas in various contexts, including daily life situations, professional/occupational situations, and scientific contexts.

Since the main focus in each assessment cycle can be different, we only draw on materials from the two PISA cycles that focused on mathematical literacy, namely, PISA 2003 and 2012. Official assessment frameworks (AF) are reviewed, along with 27 published items, collected from official sources from different countries (PISA2003AF, PISA2012AF, USA 62RMI, USA 03RMI, Dec06, HKS, DEU). Indeed, there are 213 items that have been used in one of the PISA assessments from 2000 to 2012, or in one of the field trials in 1999 and 2002, but there are only 124 published items that are available. We have chosen all items that fit one of the following groups: (1) Published items that are used in PISA 2003 and at least one more cycle (20 items); (2) Published items that are illustrated as examples in the official assessment frameworks (7 items). It turned out that two groups do not overlap. Also, it was found that no published item was tested in both PISA 2012 and another cycle. A list of reviewed items is presented in the following tables.

In this study, we have chosen three participants of PISA, including USA, Germany, and Hong Kong. According to OECD's reports (OECD, 2003, 2015), United States' overall performance has

Table 5.7: Published items, tested in both PISA 2003 and 2000 (available in Dec06, USA 03RMI, HKS, and DEU)

M124Q01	P2000 Walking
M124Q03T	P2000 Walking
M145Q01T	P2000 Cubes
M150Q01	P2000 Growing Up
M150Q02T	P2000 Growing Up
M150Q03T	P2000 Growing Up
M179Q01T	P2000 Robberies
M266Q01T	P2000 Carpenter

Table 5.8: Published items, tested in both PISA 2006 and 2000 (available in USA 62RMI and USA 03RMI)

M302Q01T	Car Drive
M302Q02	Car Drive
M302Q03	Car Drive
M421Q01	Height
M421Q02T	Height
M421Q03	Height
M598Q01	Making A Booklet
M710Q01	Forecast of Rain
M810Q01T	Bicycles
M810Q02T	Bicycles
M810Q03T	Bicycles
M833Q01T	Seeing the tower

Table 5.9: Items that were illustrated in PISA 2012 assessment framework (available in PISA2012AF)

M535Q01	Twisted Building Q1
M535Q02	Twisted Building Q2
M535Q03	Twisted Building Q3
M535Q04	Twisted Building Q4
M537Q01	Heart Beat Q1
M537Q02	Heart Beat Q2
M552Q01	Rock Concert

Table 5.10: Some PISA 2003 and 2012 results.

PISA	2003	2012
OECD mean	500	500
USA	493	481
Germany	490	514
Hong Kong	560	561

been average or lower than average compared to other participating countries, and is still declining. Germany had lower performance than the USA in 2003, but has been significantly improving since 2006. Hong Kong has been among the top performers and has only small changes over the years.

In the following, collections of curriculum documents from these three countries/regions will be described, so that we may study their modeling contents through an analytical framework.

USA

The legislation of education in United State of America is up to each state, and there has not been a national standard of curriculum until the Common Core State Standards is adopted by more than 40 states, as of 2014 (CCSSO, 2014). Some states had earlier adoption of the standards, but the decision is up to individual states, and it is reasonable to say that there is no national curriculum. Three states, Florida, Massachusetts, and Connecticut, are especially appropriate for a curriculum comparison because they also actively participated in PISA 2012 assessment as individual school systems. It means that they are supposed to maintain a more transparent and systematic documentation of their curriculum standards and test papers (convenient sample). For simplicity, we will focus only on Florida and Massachusetts for an investigation into their curricula. We are also able to obtain older and newer documents, roughly corresponding to the two mathematically focused PISA cycles, 2003 and 2012.

Florida documents

- FCFM96. Florida Curriculum Framework: Mathematics. PreK -12 Sunshine State Standards and

Instructional Practices (State of Florida, Department of State (1996).
- FIMS01. Florida's Instructional Materials Specifications for the 2003-2004 Adoption Grades K-8
- FCAT04. FCAT Mathematics Test Book, Grade 10, Released August 2005, Last used March 2004.
- FCAT04D. Florida Comprehensive Assessment Test Summary of Tests and Design.
- NGSSS Next Generation Sunshine State Standards (2007) and the corresponding mathematics framework
- FCAT10. FCAT2.0 Mathematics Test Item Specifications Grades 6-8
- EOCA1. Algebra 1 End-of-Course Assessment Test Item Specifications Version 2
- EOCG. Geometry End-of-Course Assessment Test Item Specifications Version 2

Massachusetts documents

- Massachusetts Department of Education (November, 2000), Massachusetts Mathematics Curriculum Framework
- Massachusetts Department of Education (March, 2011), Massachusetts Mathematics Curriculum Framework
- Massachusetts Comprehensive Assessment System, 2004 released mathematics items, grade 10
- Massachusetts Comprehensive Assessment System, 2012 released mathematics items, grade 10

In both Florida and Massachusetts, the public examinations cover up to grade 10. There is no known public examinations for grade 12 students. Those who are pursuing for higher education are subject to the requirements of individual universities or colleges. This is very different from the case of Germany and Hong Kong, both having a public examinations for university entrance.

Germany

Germany has 16 federal states, which also have their own legislation for education. However, these states already adopted the same national standards in 2004, shortly after the "PISA shock", or "TIMSS shock" occurred earlier (Lehmann, 2011). The TIMSS is also a large scale assessment that began its assessment cycle in 1995. The first released results annoyed many of the German educators for they were quite confident in their long tradition of education. Indeed, German education system was not greatly influenced by the reunification in 1990. The first PISA results, released in 2001, triggered a major reform in German education. Indeed, many of the curriculum documents, which are related to the reform, discussed PISA's assessment standards and emphasized the importance of application.

Therefore, it can be assumed that Germany had a more diverse system before 2004, and more standardized after 2004. Even within a state, the system is still more complicated than in other countries since there are different educational paths existing within one system. After graduated from Grundschule (primary school, ranging from 4 to 6 years in different places), a German student can continue upon their secondary education in a *Hauptschule*, a *Realschule*, or a *Gymnasium*. According to *Encyclopedia Britannica*, a *Hauptschule* is "a five-year upper elementary school that prepares students for vocational school or trade apprenticeship", a *Realschule* is "a six-year secondary course preparing students for higher business and technical schools", and a *Gymnasium* is "a nine-year university preparation school". However, these schools are also regulated differently in different states. Approximately saying, the *Gymnasium* is the most demanding and the *Hauptschule* is the least. To generate less potential confusions, we will keep the names of different kinds of schools untranslated. As a representative, we choose Bavarian *Hauptschule* and *Gymnasium*, and study their curricula, including older and newer ones.

The state of Bavaria is largest in area, and is the second most popular state. The documents that are chosen are listed as follows:

- *Lehrplan für die Hauptschule 1997* (Curriculum plans for the *Hauptschule* 1997), published by the Bavarian State Ministry for Education, Culture, Science and Art.
- *Besondere Leistungsfeststellung (Mathematik) zum Erwerb des qualifizierenden Hauptschulabschlusses 2001* (Specific assessment of mathematics for the acquisition of the Hauschule diploma qualification in 2001)
- *Lehrplan für die Hauptschule 2004* (Curriculum plans for the *Hauptschule* 2004)
- *Besondere Leistungsfeststellung (Mathematik) zum Erwerb des qualifizierenden Hauptschulabschlusses 2012* (Specific assessment of mathematics for the acquisition of the Hauschule diploma qualification in 2012)
- *Lehrplan für das bayerische Gymnasium 1990* (Curriculum plans for Bavarian *Gymnasiums* 1990), published by the Bavarian State Ministry for Education, Culture, Science and Art.
- *Abiturprüfung 2001 Mathematik* (Abitur examination in 2001, mathematics)
- *Lehrplan für das bayerische Gymnasium 2004* (Curriculum plans for Bavarian *Gymnasiums* 2004).
- *Abiturprüfung 2012 Mathematik* (Abitur examination in 2012, mathematics)

Hong Kong

Hong Kong is a small region that has a population of seven to eight millions. It was once a British colony, but its sovereignty has been handed over to the People's Republic of China in 1997. Even so, the Government of Hong Kong has its independent rights and responsibility on education affairs. Being a highly centralized education system, Hong Kong "exam culture" has been dominant since the twentieth century (Tang & Bray, 2000). The education system of Hong Kong is unified and stable enough so that a systematic record of documents is available through the government's website. The following documents will be reviewed:

- Education Department, Syllabuses for Secondary Schools, Mathematics, Secondary 1-5, 1999
- Mathematics Education Key Learning Area Curriculum Guide (Primary 1 - Secondary 3) (2002)
- Hong Kong Certificate of Education Examination 2002, mathematics (paper 1 and paper 2)
- Hong Kong Special Administrative Region Government, Mathematics Education Key Learning Area, Curriculum and Assessment Guide (Secondary 4 - 6), 2007
- Hong Kong Diploma of Secondary Education, 2012 Examination, mathematics

Analysis of PISA tasks

The analysis that was done on PISA tasks proceeded in two parts. Firstly, it is necessary to apply the analysis tool 5.4 to the official framework of PISA, since the role of modeling was explicited stated there. After that, the tool was applied to some of the published PISA items.

In the *PISA 2003 Assessment Framework* (OECD, 2004), 64 fragments are extracted and saved in a file. Each fragment is either a few sentences, some bullet lists, or a figure that is related to one or more of the questions in the analysis tool. The core concept of PISA's mathematics assessment was *mathematical literacy*, which was defined as

> ... an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgements and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen. (OECD 2004, p. 15)

Although PISA is not a direct assessment of modeling, there is much overlap between modeling and PISA's definition of mathematical literacy. In the analysis, it was found that modeling plays a role in many parts of PISA's assessment design. In terms of the construct that PISA assesses, mathematical literacy emphasizes the ability of using mathematical ideas to solve problems which can occur in situations in "the world". As an activity, this is very similar to mathematical modeling. In this respect, mathematical literacy can be seen as an attribute of an ideal citizen, who is able to use his/her mathematical knowledge in order to make judgments and evaluations, based on the real situation. Modeling ability is therefore required for a "mathematically literate individual". In terms of task design, modeling served as a great source of idea in problem posing. Indeed, what PISA's description of their item design is very closely related to a modeling cycle:

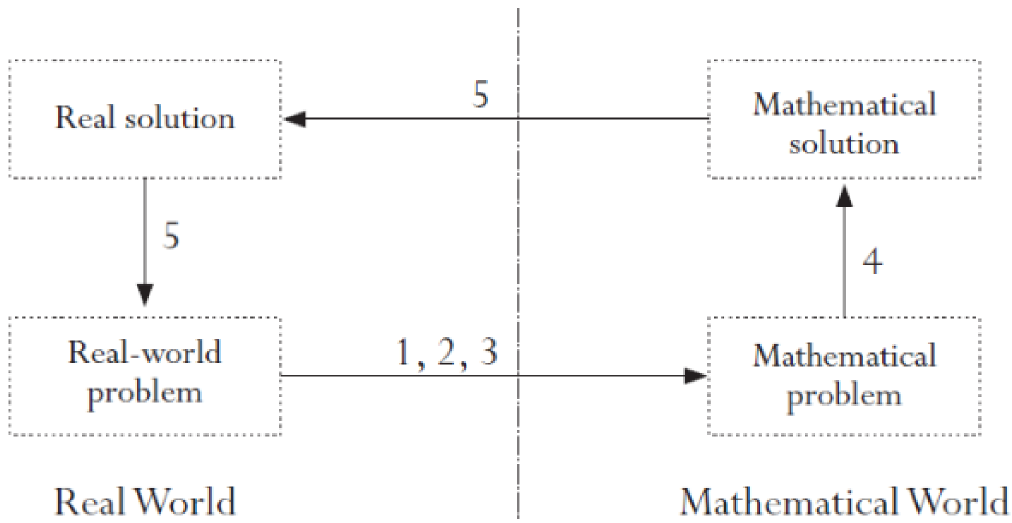
> [To] judge whether 15-year-old students can use their accumulated mathematical knowledge to solve mathematical problems they encounter in their world, one would collect information about their ability to mathematise such complex situations. Clearly this is impractical. Instead, OECD/PISA has chosen to prepare items to assess different parts of this process. (OECD 2004, p. 29)

The *different parts* of the mathematization process include five aspects, which fits the modeling cycle quite well, though PISA actually called it "the mathematisation cycle" (Figure 5.1). The five aspects include:

- Starting with a problem situated in reality;
- Organizing it according to mathematical concepts;
- Gradually trimming away the reality through processes such as making assumptions about which features of the problem are important, generalizing and formalizing;
- Solving the mathematical problem; and
- Making sense of the mathematical solution in terms of the real situation.

These aspects are obviously aligned with the modeling cycle, and therefore PISA's mathematization process can be seen as identical to a modeling process. It is then possible to use Table 5.4 to analyze the PISA materials in modeling terms. The first group of questions about modeling competence is partially answered in the above discussion, that modeling cycle is equivalent to the "mathematising process", which was stated as a crucial part of the PISA tasks to assess mathe-

Figure 5.1: PISA's Mathematisation Cycle



This figure shows PISA's version of modeling cycle (they called it the "mathematisation cycle") (Organisation for Economic Co-operation and Development [OECD], 2004, p. 38). It is treated as a "fundamental process that students use to solve real-life problems". The five aspects of mathematisation are indicated by a number (1-5) in the figure, connecting the "Real-world problem", "Mathematical problem", "Mathematical solution", "Real solution", then back to "Real-world problem", in that order.

matical literacy. Since modeling is only a tool to assess mathematical literacy in PISA, modeling competencies were not discussed directly. Instead, we may take a look at the competencies that mathematical literacy is concerned, and see how well they match with modeling competencies. The competencies for mathematical literacy were adopted from the work of Niss (2003):

1 Thinking and reasoning
2 Argumentation
3 Communication
4 Modeling
5 Problem posing and solving
6 Representation
7 Using symbolic, formal and technical language and operations
8 Use of aids and tools

Modeling is listed as one of the 8 competencies, but one may also see that, in other competencies definition such as in Maaß (2007), most of the above (other than (4) modeling) can be seen as modeling competencies, especially (2), (3), (6), and (8). Note that (1) *Thinking and reasoning* and (5) *Problem posing and solving* are too general to be treated as a competency that is specific for modeling. The (7) *using symbolic, formal and technical language and operations* is important for creating and working within a mathematical model, but is usually not the main concern of modeling competencies. To answer the question *what behaviors are expected of competent modelers*, we refer to PISA's three *competency clusters*, including reproduction, connections, and reflection. Each competency cluster involves all 8 competencies above, but here are descriptions of modeling competency within each cluster:

> The reproduction cluster, modeling:
> This involves recognising, recollecting, activating, and exploiting well structured familiar models; interpreting back and forth between such models (and their results) and "reality"; and elementary communication about model results.
>
> The connections cluster, modeling:
> This involves structuring the field or situation to be modelled; translating "reality" into mathematical structures in contexts that are not too complex but nevertheless different from what students are usually familiar with. It involves also interpreting back and forth between models (and their results) and "reality", including aspects of communication about model results.
>
> The reflection cluster, modeling:
> This involves structuring the field or situation to be modelled; translating "reality" into mathematical structures in contexts that may be complex or largely different from what students are usually familiar with; interpreting back and forth between models (and their results) and "reality", including aspects of communication about model results: gathering information and data, monitoring the modelling process and validating the resulting model. It also includes reflecting through analysing, offering a critique, and engaging in more complex communication about models and modelling.

These descriptions are of interest if compared to Greer and Verschaffel (2007), who describe three levels of modeling activities:

> Implicit Modeling: Student is essentially modeling without being aware of it
> Explicit Modeling: Student's attention is drawn to the modeling process
> Critical Modeling: The roles of modeling relative to the different realms are critically examined

Greer and Verschaffel's levels are mostly related to the consciousness of a modeling process, whereas PISA focuses more on the complexity of the modeling tasks, and the distance from students' familiar contexts. The next group of questions in the analysis tool are about the design of tasks. In PISA, mathematical literacy is assessed in relation to the mathematical *content*, the *process*, and the *situations* (OECD, 2004, pp. 15-16), and each PISA item is classified by each of these aspects. PISA's mathematical content is defined through a "phenomenological approach" (p. 34), meaning that PISA does not directly adopt traditional school mathematics as a definition of mathematical content, but looks for groups of mathematical idea that are fundamental to mathematical activities in the real world. These groups, called "overarching ideas", include *quantity*, *space and shape*, *change and relationships*, and *uncertainty*. The cognitive processes which undergo with students' work on the PISA mathematics items are generally classified as *reproduction*, *connection*, and *reflection*, roughly depending on how much creativity is needed during a typical 15-year-old students' thinking process. Finally, it is preferred that an item is authentic, namely, the item has a context in a real-world situation. PISA's definition of *situation* and *context* are clear: "The situation is the part of the student's world in which the tasks are placed. It is located at a certain distance from the students. [...] The context of an item is its specific setting within a situation. It includes all the detailed elements used to formulate the problem. (p. 32)"

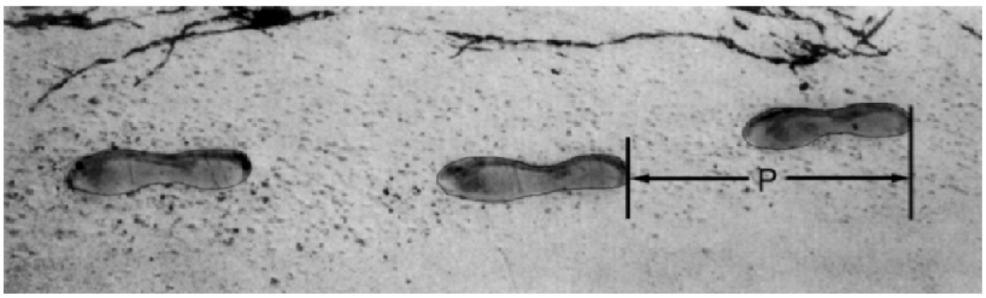
PISA mathematics items are classified in 4 types of situations: personal, educational/occupational, public, and scientific. Besides content (one of the 4 overarching ideas), process, and situation, a PISA item also needs a specific format of testing, which can be multiple choice, open-constructed response, or closed-constructed response. Furthermore, it is a fairly common practice that two or more items (questions) are put together as a bundle, sharing the same context and under the same title, but it is likely that the questions assess different competencies or require different levels of thinking. In the following, we present some of the PISA items under our analysis. There are currently 124 published items, under 62 titles. The released items can be found on PISA's official website at OECD, mainly in the following two documents:

- PISA Released Items – Mathematics (OECD, December 2006)
- PISA 2012 Released Mathematics Items (OECD, May 2013)

Figure 5.2, 5.3, and 5.5 (pp. 121, 122, & 124) show three PISA items: *Walking*, *Carpenter*, and *Height*. None of these involve a whole modeling cycle. *Walking* has a picture and a short description about what the pacelength mean, and then presents a formula that approximates the

Figure 5.2: PISA item: *Walking*

M124: Walking



The picture shows the footprints of a man walking. The pacelength *P* is the distance between the rear of two consecutive footprints.

For men, the formula, $\frac{n}{P} = 140$, gives an approximate relationship between *n* and *P* where,

n = number of steps per minute, and

P = pacelength in metres.

Question 1: WALKING *M124Q01- 0 1 2 9*

If the formula applies to Heiko's walking and Heiko takes 70 steps per minute, what is Heiko's pacelength? Show your work.

Question 3: WALKING *M124Q03- 00 11 21 22 23 24 31 99*

Bernard knows his pacelength is 0.80 metres. The formula applies to Bernard's walking.

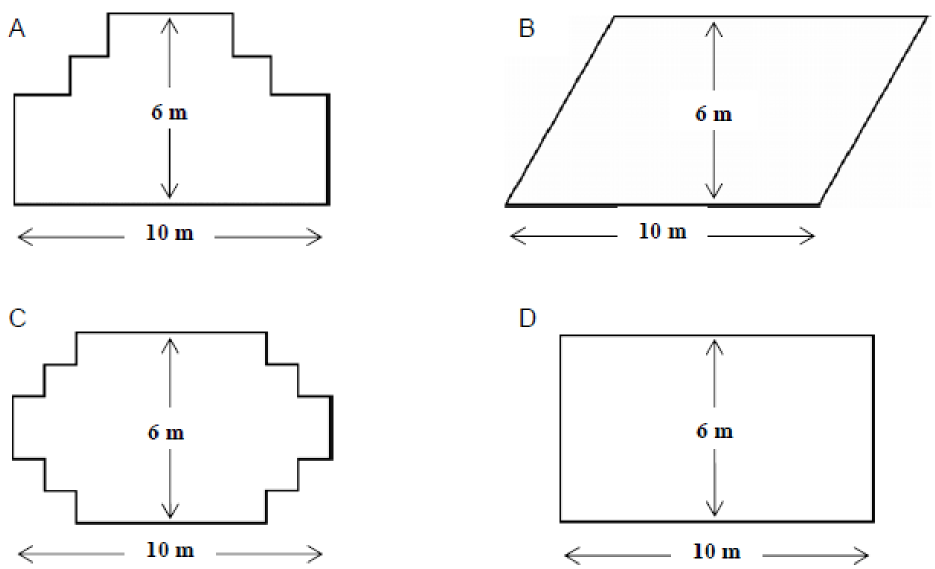
Calculate Bernard's walking speed in metres per minute and in kilometres per hour. Show your working out.

This item, *M124: Walking*, contains a picture of footprints followed by a short description of pacelength, P, and its approximate relationship with the number of steps per minute, n. After the description, two quantitative questions were asked on the situation. Adapted from "PISA released items - Mathematics" (p. 8), by Organisation for Economic Co-operation and Development [OECD], 2006, OECD Publishing. Copyright 2006 by OECD.

Figure 5.3: PISA item: *Carpenter*

Question 1: CARPENTER *M266Q01*

A carpenter has 32 metres of timber and wants to make a border around a garden bed. He is considering the following designs for the garden bed.

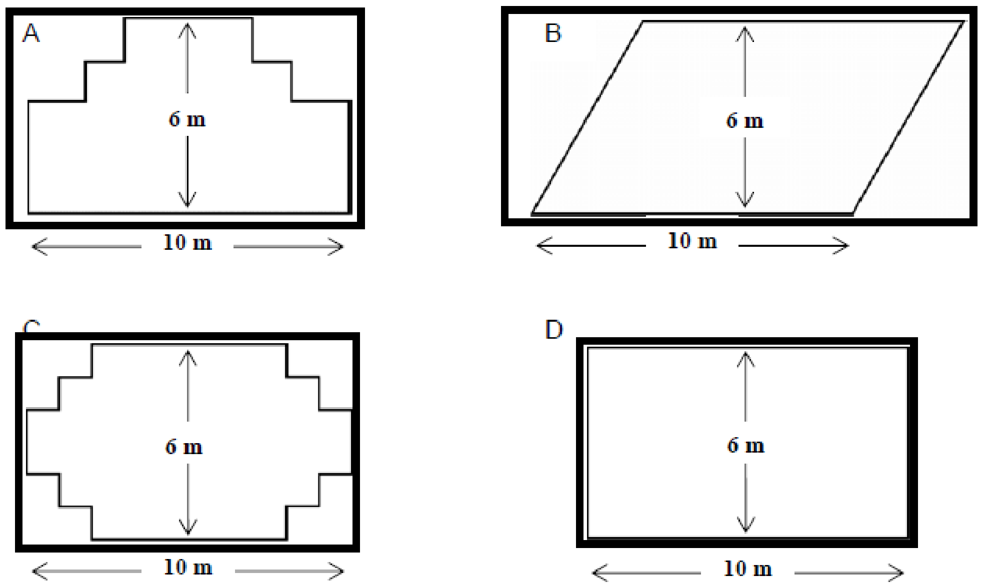


Circle either "Yes" or "No" for each design to indicate whether the garden bed can be made with 32 metres of timber.

Garden bed design	Using this design, can the garden bed be made with 32 metres of timber?
Design A	Yes / No
Design B	Yes / No
Design C	Yes / No
Design D	Yes / No

This item, *M266: Carpenter*, describes a design problem of a carpenter who wants to make a border around a garden. The short description is followed by four different graphs of design, and then four yes-no questions about the design. Adapted from "PISA released items - Mathematics" (p. 39), by Organisation for Economic Co-operation and Development [OECD], 2006, OECD Publishing. Copyright 2006 by OECD.

Figure 5.4: Wrong solution *Carpenter* with correct answer



The figure shows how the *Carpenter* item can still be answered correctly by mistakenly drawing rectangular borders.

relationship between pacelength and the number of steps per minute. Therefore, a mathematical model is already given, so that a student does not have to choose important variables or make assumptions, but only need to identify the meaning of the variables. Indeed, in question 1 of *Walking*, even the meaning of the variable is not needed [4], and therefore it was classified by PISA as in the reproduction cluster. In the modeling competence framework, this question tests if a student can understand the mathematical equations and work on it, without much thinking about the real context. Question 2 requires a little more connection between mathematics and reality because the student has to figure out that speed is the product of pacelength and the number of steps per minute; this information must be reasoned from the context, but not mathematics alone.

[4] Indeed, a few things can *perplex* a student: (1) Why, for all men, n/P should be a constant? (2) If one calculates the speed, $nP = 140P^2$, it is found to be proportional to P^2, which does not seem to be reasonable; and (3) The unit of n/P does not match the right side, 140, unless one realized that the pacelength P is redefined as the *number of metres* of pacelength. Students do not need to think about any of these. Actually, pondering anything else than "plugging in" the numbers directly is implicitly discouraged in a timed test.

Figure 5.5: PISA item: *Height*

HEIGHT

There are 25 girls in a class. The average height of the girls is 130 cm.

Question 1: HEIGHT *M421Q01 - 0 1 9*

Explain how the average height is calculated.

Question 2: HEIGHT *M421Q02*

Circle either "True" or "False" for each of the following statements.

Statement	True or False
If there is a girl of height 132 cm in the class, there must be a girl of height 128 cm.	True / False
The majority of the girls must have height 130 cm.	True / False
If you rank all of the girls from the shortest to the tallest, then the middle one must have a height equal to 130 cm.	True / False
Half of the girls in the class must be below 130 cm, and half of the girls must be above 130 cm.	True / False

Question 3: HEIGHT *M421Q03*

An error was found in one student's height. It should have been 120 cm instead of 145 cm. What is the corrected average height of the girls in the class?

A 126 cm
B 127 cm
C 128 cm
D 129 cm
E 144 cm

This item, *M421: Height*, provides the average height of a class and then asked conceptual questions. Adapted from "PISA 2012 released mathematics items" (p. 3), by Organisation for Economic Co-operation and Development [OECD], 2013b, OECD Publishing. Copyright 2013 by OECD.

Therefore, student needs to work on one more part of the modeling cycle (from the *conceptual model* to the *mathematical model*).

The item *Carpenter* is arguably harder than *Walking*, for the calculation of perimeter of A and C is a challenging problem (if the student has not seen it before), and also the perimeter of B can only be bounded below but not determined. A student who successfully solve this question needs some skills of solving geometric perimeter problems, but the need of mathematization is minimal, since the geometrical shapes are well-defined mathematically, and there is no need to going "back to the real world". Furthermore, the mentioning of alternative designs of a garden does not make the problem more "real". Therefore, *Carpenter* can be seen as a "camouflaged" problem of mathematics. Note also that the problem can be answered correctly without right understanding of boundary (Figure 5.4), that is, even a student misunderstanding it as *area* can choose the correct answer. The item *Height* focuses on the concept of average, where Question 2 actually tests if a 15-year-old has misunderstanding about mode, median, and mean. For the Questions 2 and 3, there are two possible scenarios: (1) The student only knows the mathematical definition of average, and needs to judge if a certain property is always true, and to calculate a new average if some number changed, which may require some decent algebraic skills. (2) The student not only knows the definition, but also have much contextual understanding (or misunderstanding) of it, which helps (or hinders) him to solve the problem. Since such scenarios exist, this item is quite ambiguous about which sub-competence it measures. For a student who does not care much about the context, the average is a purely mathematical concept and so the question again only assesses how the student works mathematically; for a student who tries to understand the average in context, he may work on a conceptual model directly. In either case, only one step of the modeling process is concerned. Mathematization is again not necessary for the student. Appendix B (p. 223) shows how each PISA released item is related to the modeling steps, with

Table 5.11: Percentage of students who are able to do the problems

Walking, Question 1	36.7 % (getting 2 points)
Walking, Question 3	17.3 % (getting at least 2 points)
Carpenter	20.2 % (getting all correct)
Height, Question 1	65.7 % correct
Height, Question 2	18.1 % (getting all correct)
Height, Question 3	38.3 % correct

Table 5.12: Number of PISA items each modeling step involves, out of a sample of twenty-eight items

[1] real model	3
[2] mathematical model	8
[3] mathematical work	22
[4] interpretation	9
[5] verification	1

comments that explains such relation.

In the analysis, it is evident that PISA tests students "atomistically", focusing on each part of the modeling process and the related sub-competencies. Most of the mathematization was done, so that students only need to perform interpretations of small parts. However, even small interpretations turn out to be quite hard. For the item *Walking*, Question 3, only 17.3 % of the student (OECD average) obtained at least 2 points out of 3. According to the official coding scheme (OECD, 2006, p. 10), it means that more than 80% of the students were not able to get beyond the mere substitution of $n = 140P = 140 \times 0.80 = 112$. In other words, they never connected the steps per minute to the speed of the person.

Out of the 28 items that are reviewed, the vast majority of them (24) involve one or two steps of a modeling cycle, whereas 3 of them have three steps; one item has no modeling step at all. Each modeling step is involved in some of the items, and the number of them is shown in Table 5.12. In the following section, we will see how performance varies on different kinds of questions of different countries of students.

Analysis of Florida's curriculum materials

Florida, 1995-2003.

In this period, the *Sunshine State Standards* (1996) must be included in the instruction of Florida's public schools, and a public assessment, called the *Florida Comprehensive Assessment Test* (FCAT), is administered in grades 3 to 10 so as to measure if students were meeting the standards. Students must pass the FCAT of grade 10 in order to obtain a high school diploma. These are treated as Florida's intended and assessed curriculum in the period before 2003. It is emphasized in the documents that their framework paid much attention to the NCTM (National Council of Teachers of Mathematics) standards published in 1989. Although NCTM standards do not represent as a national standard, they are very influential, and many states had followed the ideas at least at the intended curriculum level.

Ideology is not usually stated, but may be inferred from some excerpts of the documents, and an assignment of multiple ideologies is possible. Florida's Curriculum Framework: Mathematics (FCFM1996) matches three of the six ideologies, namely, Social and economic efficiency, Orthodoxy, and Cognitive pluralism.

FCFM1996. Ideology: Social and economic efficiency.

An increasingly service-oriented, information based society that is virtually exploding with expanding knowledge demands that everyone have the opportunity to acquire the necessary skills to succeed in the information age. (p. 1)

These new conditions require citizens who are prepared to make well-reasoned, thoughtful, and healthy lifelong decisions in an ever-changing world. Students must learn how to locate, comprehend, interpret, evaluate, manage, and apply information from a variety of sources and media. They must learn how to communicate effectively in a variety of settings and for a variety of purposes through many different media. They must develop mathematical skills to analyze information, solve problems, and create products to meet new needs. They must become creative and critical thinkers, skilled in systematic problem solving. They must learn to wisely allocate resources used to solve problems. They must learn to understand systems and to use technology. They must develop the integrity to work cooperatively and effectively with people from many diverse backgrounds. (p. 2)

FCFM1996. Ideology: Orthodoxy.

Infusing a Multicultural Perspective: Florida students appreciate their own culture and the culture of others, understand the concerns and perspectives of members of other ethnic groups, reject the stereotyping of themselves and others, and seek out and utilize the views of persons from diverse ethnic, social, and educational backgrounds. (p. 158)

A carefully designed and continuous curriculum (preschool through 12th grade) can create the multicultural literacy so necessary for a healthy nation. Each cultural group has its own set of values and perspectives. Many of these values are shared with other cultures and form the basis of American national unity. (p. 158)

FCFM1996. Ideology: Cognitive Pluralism.

Seven types of intelligences have been identified by Howard Gardner (1985): verbal/linguistic, logical/mathematical, visual/spatial, body/kinesthetic, musical/rhythmic, interpersonal (dealing with other people), and intrapersonal (knowing oneself). Each student has a dominant learning style that consists of a unique combination of these intelligences. It is important for teachers to understand the learning styles of their students so that they can structure their teaching in a way that incorporates these seven ways of knowing. The mathematics program that matches teaching to learning styles allows students to process material more efficiently, thereby reaching all students and providing the opportunity for deeper and more thorough learning. (p. 162)

The "Orthodoxy" curriculum ideology may need further discussions since it usually refers to religious or political views manifested in curriculum materials. The view of multiculturalism is an important value that is to be pertained and reproduced through the curriculum. One can also find examples in more detailed materials that specify how to infuse multiculturalism into mathematics instruction:

FIMS01: Possible multicultural links in mathematics instructions

Mathematics provides abundant opportunities for multicultural links. Possibilities include historical connections and mathematical puzzles and games. For example, the mathematics we use today has contributions from all continents and peoples. Our numeral system is called the Hindu-Arabic system. The numerals, the zero concept, and the place-value system were invented by the Hindus and transmitted to northern Africa and Europe by Arabic traders and scholars. In Italy, the system eventually replaced the Roman numeral system and has been passed on to us. (p. 29)

These serve as background understanding of the curriculum, and now we turn to the role of modeling and application. The word *model* does not usually refer to a mathematical model, but rather a physical, concrete model, or a graphical model to show more abtract relationship. It does mention modeling in the mathematical sense, but only in two unimportant instances. Modeling cycle is not observed. The document does appreciate real-world application of mathematics. For example, it suggested that a good mathematical task is one that has "daily application in people's lives", and a teacher should be proficient in "providing opportunities to deepen students' understanding of mathematics and its applications" (p. 145).

At a deeper level, the curriculum document also recognizes the need to facilitate transfer of learning, and one way to do so is to provide curricular connections (p. 178). Many introductions were done on multidisciplinary instruction and transdiciplinary instruction. The purpose of making such connections is to enable students to solve practical problems in the real-world. Indeed, the word *real-world* is mentioned 75 times in the document. It is of particular interest when the document introduces *authentic assessment*, which is "a form of performance assessment structured around a real-life problem or situation" (p. 193). One example of a hypothetical situation was provided as follows:

FCFM1996. Authentic assessment.

```
   Students in Mrs. Martinez's class think that the price of pizza in the school
cafeteria is too high. Mrs. Martinez decides to use this issue in a problem-solving
activity in which her students will propose a fair pricing system for school-lunch
pizzas. This activity allows students to apply their computational and analytical
skills to an issue of direct relevance to them. Before the activity, Mrs. Martinez and
her students establish criteria for the pricing proposals and develop a scoring rubric.
Mrs. Martinez has her students research the prices of ingredients and estimate the cost
of making each pizza, the cost of labor, and overhead expenses. She also asks them to
incorporate the concepts of supply and demand that they are learning in social studies.
Students come up with fair pricing proposals and present them to the class. (p. 194)
```

The task involves an interesting real-world problem of pricing, and can potentially develop into a modeling task. It actually specified elements of *conceptual modeling*, since the teacher asks the students to research the important factors that comprise the final cost of making a pizza. The supply and demand model in economics can also be used. It can be considered as an open task of modeling, since the actual pricing method can be very diverse. However, standards of evaluating a proposal is not specified, and are supposed to be dependent on the teacher's scoring rubric and subjective evaluation, which is a common problem of project-based assessment (cf. Table 4.1). Furthermore, since project-based assessment is still nonexistent in public examinations, which is believed to dominate the *acquired* curriculum, we will not have further discussions on project-based or other subjective forms of assessment.

In the following, sample test items are reviewed from the persepctive of testing the ability to apply mathematical knowledge. In particular, a tool is developed to reveal the modeling steps involved in the tasks. The analysis has to be done more exploratively on Florida's tests in order to achieve inherent consistency of the analytical process, and make necessary adjustments. Hence, the presentation of Florida's results involves discussions of the analytical tools. Subsequent results of other regions will be presented more summatively, highlighting the differences found that can improve understanding of the thesis.

The public examination in Florida, FCAT, was designed to meet the *Sunshine State Standards* since 1998, and is replaced by FCAT 2.0 after 2010 (FDOE website). We choose FCAT04 (Florida Comprehensive Assessment Test, Grade 10 Mathematics, last used March 2004) because it is the earliest available one that is released on the website of Florida Department of Education. Assuming that the test contents are consistant from 1998 to 2010[5], FCAT04 can be seen as a representative of the public assessment in this period.The test has five content categories: (1) Number Sense, Con-

[5]There is one recorded change in 2004: the way to classify complexity level was based on Bloom's Taxonomy, but later changed to a method used in NAEP. It is emphasized that there is no change of the items themselves (Document FCAT04D, p. 4)

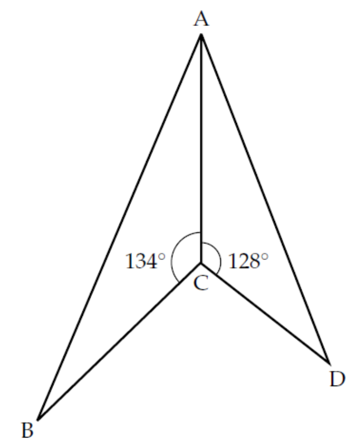
cepts, and Operations; (2) Measurement, (3) Geometry and Spatial Sense, (4) Algebraic Thinking, and (5) Data Analysis. Item types are extended response, gridded response, multiple choice (most frequent), or short response.

FCAT is consistent with the curriculum framework in the emphasis of authentic or real-world scenarios. In the broadest sense, if a mathematical task refers to any real-world ideas or contexts, we call it an *application task*, though further study is usually needed to see how relevant is the task to the real-world, and to what extent it tests the ability to apply mathematics. Among all the 50 problems in FCAT04, 47 of them (94%) are application tasks, and the other three do not present any real-world idea or context. However, the solution to some of the 47 application tasks do not seem to require any connection between the real-world context and the mathematical work. For example, in the *Nigerian Mask*[6] item (Figure 5.6), the actual mask is not given, but standard geometric shapes instead. For this problem, neither the problem writer nor the student need to perform the modeling process. Therefore it can be treated as a purely mathematical problem without a model. The mentioning of Nigeria is only a result of the multicultural requirements. Another kind of problem that does not need any modeling from the problem poser and the problem solver is like the following: "Arturo was evaluating some formulas as part of a science experiment. What is the value of the following expression? $(-2.1)^2 + (-0.5)^3$" (#36). The non-mathematical message that is conveyed may be "doing science experiment requires the mathematics you learned", but since the context is abstracted, this is still a "pure math" problem. These two problems can be classified as *context-detachable problem*, where the context can be daily-life, cultural, geographical, or professional. The problems does not involve any of the modeling steps. Though [5] solve mathematically is required, it is not solved within a model, since a mathematical model either does not exist, or is only assumed to exist. In FCAT04, four tasks are of this type.

[6]Names are sometimes given by the thesis author simply for convenience.

Figure 5.6: FCAT 2004 item, #44, Nigerian Mask.

44 The diagram below shows a design found on a mask from Nigeria. In the diagram, ∠ACB measures 134°, and ∠ACD measures 128°.



What is the measure, in degrees, of ∠BCD?

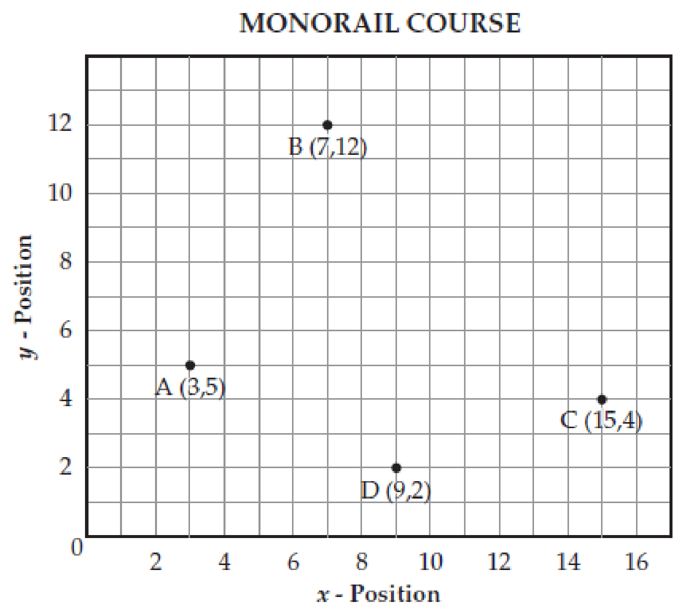
In another type of items, some real situation is already modeled, and the student has to do the mathematical work within the model. In the *Amusement Park Construction* task (Figure 5.7), the reason why the engineers want to use a parallelogram is not known, but student has to understand that point C is to be moved to a new location C' in order that the four points make a parallelogram. After such understanding of the geometric model, the student may solve it mathematically, by various available methods. The student also need to understand the idea of *relocating a supporting column located at a point* and realize that it is to say that point C has to be *deleted* and point C' is to be *created*. So the solution of the task also requires the student to (somehow) understand the conceptual model of the situation. The modeling spectrum of this item is shown in Table 5.13. All rows with entries (No, NA, No) are not shown.

Such items can be classified as a variation of *classical* applied mathematical tasks (or simply *classical* tasks), in which the mathematics that is used is apparently applied and it cannot be solved

Figure 5.7: FCAT 2004 item, #45, Amusement Park Construction.

The course of the monorail at an amusement park must be changed to make room for a new parking lot. Engineers have decided that only the main supporting column located at point C on the grid below should be relocated. They have also decided that the rebuilt course should be in the shape of a parallelogram.

Part A Plot the new location of the supporting column and write its coordinates. Label the new location C′.



Part B Use the definition or properties of a parallelogram to verify that the new monorail course is a parallelogram. You must use the slopes of the sides, the lengths of the sides, or both, to help verify your answer.

Table 5.13: Modeling spectrum of FCAT 2004, problem #45

Task Name: Amusement Park Construction
Context: Civil Engineering
Mathematical knowledge: Planar Geometry

Modeling Steps	Done by the task description:	Student needs to understand:	Student needs to perform:
[2] Model Conceptually	Yes	Yes	No
[4] Mathematize	Yes	Yes	No
[5] Solve Mathematically	No	NA	Yes

without some understanding of the situation. However, the student does not have to take part in the conceptual modeling or the mathematization steps, but has to understand one or both of them. Three possible spectra are of this kind as shown in 5.14. Since there are 32 classical tasks, they constitutes the majority type in FCAT04. What if some conceptual modeling and mathematization is presented in the task but a student does not have to understand them to solve the task? Two such task exist in FCAT04 and are presented in Figures 5.8 (FCAT04 item # 30, p. 135) and 5.9 (FCAT04 item # 46, p. 135). Such tasks can be called *camouflaged word problems*. In the first item, *Fishing Records*, the meaning of "Cost per Pound" is not clear: is it the cost of allocation or the cost for a customer who buys it in the market? The meaning of "Earned" is also not clear enough. Once again, the student does not have to think about all these to solve the problem. One way to *improve* the item is to ask the student to interpret what was meant by the words using the tables and equations.

In all the problem types described above (with 41 tasks, including 3 non-application tasks, 4 context-detachable, 32 applied mathematical tasks, and 2 camouflage word problems), the only actual step that a student needs to perform is to solve mathematically. The rest of the 9 problems require more modeling competence to solve the problem. Their modeling spectra (Table 5.16 are more complicated than the other problems, and are presented using the following scheme. For any row (i.e. each modeling step), the triple of the entries can be one of the five possibilities: (Yes Yes Yes), (Yes, Yes, No), (Yes, No, Yes), (Yes, No, No), (No, NA, Yes), (No, NA, No). They can be coded as if they are binary representations, namely 111, 110, 101, and so on.

The mode of these problems (four occurrences) is a type that can be named *standard mathematization* problem, where the problem statement has done the conceptual modeling; the student has to understand it, and to decide and perform the sort of mathematics that should be done. This type of items can be contrasted with the *applied mathematical problem (3)*, which provides the

Figure 5.8: FCAT04, #30, Fishing Records.

A fishing boat captain organized his fishing records in the table below.

APRIL FISHING RECORDS

Type of Fish	Cost per Pound	Pounds Sold	Total Dollars Earned
Snapper	\$5.50	s	\$8,250.00
Grouper	\$4.75	g	
Total		2,600	\$13,475.00

The following equations represent the information in the table, where s is the number of pounds of snapper and g is the number of pounds of grouper.

$$s + g = 2{,}600$$

$$5.50s + 4.75g = 13{,}475$$

The captain used these equations to determine that \$8,250.00 was received from the sale of the snapper. How many pounds of grouper, g, were sold in April?

Figure 5.9: FCAT04, #46, Volleyball Court.

The dimensions and shape of a volleyball court are shown in this picture. What is the approximate distance of a serve that is hit diagonally from one corner of the court to the other?

F. 27.0 meters
G. 20.1 meters
H. 15.6 meters
I. 12.7 meters

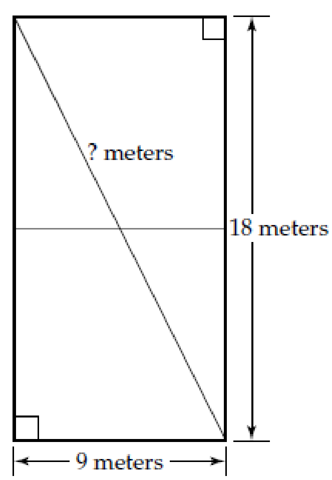


Table 5.14: Modeling spectra of applied mathematical problems

Classical Applied Mathematical Task (1) (18 out of 50 in FCAT04)

Modeling Steps	Described by the task	Student needs to understand	Student needs to perform
[2] model conceptually	Yes	Yes	No
[4] mathematize	Yes	Yes	No
[5] solve mathematically	No	NA	Yes

Classical Applied Mathematical Task (2) (4 out of 50 in FCAT04)

Modeling Steps	Described by the task	Student needs to understand	Student needs to perform
[2] model conceptually	Yes	Yes	No
[4] mathematize	Yes	No	No
(or:	No	NA	No)
[5] solve mathematically	No	NA	Yes

Classical Applied Mathematical Task (3) (10 out of 50 in FCAT04)

Modeling Steps	Described by the task	Student needs to understand	Student needs to perform
[2] model conceptually	Yes	No	No
(or:	No	NA	No)
[4] mathematize	Yes	Yes	No
[5] solve mathematically	No	NA	Yes

modeling step so that the student only needs to understand the mathematical model but not to mathematize themselves. Two examples (FCAT04, items # 37 & # 19) are shown in Table 5.15, one of each type[7].

All items recorded in Table 5.16 require a step of mathematization (the last "1" in [4] mathematize), and also some understanding of steps presented by the task (the middle "1" in the codes). An exception is problem #25 (as shown in the Figure 5.10, p. 138), where the student only has to perform the step of mathematical calculations, but has to understand three steps of modeling. We expect to see more examples of this type when we study test items from other school systems. In particular, the task describes that some snakes are spread in a desert preserve that is represented by a gridded map (*conceptual modeling*), and assumed that the snakes are evenly distributed

[7]In FCAT04 item # 37, the label "Calories Consumption" is quite misleading, since the unit that is in use is *kilojoule* instead of Calorie. A better word choice is "Energy Consumption".

Table 5.15: Standard mathematization (#37) and Applied mathematical task (3) (#19)

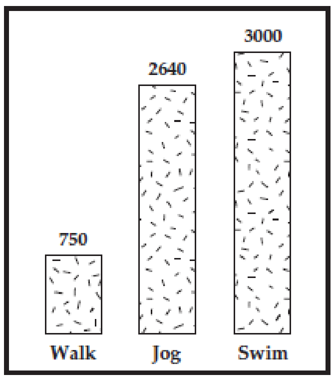
	#37 Calories Consumption	#19 Spring Stretch
[2] model conceptually	Yes Yes No	Yes No No
[4] mathematize	No NA Yes	Yes Yes No
[5] solve mathematically	No No Yes	No No Yes

37 A nutritionist has a female client who has been inactive, but plans to begin swimming one hour each day. This change in her level of activity will result in her daily energy requirement increasing by the same amount she expends swimming. According to the information below, what will be the woman's new daily energy requirement in kilojoules (kJ) after she begins swimming one hour each day?

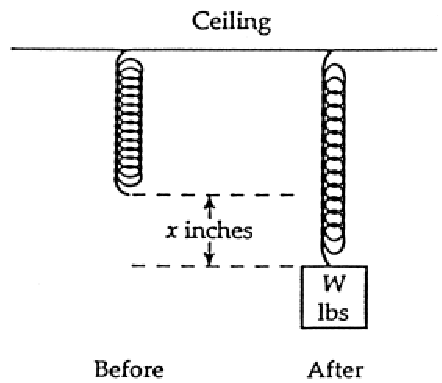
AVERAGE DAILY ENERGY REQUIREMENT

Subject	kJ/day
Infant, 9–12 months	4200
Child, 8 years	8770
Boy, 15 years	12 560
Girl, 15 years	9560
Woman (inactive)	7950
Man (inactive)	10 460

ENERGY EXPENDED DURING CERTAIN ACTIVITIES (in kJ per hour)



19 When an object that weighs W pounds is hung from a spring, the spring stretches x inches, as shown in the picture below.



Use the equation below to determine how many inches the spring will stretch if an object weighing 5 pounds is attached to the end of the spring.

$$W = \frac{2}{3}x$$

These two items are reprinted from FCAT Grade 10 Mathematics, Sunshine State Standards, Test book (released Aug 2005), copyright 2005 by Florida Department of Education. Permission granted.

throughout the preserve (*idealization*). The preserve is modeled by a simple closed curve, and some shaded area is given to represent the region that has been counted (*mathematization*). All of the above steps are not performed by the student but has to be understood in order to proceed. It is worth noting that the idealization step is rarely seen in FCAT04, only in #25 as discussed above, and also in #34 (see Figure 5.11, p. 139). All other steps ([1] posing a real-world question, [6] interpretation of the result, [7] verify or criticize the model, and [8] reiteration of the process) are neither present in the task description nor as a requirement for the student.

To sum up briefly, the FCAT04 mainly focuses on testing students' understanding of some

Table 5.16: Modeling spectra of modeling problems in FCAT04

Task Name: Amusement Park Construction
Context: Civil Engineering
Mathematical knowledge: Planar Geometry

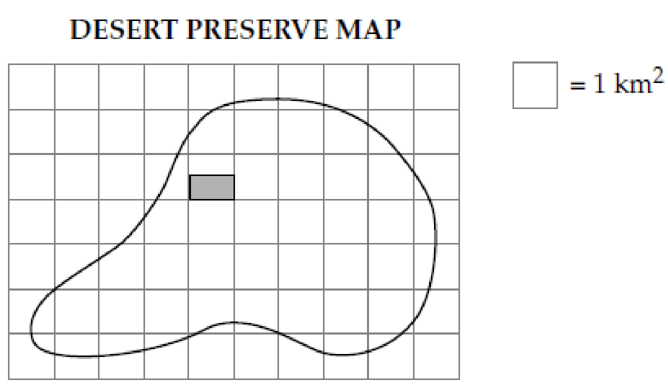
Problem #	4, 10, 37, 39	7	23	25	34	49
[2] model conceptually	110	110	110	110	110	100
[3] idealize	000	000	000	110	110	000
[4] mathematize	001	111	111	110	001	111
[5] solve mathematically	001	001	000	001	001	000

For a definition of modeling spectrum, see page 107. The "1" and "0" in the entries are symbols referring to "Yes" and "No", respectively. For example, "110" in the row "model conceptually" means that the item includes a description of how a real situation is conceptually modeled, and the student needs to understand such conceptual modeling in order to solve the problem, but the student does not need to perform this modeling step.

Figure 5.10: FCAT 2004 item, #25, Desert Preserve.

For the following problem, you will be required to use ESTIMATION strategies.

Luisa, a ranger at a desert preserve, is estimating the number of snakes in the preserve. She counted 25 snakes in a 0.5-square kilometer area as represented by the shaded area on the grid below. The total area of the preserve is also outlined on the grid.

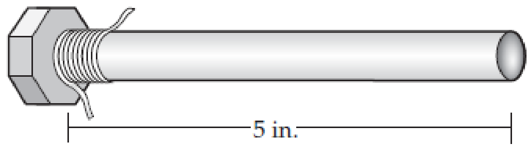


If Luisa assumes the snakes are evenly distributed throughout the preserve, how many snakes should she ESTIMATE are in the entire preserve?

Show your work or explain in words how to make an estimate.

Figure 5.11: FCAT 2004 item, #34, Iron Bolt.

Jenny is making an electromagnet by wrapping wire around an iron bolt, as shown in the picture below.



The wire is 0.1 inch in diameter so each wrap is 0.1 inch wide. If Jenny makes one wrap every 3 seconds, how many **seconds** will it take to wrap the 5-inch bolt?

This item is reprinted from FCAT Grade 10 Mathematics, Sunshine State Standards, Test book (released Aug 2005), copyright 2005 by Florida Department of Education. Permission granted.

modeling steps, which are usually necessary for solving the presented task, and it is a minor concern to have students performing the mathematization steps themselves. Students never have to comprehend earlier or later steps of modeling. The following analysis of other education systems' documents emphasizes more on how the "earlier or later steps of modeling" are assessed.

In FCAT04, many questions echoed the requirement of multiculturalism by introducing names and places from different traditions. *Real-world* context, as Florida's curriculum framework emphasized, is present almost everywhere throughout the test paper, but the connection between the context and the mathematics is not strong. In other words, it may be seen that the mathematics problems can be posed using some real-world context, but the solution often has little to do with such context.

Through the analysis, we have also presented a tool that may expose a *modeling spectrum* of an application task. An application item is assumed to be related to one or more modeling steps in one of the following ways: (i) the task statement describes a certain step; (ii) the student has to understand the description of a step; (iii) the student has to perform the modeling step. One may classify application problems using this tool in order to see what kind of modeling competence they could assess.

Florida, 2004-2012.

Florida's *Next Generation Sunshine State Standards* (NGSSS) is a response to standard-based reform movement and the demand for a higher educational standard. In particular, the NGSSS (Mathematics Standards) is designed to be narrower, deeper, more concise than the *1996 Sunshine State Standards* (FLDOE, 2013) [8]. The NGSSS was published in 2007 to replace the older curriculum, but was quickly altered in 2010 due to the adoption of Common Core State Standards, and was totally replaced in 2014 by the current *Mathematics Florida Standards*. Although the life span of NGSSS was quite short, it is the focus of our analysis, since it covered the most part of the period 2004-2012.

The public examination under NGSSS was under a major change. The older examination, FCAT, was operated up to grade 10, but the newer FCAT 2.0 was up to grade 8 only. In addition, there are individual examinations for Algebra 1 and Geometry, called the End-Of-Course (EOC) assessments. The FCAT 2.0 and EOC Algebra 1 assessments commenced in 2011, and EOC Geometry in 2012 (FLDOE, 2011; 2012; 2013). In the following, we briefly review the NGSSS mathematics standards, and some sample items of FCAT 2.0 (FCAT10), EOC Algebra 1 (EOCA1), and EOC Geometry (EOCG). Although Common Core was adopted in 2010, these public assessments are still based on the 2007 standards.

Comparing to the 1996 standards, the NGSSS still mentioned *real-world* throughout the document, along with the emphasis of *context*, and also started to mention *modeling*. Moreover, there are more specified real-world applications that are required in the standards, including financial

[8] FLDOE (2013), Florida's Adoption of State Standards: Background, in *Governor Rick Scott's Education Accountability Summit - Collaborative Engagement, August 26-28, 2013*

literacy concepts, such as simple and compound interest, net present value and net future value of income streams, and ideas in economics, such as the relationship between price and quantity supplied. Requirements of data analysis and probability are also largely expanded. Therefore we expect to see even more real-world applications in the public examination items.

Since the NGSSS used benchmarks for the description of the standards, it does become much more concise without the specification of how the standards are developed step by step. As FLDOE (2013) presented, it is the *what* but not the *how* that is the focus of the standards documents.

In the three public examinations (FCAT 2.0, EOC Algebra 1, and EOC Geometry), the item formats are similar to FCAT (2004), but there are some important changes in the item specification: each item has an *Item Context*, which has to refer to the real-world situation of the item. It is also required that if specific data or information is included, they have to be accurate and documented (p. 32). Possible contexts of FCAT 2.0 items include *Mathematics*, *Health/Physical Education*, *Social Studies/Consumerism*, *Science*, and *The Arts*. For EOC Algebra 1 and Geometry, the contexts are *Science, Technology, Engineering, and Mathematics*, , However, these labels are not very helpful, as can be seen in one task that is labeled as Science context: "Last year, Nicholas paid $7.25 for the price of a movie ticket at his favorite theater. This year, the price of a movie ticket at the same theater is $9.50. Which is closest to the percent of increase in the price of a movie ticket?''. The context is not from a scientific, but a daily activity.

As was done for the FCAT04, we can also check the modeling spectrum of the FCAT 2.0, EOC Algebra 1, and EOC Geometry items. Two items of each test are presented below, highlighting the modeling steps *other than* [5] to solve mathematically. We also discuss how these items can be connected (or *disconnected*) with the ability to apply mathematical knowledge.

Sample item 61 (*Supply and Demand*) of FCAT 2.0 is one that uses economics concepts mentioned in the curriculum standards. However, student can simply ignore the meaning of the supply

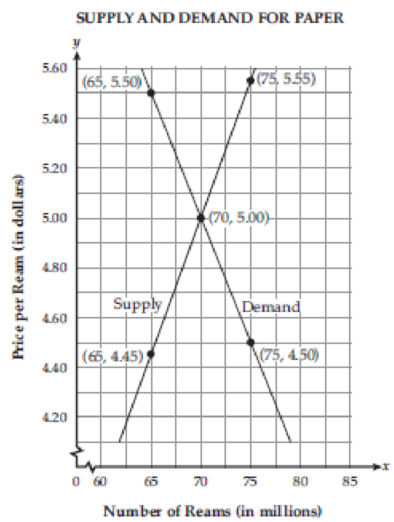
Figure 5.12: FCAT 2.0 item, #61, Supply and Demand.

An economist is helping a paper company evaluate the demand for reams of paper at different selling prices. The point at which the supply and demand graphs intersect is referred to as *market equilibrium*.

The economist graphed the supply and demand equations shown below.

Demand equation: $y = {}^{-}0.1x + 12$

Supply equation: $y = 0.11x - 2.7$



What is the price per ream, **in dollars**, of the *market equilibrium*?

Figure 5.13: FCAT 2.0 item, #78, Golf Rounds.

Austin saved \$455 from his pay and joined a golf club to improve his game. He paid a \$100 membership fee and will pay \$15 for each round of golf he plays.

Austin used the following inequality to determine the number of rounds of golf he could play.

$$100 + 15r \leq 455$$

What is the maximum number of rounds of golf Austin can play?

and demand, and finish the task by solving for the linear system in two variables. It can be reasonably understood that the item focuses more on required mathematical skills rather than the understanding of economics. In Sample Item 78, a golf playing situation is presented, and the mathematical is also given (as an inequality), requiring that the student understands and solves the inequality in positive integer. Since the real-number solution to the inequality is $r \leq 23.67$, a solution that ignores the context may result in an answer 23.67 or 24. These two items provide some interesting contrast, that an economics model can more sophisticated than the "golf playing inequality", but it depends on how the task is presented in order to see what competence is needed. Indeed, the item 78 needs somewhat more modeling than item 61. The modeling spectra of these two items, along with the four that are presented below, can be found in Table 5.17).

For the End-of-Course assessments of Algebra 1 (EOCA) and Geometry (EOCG), we found two interesting items with modeling steps that were not discussed before. Sample Item 35 of EOCA (Figure 5.15, p. 147) does not ask student to calculate mathematically, but presented a graph about the water balloon's height and ask student for an option that interpreted the graph correctly. This is known as an *interpretation* step because the mathematical results (as the equation, its graph, and some labeled coordinates) are already presented, so that student can use any of these to check whether the statements in real-world terms are right or not. Sample Item 27 of EOCG (Figure 5.17, *Plastic Wrap*, p. 149) poses an interesting and authentic question: "How much wrap is needed to wrap the irregularly shaped container?" It is a good sign that the first modeling step is involved. Indeed, the problem situation provided much opportunities for modeling: to wrap an object with "irregular'' shape, what is the area of plastic wrap that is sufficient? If the plastic wrap is only available in one-foot-wide rolls, what is the *length* that is necessary? What method of wrapping minimizes the consumption of plastic wrap rolls? If the object is to be boxed, what is the smallest surface area of the box? Nonetheless, we also see that the question can be ignored to solve the problem.

Actually, we see quite a number of item that allows students to solve without taking care of the context. Both *Supply and Demand* and *Plastic Wrap* are two of these items, but there are even more. Figure 5.16 (p. 148) shows an item that described a boat that is to change its planned route and needs to find the correct turning angle, but since a right triangle, even with the right angle sign drawn, is given, so the long descriptions are not useful in the solution. From the perspective of solving problems in a timed test, skipping the description totally and focusing solely on the mathematical symbols may be the best strategy. This point is even strengthened by the official description that "extraneous information may be included in items" (see, e.g., FCAT10, p. 32). Now the context *per se* can be treated as "extraneous information".

Creating tasks that require students to use the contextual information is challenging. For example, in the *Plastic Wrap* item, the task writere probably noticed that the area of the wrap needed should be more than the surface area, and it can be difficult to figure out what the total area of plastic wrap is needed. To guarantee that an answer is consistently calculated, it is the least risky to ask what the exterior surface area is. However, there are still simple chances to test modeling competence. For example, "what are the options that must not be enough for the area of plastic wrap that is necessary to cover the whole container?" (Answer: A and B).

Although it may not be the intent of the test author to test students' ability to apply mathematics, it may be hypothesized that doing a large number of such items can discourage students' consideration of real-world constraints in solving mathematics problems.

The last example for Florida is about health insurance (EOCA #11, Figure 5.14, p. 146). The item ambitiously listed definitions of insurance terms, which can be seen as *conceptual modeling*. A student really has to understand how these terms work in order to understand the next step when the total annual health care *costs* is modeled by a mathematical expression, assuming that the

Table 5.17: Modeling spectra of sample problems in Florida's 2012 tests.

Test	FCAT 2.0		EOCA		EOCG	
Problem #	61	78	11	35	32	27
[1] pose r.w. question	000	000	000	000	000	100
[2] model conceptually	100	110	110	000	100	100
[3] idealize	000	000	000	000	000	100
[4] mathematize	100	110	111	110	100	100
[5] solve mathematically	001	001	000	110	001	001
[6] interpret	000	000	000	111	000	000

total medical *expenses* is not less than $500. Conceptually, this problem is relatively sophisticated comparing to other tasks. It is also a realization of what the curriculum documents mentioned. Some may argue that student can be advantaged if they already knew the meaning what premium, deductible, and copayment, but if the purpose is to test whether students understand the context, this problem is appropriate. All we need is more such problems in various contexts (which students need to understand in order to solve).

Analysis of Massachusetts' curriculum materials

Massachusetts, 1995-2003.

The *Massachusetts Comprehensive Assessment* (MCAS), a statewide assessment system, was introduced in Massachusetts in 1993 (Anthony & Rossman, 1994). Students have to take the test and exceed a threshold score in order to fulfill part the graduation requirements. The 2004 released test (MCAS04) can be seen as an assessment of students' accumulative achievement due to their schooling experiences within in the 1995-2003 period. The MCAS04 was based on the 2000 Massachusetts' mathematics curriculum framework (MACF), which is to be reviewed below, following by an analysis of the test items.

Figure 5.14: EOC Algebra 1 item, #11, Health Insurance

The out-of-pocket costs to an employee for health insurance and medical expenses for one year are shown in the chart below.

EMPLOYEE'S ANNUAL HEALTH CARE COSTS

Type of Cost	Definition	Cost to Employee
Premium	Total amount employee pays insurance company for the policy	\$3,626
Deductible	Amount of medical expenses employee pays before insurance company pays for anything	\$ 500
Copayment	Percentage of medical expenses employee has to pay after the first \$500	20%

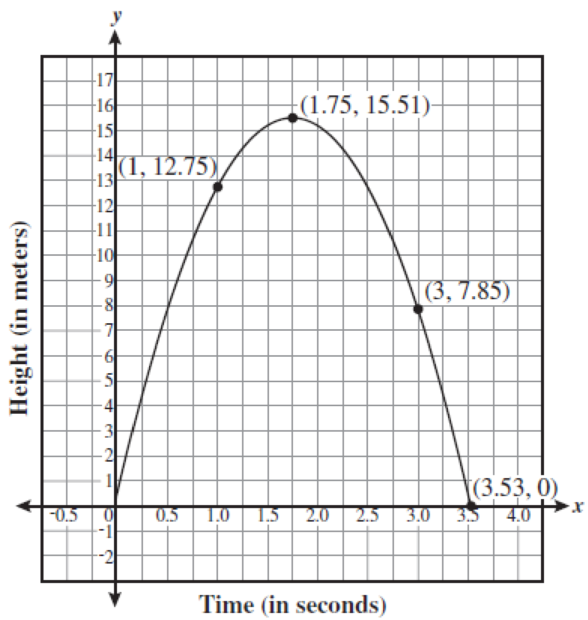
According to the plan outlined in the chart, total annual health care costs, C, depend on the employee's medical expenses for that year. If x represents the total medical expenses of an employee on this plan and $x \geq 500$, which of the following equations can be used to determine this employee's total health care costs for that year?

A. $C = 3626 - 500 + 0.20(x - 500)$
B. $C = 3626 - 500 + 0.20x$
★ **C.** $C = 3626 + 500 + 0.20(x - 500)$
D. $C = 3626 + 500 + 0.20x$

As was in Florida, MACF was influenced by the NCTM standards. The curriculum makers also consulted reports of the MAA, AMS, and many other professional associations and research groups. In the introductory sections of MACF 2000 version (MACF00), it emphasized *problem solving*, *communicating*, *reasoning and proof*, *making connections*, and *representations*, and some educating principles, but it did not have statements of why mathematics is to be learned as a required subject. Nevertheless, some excerpts help understanding of how mathematics was viewed. The excerpts below express that mathematics is a fundamental subject and is necessary for many

Figure 5.15: EOC Algebra 1 item, #61, Water Balloon:First item with interpretation of mathematical results

Timmy and Kelli had a water balloon launcher. When launched, the water balloon's height could be modeled by the quadratic equation $y = -4.9x^2 + 17.15x + 0.5$. The graph shown below represents the water balloon's height.



Which of the following is true about the water balloon?

A. The water balloon reaches a height of 16 meters.
★ **B.** The water balloon reaches the height of 7.85 meters twice.
C. The water balloon has a maximum height of 17.15 meters.
D. The water balloon travels for 4.9 seconds before it hits the ground.

This item is reprinted from Algebra 1 End-of-Course Assessment Test Item Specifications Version 2, copyright 2012 by Florida Department of Education. Permission granted.

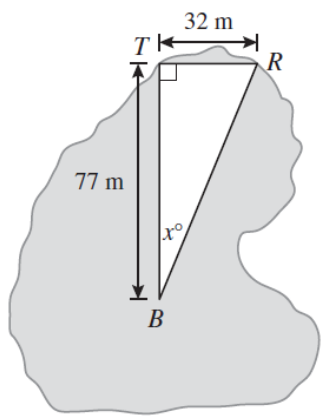
academic branches. This can be taken as *Academic Rationalism.*

MACF00. Ideology: Academic Rationalism.

```
    Because mathematics is the cornerstone of many disciplines, a comprehensive
curriculum should include applications to everyday life and modeling activities
that demonstrate the connections among disciplines. Schools should also provide
opportunities for communicating with experts in applied fields to enhance students'
knowledge of these connections. (p. 9)
    The geometry in Euclid's Elements of Geometry was a logical system based on
ten assumptions. Five of the assumptions were called common notions (axioms, or
self-evident truths), and the other five were postulates (required conditions). The
resulting logical system was taken as a model for deductive reasoning and profoundly
```

Figure 5.16: EOC Geometry item, #32, Boat Direction

A tackle shop and restaurant are located on the shore of a lake and are 32 meters (m) apart. A boat on the lake heading toward the tackle shop is a distance of 77 meters from the tackle shop. This situation is shown in the diagram below, where point T represents the location of the tackle shop, point R represents the location of the restaurant, and point B represents the location of the boat.



The driver of the boat wants to change direction to sail toward the restaurant. Which of the following is closest to the value of x?

★ **A.** 23
B. 25
C. 65
D. 67

This item is reprinted from Geometry End-of-Course Assessment Test Item Specifications Version 2, copyright 2012 by Florida Department of Education. Permission granted.

```
influenced all branches of knowledge. Indeed, the development of the axiomatic approach
to geometry extends to the present. (p. 14)
```

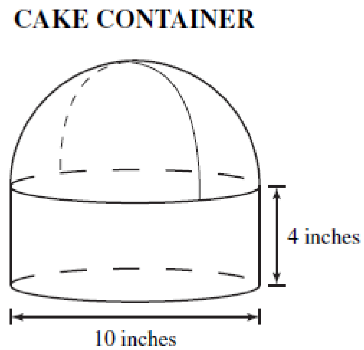
Another excerpt shows that the learning of mathematics is important for success in college and the workplace, and can be taken as the ideology of *Social and Economic Efficiency*. Other ideological views were not found in MACF00.

MACF00. Ideology: Social and Economic Efficiency.

```
   At every level of the education system, teachers should act on the belief that
every child should learn challenging mathematics. Teachers and guidance personnel
should advise students and parents about why it is important to take advanced courses
```

Figure 5.17: EOC Geometry item, #27, Plastic Wrap

Abraham works at the Delicious Cake Factory. He packages cakes in cardboard containers shaped like right circular cylinders with hemispheres on top, as shown in the diagram below.



Abraham wants to wrap one cake container completely in colored plastic wrap and needs to know how much wrap he will need. What is the total exterior surface area of the container?

> in mathematics and how this will prepare students for success in college and the workplace. (p. 9)

The role of modeling and application was secondary, since there was not a section that made an emphasis of teaching that. Application was only listed as a *feature* (p. 118) of instructional materials, but not a part of the learning standards. The word *modeling* was rarely mentioned, but it was indicated three times that students should be able to use *geometrical modeling* to solve problems. The modeling spectra of some sample items from MCAS 2004 assessment (Spring 2004, Grade 10) are summarized in Table 5.18, and the analysis is presented below.

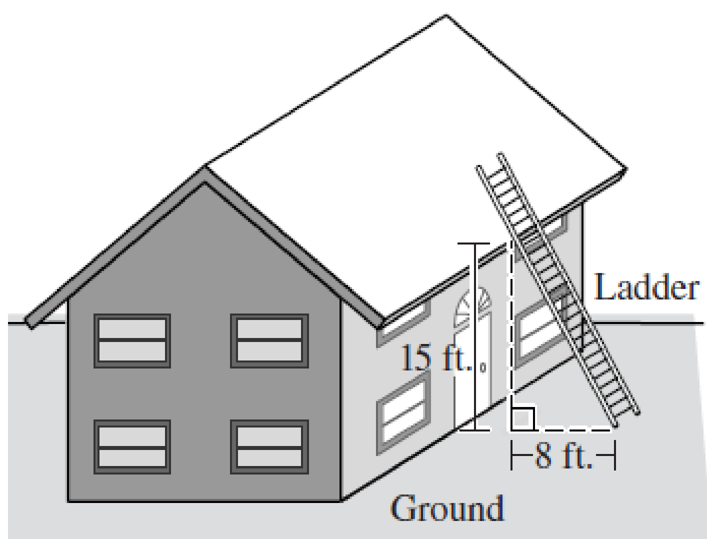
Out of the 42 assessment items in MCAS04, only 18 (or 43%) of them are written in a real-world context. As was done in the analysis of Florida's items, we simply call them *application tasks*. The proportion of application tasks is much less than Florida's. None of them are *context-detachable* pure mathematics tasks or *camouflage word problems*, meaning that there is always some conceptual model or mathematical model to be understood or worked on by the student.

Table 5.18: Modeling spectra of sample problems in MCAS 2004, Grade 10 Mathematics.

Problem #	33	37	5	42
[1] pose r.w. question	000	100	000	110
[2] model conceptually	110	110	111	110
[4] mathematize	110	000	110	001
[5] solve mathematically	001	001	001	001
[6] interpret	000	000	000	111

Figure 5.18: MCAS04 item, #33, Ladder Section.

33 Using the measures shown in the sketch, what is the length of the section of the ladder from the point where it rests on the ground to the point where it touches the house?



Applied mathematical problem is still a typical type of application tasks (10 out of 18), where students have to undestand the context and/or the mathematical model, but only need to answer mathematical question within the model. An example is Item 33 (Figure 5.18, p. 150). Student has to understand both conceptually – "the section of the ladder from the point where it rests on the ground to the point where it touches the house", and to understand the mathematical model – that section of the ladder is modeled as the hypotenuse of the right triangle given in the figure. Note that this task is slightly different from Florida's *Boat Direction* task (Figure 5.16, p.148), since the latter can be done by solely focusing on the x° marked in the figure, whereas the former task at least requires the understanding of the "ladder section". However, the practical effect may be similar, since the description about the ladder section is not really important to solve the problem – a test-smart student could simply *assume* that the unknown quantity is the hypotenuse, and quickly solve the problem. This is a hypothesis to be tested, but is theoretically supported by Verschaffel et al. (2000).

Another task that can be directly compared with Florida's tasks is #37: "A right circular cylindrical can is 6 inches high, and the area of its top is 36π square inches. What is the minimum number of square inches of construction paper it would take to cover the lateral surface of this can?" This task is about a similar situation as in the *Plastic Wrap* task (Figure 5.17, p. 149), though it does not ask for the *surface area*. It does not generate a problem because the lateral surface of a cylinder can be "perfectly wrapped" by a piece of paper[9].

In MCAS04, there are three questions where conceptual modeling is needed to solve the problem, and it was not seen in Florida's items. Two of them are questions about understanding the bar chart and the box-and-whisker chart, which are treated as conceptual models to represent statistical concepts. The third one is #5, *Bicycle Movement*: "The wheels on Bill's bicycle each have a radius of 35 centimeters. Which of the following is closest to the distance the bicycle moves along

[9]Technically, the lateral surface of a cylinder is a *developable* surface, which can be transformed to a plane without distortion, but a hemisphere can not.

the ground in one complete revolution of the wheels?" One has to understand how the bicycle wheels roll along the ground (without sliding) in order to understand that the bicycles' travel distance is equal to the circumference of a wheel. There is no picture given, so the student has to go through the process of conceptual modeling.

Figure 5.19: MCAS04 item, #42, Ski Trip.

A local ski club plans to charter transportation for a ski trip. Two different bus companies are available for charter services.

SNOWBIRD CHARTER	MOUNTAIN CHARTER
Roundtrips Depart Daily 6 and 8 A.M.	*Roundtrips Daily at 6 and 8 A.M.*
\$300, plus \$12 per person	*\$15 per person*
Reservations are required.	**Call for reservations.**

a. If 72 club members sign up for the trip, what would be the total transportation cost for each of the two charter companies? Show your work and label each answer with the company name.

b. Write an equation that expresses c, the total cost of using Snowbird Charter, in terms of p, the total number of club members who go on the trip.

c. Write an equation that expresses c, the total cost for using Mountain Charter, in terms of p, the total number of club members who go on the trip.

d. If the club members want to choose the less expensive of the two bus companies, which company should they choose? Justify your answer by explaining how the number of club members who go on the trip should affect their decision.

The last example is #42 (Figure 5.19, p. 152), which is also the last problem in MCAS04, involves five modeling steps. It poses a real-world relevant question: there are two plans offered by bus companies, and a club of ski members is going to choose the less expensive one. Which one should they choose? Then, conceptually, student has to understand the contents in the two offers. After listing some equations, they have to solve, and finally interpret their mathematical results by explaining how the number of club members who go on the trip should affect their decision. Note that the actual number of club members who go on the trip is only given in part (a), but becomes a variable in (b) through (d). It is a simple but realistic modeling problem that student may take in a timed public examination.

Massachusetts, 2004-2012.

The 2011 version of Massachusetts Curriculum Framework for Mathematics (MACF11) continued to draw on NCTM standards, and proficiency urged in the National Research Council's reports. More importantly, the new framework was based on the *Common Core State Standards* (CCSS), which means that the *8 Standards for Mathematical Practice* from CCSS were adopted, including the 4th standard, "Model with Mathematics"

> Mathematically proficient students can apply the mathematics they know to solve problems arising in everyday life, society, and the workplace. In early grades, this might be as simple as writing an addition equation to describe a situation. In middle grades, a student might apply proportional reasoning to plan a school event or analyze a problem in the community. By high school, a student might use geometry to solve a design problem or use a function to describe how one quantity of interest depends on another. Mathematically proficient students who can apply what they know are comfortable making assumptions and approximations to simplify a complicated situation, realizing that these may need revision later. They are able to identify important quantities in a practical situation and map their relationships using such tools as diagrams, two-way tables, graphs, flowcharts and formulas. They can analyze those relationships mathematically to draw conclusions. They routinely interpret their mathematical results in the context of the situation and reflect on whether the results make sense, possibly improving the model if it has not served its purpose. (MACF11, p. 16. See also CCSSI (2010)).

This also adds on to the curriculum ideology of Social and Economic Efficiency, since

"[s]tudents learn to model with mathematics and to apply the mathematics that they know to solve problems arising in everyday life, society, and the workplace. (See Standard for Mathematical Practice 4: Model with mathematics.)" (p. 10)

Overall, the curriculum ideologies did not change in the 2011 version of MACF.

Another standard of CCSS, "make sense of problems and persevere in solving them" is also important for modeling from the other direction, since it expects that mathematically proficient students should make sense of the mathematics that they do, which may involve using concepts from real-world situations.

Consequently, the use of words *context*, *modeling*, and *application* become more in MACF11.

As expected from *Academic Rationalism*, modeling is mostly seen in the description of individual topics in mathematics:

> Students build on this idea to develop the notion of radian measure for angles and extend the domain of the trigonometric functions to all real numbers. They apply this knowledge to model simple periodic phenomena. (p. 147)
> [Students r]epresent constraints by equations or inequalities, and by systems of equations and/or inequalities, and interpret solutions as viable or non-viable options in a modeling context. For example, represent inequalities describing nutritional and cost constraints on combinations of different foods. (p. 151)

Figure 5.20: MCAS12 item, #34, Driveway Area.

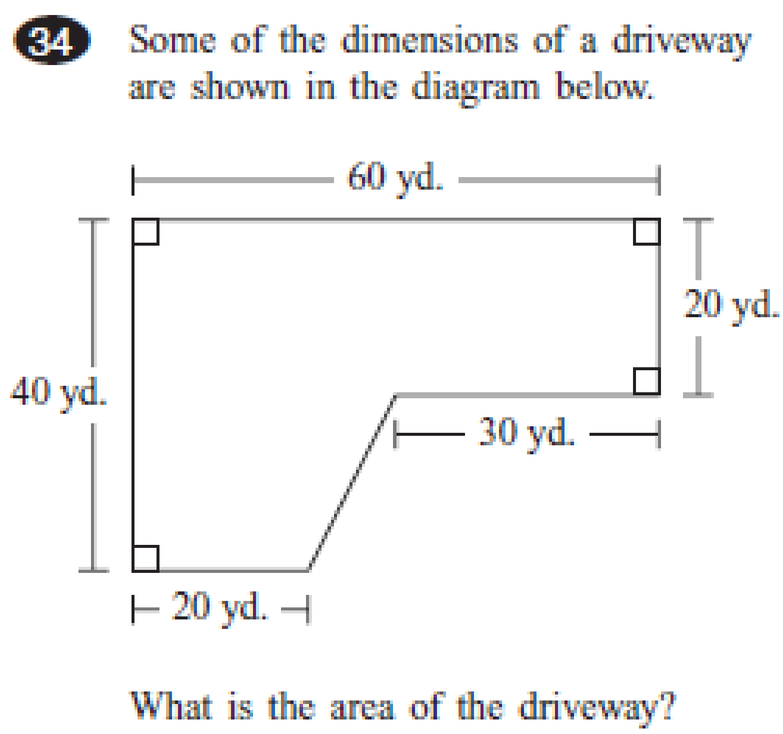
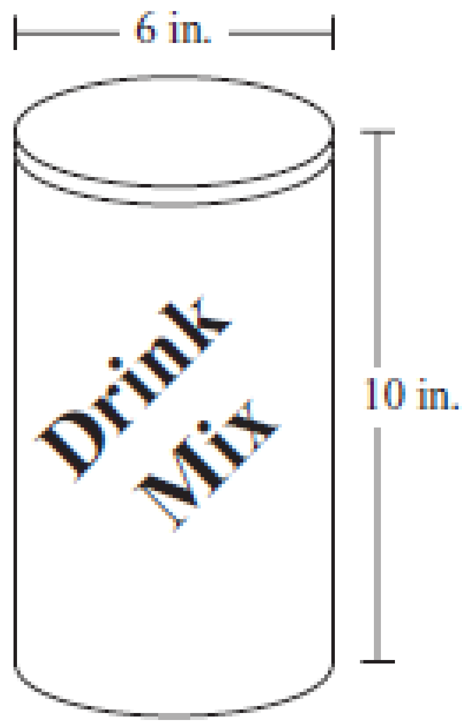


Figure 5.21: MCAS12 item, #22, Cylindrical Container.

A powdered drink mix container is in the shape of a right circular cylinder. The dimensions of the container are shown below.

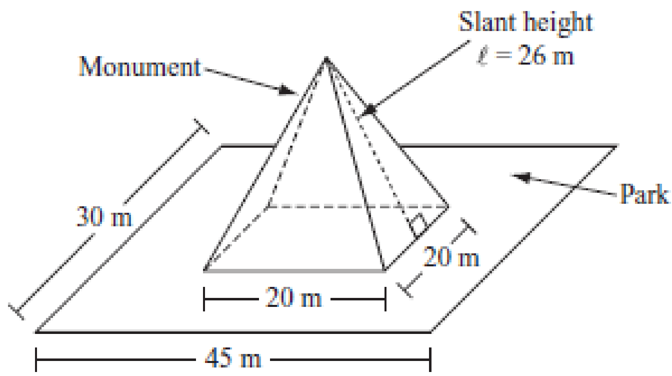


Which of the following is closest to the volume of the container?

This item is reprinted from *Massachusetts Comprehensive Assessment System, 2012 released mathematics items, grade 10*, copyright by Massachusetts Department of Elementary and Secondary Education (Department), 2012. Permission granted.

Figure 5.22: MCAS12 item, #42, Monument Park.

A monument in the shape of a right square pyramid is located in a park. The park is in the shape of a rectangle. The measurements of the monument and the park are shown in the diagram below.



a. What is the area, in square meters, of the base of the monument? Show or explain how you got your answer.

b. What is the area, in square meters, of the park, **not** including the base of the monument? Show or explain how you got your answer.

c. What is the lateral surface area, in square meters, of the monument? Show or explain how you got your answer.

The height of the monument is 24 meters.

d. What is the volume, in cubic meters, of the monument? Show or explain how you got your answer.

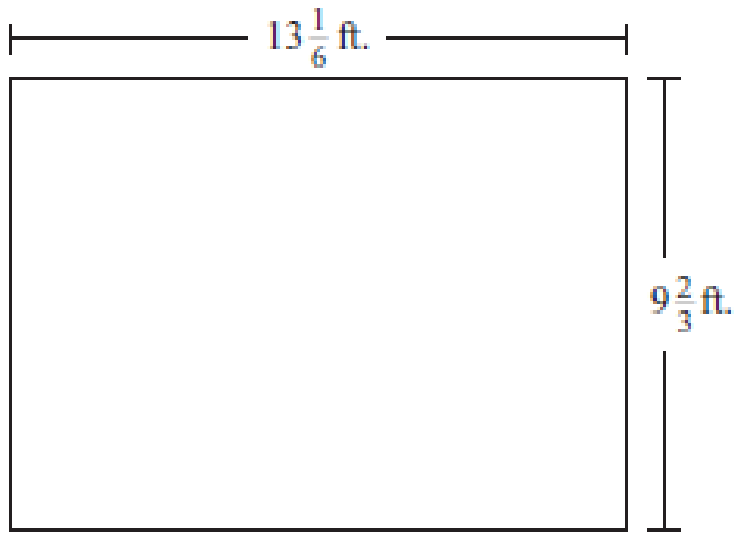
This item is reprinted from *Massachusetts Comprehensive Assessment System, 2012 released mathematics items, grade 10*, copyright by Massachusetts Department of Elementary and Secondary Education (Department), 2012. Permission granted.

In the following, we take a look at the public examination, MCAS 2012 (Grade 10 Mathematics), which is based on the 2011 Curriculum Framework. The format of the test is similar to MCAS04 reviewed before. Out of the 42 items in MCAS12, 14 of them (33%) are application tasks. It is a lower proportion than the 43% in MCAS04. In a strict sense, no *context-detachable* tasks were found. Classical tasks continue to be observed, but some of them only barely needs mathematization, such as such as #34 (*Driveway Area*), #22 (*Cylindrical Container*) and #42 (*Monument Park*); see figs. 5.20 to 5.22, pages 154 to 156. In these three problems, students do have to understand that the driveway is modeled as a polygon, the container as a cylinder, the monument as a square pyramid, and the park as a rectangle. However, since the shapes are already drawn with clear markings of dimensions and angles, the context can actually be abstracted away. For example, in item #34, the first sentence can simply be ignored, and "what is the area of the driveway?" can be seen as "what is the area?" If so, the context is really detachable and the task is still solvable. As discussed in Florida's items, the context can be treated as *extraneous information* since a complete mathematical task is already presented *among* the context. In these tasks, conceptual modeling and/or mathematization were present, and they were supposed to be related to the task and expected to be understood, but they might well be neglected without affecting the solution. In the modeling spectrum, an asterisk mark is attached to indicate that the context can be *extraneous*.

As was in MCAS04, there are tasks in MCAS12 that require some conceptual modeling, and all of them are about graphical representation in statistics, such as the *histogram* (#20) and the *stem-and-leaf plot* (#31) (figures of these two problems are not shown here). Finally, there is a task that would be a simple application but still posed interesting real-world question. In Figure 5.23, the item describes that "each tile covers one square foot and costs $2.49", and asked what is closest to the total cost needed. The question can simply be solved by calculating the rectangular area in square feet and multiply it by $2.49, but the problem's original intent may be an estimation: if one rounds the dimensions to get 13 and 10, the area would be approximately 130 square feet, and therefore the total cost is about 130 times 2.5 = 325 dollars, closest to the choice (D). However, there can be different understanding of the description "each tile covers one square foot". Does it

Figure 5.23: MCAS12 item, #4, Floor Tiling.

4 A contractor will use tiles to cover the floor represented by the rectangle below.



Each tile covers one square foot and costs $2.49.

Which of the following is closest to the total cost of the tiles needed to cover the entire floor?

A. $100

B. $234

C. $250

D. $320

Table 5.19: Modeling spectra of sample problems in MCAS 2012, Grade 10 Mathematics.

Problem #	34	22	42	20	31	4	12
[1] pose r.w. question	000	000	000	000	000	110	110
[2] model conceptually	000	000	100	111	111	000	110
[4] mathematize	110*	110*	110*	000	000	000	110
[5] solve mathematically	001	001	001	001	001	000	110
[6] interpret	000	000	000	000	000	000	000

* indicates that the context can be extraneous

mean that the tile is in the shape of a unit square? or the tile can be in other shape as long as the *area* of it is one square foot? Is it allow to use part of the tile? If so, the cost can be higher, though it would not change the answer (D) since it is the largest choice already.

Table 5.19 shows the modeling spectra of the discussed tasks. To sum up briefly, Massachusetts' tests have less on application items, but there are a few real-world relevant questions that students need to solve with understanding of the context. Idealization step is not tested, and it is also rare to see tasks that require interpretation of the mathematical results. Finally, the more recent FCAT apparently tested *less* on applications; when there is an application task, it is also more often to see that the context is weakly relevant to the solution of the task. It is contradictory to the fact that modeling was emphasized more in the 2011 curriculum documents, and requires further investigation. It will be beneficial to see the how the development of the Next-Generation MCAS (to be available in 2019) involves more in modeling and application.

Analysis of German curriculum materials (Bavaria)

Bavarian Haupschule and Gymnasium, Germany, 1995-2003

In the federal state of Bavaria (or *Bayern*), education affairs are administered by the Bavarian Ministry of Education and Cultural Affairs (KWMBl). Like other federal states of Germany, Bavaria has multiple paths of schooling. Starting from a four-year *Grundschule*, pupils may enter a *Mittelschule*, which is more professionally oriented, a *Realschule*, or a *Gymnasium*, which is the most academically oriented. The *Mittelschule* is a result of an administrative movement since the 2011/2012 school year, that if a *Hauptschule* or a conjunction of more than one schools fulfill certain criteria, then it can be named *Mittelschule*. We focus only on *Hauptschule* and *Gymnasium*.

Separate final examinations apply to the three types of schools. *Hauptschule* students generally take the *Qualifizierenden Abschluss* ("Quali") upon graduation, and the *Abiturprüfung* ("Abitur") is generally for *Gymnasium* students, and also serves as a qualification for university entrance. In the following, we review the curriculum documents (*Lehrplan*) for the Bavarian *Hauptschule* and *Gymnasium* in the period 1995-2003.

According to the nature of the school types, it is reasonable to expect that a more professionally oriented curriculum would take on a *Social and Economic Efficiency* ideology, which is the case for *Hauptschule* and can be seen in the following excerpt:

> BY97Hau (Lehrplan für die Hauptschule, Bekanntmachung des Bayerischen Staatsministeriums für Unterricht, Kultus, Wissenschaft und Kunst vom 29. Oktober 1997).
>
> Curriculum Ideology: Social and Economic Efficiency.
>
> `Der Mathematikunterricht stellt sich die Aufgabe, einen bedeutsamen Beitrag zur Allgemeinbildung der Hauptschüler zu leisten. Er soll sie befähigen, vor allem wirtschaftliche und technische Sachverhalte des Alltagslebens mit mathematischen Mitteln zu erfassen, zu durchdringen und aus ihnen erwachsende Fragestellungen und Probleme zu lösen. Der Unterricht schafft die Grundlage für die Bewältigung mathematischer Aufgaben in Arbeits- und Berufswelt sowie in weiteren Bildungsgängen.`
>
> *Translation*: `Mathematics teaching is the task of making a significant contribution to the general education of the primary pupils. It is intended to enable them to grasp, penetrate and solve problems and problems arising out of them by means of mathematical methods, especially economic and technical aspects of everyday life. Teaching provides the basis for the mastering of mathematical tasks in the world of work and professional life as well as in further education courses.`

The document also emphasized application, in the sense that students should be empowered to apply mathematical concepts and procedures to mathematize facts from the environment, profession and economy, but they should also recognize the limits of the mathematization of everyday phenomena (BY97Hau, p. 90). Following this principle, in the introduction of descriptive statistics and probability it also mentioned that students should learn how to use statistics to interpret vital situations, to look at the results critically, and to avoid misinterpretations (ibid, p. 375). Applications to common financial concepts are observed, such as *capital*, *compound interest*, and *depreciation.*

As for the curriculum of *Gymnasium*, we expected that since it is a university preparatory curriculum, it would take more on *Academic Rationalism* ideology. However, the curriculum documents of *Gymnasium* did not directly show a stance of what role mathematics plays in education. Instead, we have the following excerpt that describes a *Gymnasium*'s educational vision in general:

> BY90Gym (Lehrplan für das bayerische Gymnasium, Bayerisches Staatsministerium für Unterricht und Kultus, Juli 1990).
>
> Curriculum Ideology: Progressivism.
>
> Das Gymnasium ist deshalb eine Schule für Kinder und Jugendliche, die sich als in besonderem Maße geistig beweglich, lernbegierig und phantasievoll erweisen, die schnell, zielstrebig und differenziert lernen können, ein gutes Gedächtnis haben, sich gern selbständig, ausdauernd und von verschiedenen Seiten mit Denk- und Gestaltungsaufgaben beschäftigen und in allem die Bereitschaft erkennen lassen, die Anstrengungen auf sich zu nehmen, die der Bildungsweg des Gymnasiums ihnen abverlangt.
>
> *Translation*: The Gymnasium is therefore a school for children and adolescents, who show themselves to be especially mentally mobile, learning and imaginative, who can learn quickly, purposefully and differentiated, have a good memory, like to work independently, persevering and from different sides with thought And design tasks, and in particular the willingness to take the efforts which the education of the Gymnasium requires them.

In the mathematics curriculum of Gymnasium, the subject of mathematics is introduced as a fundamental set of knowledge that can be useful in many other disciplines, and therefore it is indispensable to master it. As discussed in Massachusetts' curriculum, the emphasis of a system of knowledge can also be taken as the *Academic Rationalism* ideology. The curriculum plan of Bavarian gymnasium systematically lists the required mathematical knowledge from Year 5 to Year 13 (roughly correspond to America's Grade 5 to "Grade 13"), pointing out the relevant topics in other disciplines. For example, when topic of prisms is presented in Year 8, it is remarked that the knowledge is related to *ray optics* in physics, and the study of *crystals* in chemistry; in Year 10, the learning of logarithmic scale has usage in the study of *radioactive decay* in physics, *pH value* in chemistry, *bacterial growth* in biology , the *Weber-Fechner law* of perception vs. stimulus, and *population explosion* in economics. Although modeling process is not mentioned, many of these applications require the understanding of mathematical models.

In the following, released *Quali* and *Abitur* examinations are reviewed. For the *Hauptschule*, the *Besondere Leistungsfeststellung zum Erwerb des qualifizierenden Hauptschulabschlusses, 2001, Mathematik* is chosen (Specific performance assessment for the acquisition of the qualifying Hauptschle leaving certificate, 2001, Mathematics. Code: BY01QH); for the *Gymnasium*, *Abiturprüfung 2001, Mathematik als Grundkursfach* (Abitur examination in 2001, mathematics. Code: BY01Abi).

The *Quali* is taken voluntarily by Year 9 students in *Hauptschule*, but it is expected that teachers and students are preparing for the examination together throughout the year (BY97Hau, p. 311). In the 2001 examination there are four parts. Students were given 100 minutes to work on two parts only, which were determined by the examination committee. There are four item in each part, thus 16 in total. Each item can have one up to five sub-questions, but we will only count each task as a whole. There are 10 application tasks, all of which require the understanding of the conceptual model. Therefore, under our classification, all of these 10 items are *classical* applied mathematical problems. Five items posed real-world relevant questions, such as #2.3:

> *Translation*: A businessman has not paid his bill over 11000 DM on time. Therefore, at an interest rate of 12%, he must pay 296 DM default interest.
> A) How many days have the payment deadline been exceeded?
> B) If the business man had paid within 30 days, he could have deducted 3% cash discount from the invoice amount.
> How much money would he have saved in comparison to the late payment?

and #3.3:

```
Translation: In the autumn, an organic farmer is storing 1.2 tons
of pears. He puts his own costs at 640 DM. During storage a weight
loss of 18% occurs. The farmer wants to sell the whole product with
a profit of 32%.
a) Calculate the weight of the pears sold.
b) Determine the selling price.
c) Determine the profit in DM.
d) Determine the final price per kg (the VAT is 8%).
Round to 2 places after the decimal point.
e) How much would his percentage profit have been if he sold the
fruit immediately after the harvest at 520 DM cost and 752 DM total
selling price?
Round to 2 places after the comma.
```

A competent student is assumed to understand concepts such as *interest rate*, *profit*, and *VAT* (*value-added tax*) in order to understand the task, and will have to mathematize them into mathematical expressions and equations. This is typical in classical application tasks. Moreover, the items presented some real-world issues like comparing the saved amount if the businessman paid on time, and comparing the profit with different sales schemes. With understanding of these real-world issues, students are supposed to solve these tasks more easily.

There are also some items that may work well as a test, but the real-world situation can be problematic. For example, item #1.3 is a typical word problem of *working together*:

```
Translation: The school-leaving certificate of a Hauptschule is to
be opened in eight weeks, where 6 pupils had to work 4 hours per
week. After two weeks, 2 of these students are on a class trip and
are leaving for a week.
How many minutes per week do all the participants have to work
after the end of the class trip so that the pupil can finish by the
scheduled date?
```

Of course, in this type of problems, we usually assume that (1) all the people work at a constant rate; (2) when more than one people are working together, their contribution to the work can be added directly. However, in reality, (1) students' working rate can be very different when the deadline is approaching; (2) the students are different in their characters and abilities. Another example is related to particle physics (#4.4):

Table 5.20: Modeling spectra of sample problems in Bavaria's mathematics *Quali* and *Abitur*, 2001.

Examination	*Quali*				*Abitur*	
Problem #	2.3	3.3	1.3	4.4	3	4
[1] pose r.w. question	110	110	110	000	110	110
[2] model conceptually	110	110	110	110	110	110
[3] idealize	000	000	000	000	000	111
[4] mathematize	001	001	001	001	111	111
[5] solve mathematically	001	001	001	001	001	001
[6] interpret	000	000	000	000	000	001

Translation: All known substances are composed of single atoms. The substances differ only in the different number of core particles.
The core is composed of electrically positive protons (mass about $1.673 \cdot 10^{-24}$ g and neutrons of approximately the same weight.
a) Calculate the mass of an electron. It weighs the 1836th part of a proton.
b) The core of a uranium atom consists of 92 protons and 146 neutrons.
Calculate the mass of the atomic nucleus.

From the problem, it is obviously assumed that the mass of the nucleus is the sum of the total mass of proton and neutrons, so the answer will be a higher than the true mass of uranium-238 by about less than 1%, due to the phenomenon of *mass defect*. One may argue that the item's modeling did not involve mass defect, so it would be reasonable to ask students to work within the model. In that case, the asssumption should either be explicitly stated by the task description, or be articulated by the student as part of the task, otherwise the result is not physically correct, noting that the given values have four significant figures. Table 5.20 presents the modeling spectra of the above-mentioned items, as well as of the Abitur 2001 items.

Abitur 2001. The *Abitur* is required for *Gymnasium* students, and is taken generally at the end of Year 12 or 13. In the 2001 *Abitur* for mathematics, there are three assessed fields (*1. Calculus, 2. Probability and Statistics, 3. Analytic Geometry*), and there are two items in each field. Each item has two up to five related sub-questions, but we will only count each task as a whole, so there are six tasks in total. Only two of these items are application tasks, and as one may correctly guesses, both item are from the second field, Probability and Statistics. The problem statements tend to be quite long, about one page per item, therefore the whole item is not cited here. Instead, we will summarize the real-world situation that an item is about, and then analyze the modeling steps involved.

Item #3 described a manufacturing group that produces microchips with a 15% defective rate. After some basic probability calculations, the item further described a testing device that is able to probabilistically detect defectiveness of a chip. For flawless chips, there is a 3% rate that the testing device report defect and therefore discarded, and it is given that 83% of all the microchips are not discarded after the detection process. The item asked student to infer the percentage of the defective chips that is discarded. Finally, the item mentioned that a bonus program was created in order to improve the quality: After the production, a sample of 200 chips is taken and sent to an expert group. The expert group is able to examine all 200 chips accurately, knowing which ones are really flawless or defective. If there are no more than 22 chips that are defective, then the production team receives a bonus. Two natural questions are: (1) What is the probability that the team will receive the bonus even though no quality improvement is occurred? (2) Even if there is actually an improvement, say, the defective rate has fallen to 10%, what is the probabilty of the team being denied the bonus?

We see that four modeling steps are involved: (1) A few real-world questions are posed; (2) the student needs to understand the situation, and (3) be able to model it using probablity (*Bernoulli trials*); (4) perform necessary calculations with aids such as calculators and distribution tables. There are hidden assumptions, such as independence, that are not addressed.

There are also extra questions that are not related to the microchip production. One of them is like the following: in the expert team there are 12 members from Germany, England, and France.

The numbers of female and male experts from each country are all given. In a restaurant, a four-table and an eight-table are reserved. How many options are there to occupy the tables sot that at each table, (a) there are equal numbers of men and women? (b) at least two German members should be present?

Such extra questions can actually be seen as a separate item, and is more like a problem of "pure combinatorics".

The item #4 began with the following description:

> A singing association has 30 female and 20 male members. The number of participants in the weekly choir rehearsal fluctuates from time to time. To describe this variation by a model, it is to be assumed that the members participate independently of each other and with a probability of 85% of presence in a rehearsal.

After that, a few basic probability calculations were required. Some are easier, such as "what is the probability of a choir rehearsal having at least 40 members present?" Some looks more complicated, like "How many choral rehearsals must at least be held, so that, with a probability of more than 99%, at least once all 5 tenors of the club are to be found present together?" After that, a hypothesis testing question follows like this:

> *Translation*: The chorus suspects that the attendance rate has fallen below 85%. In order to check this hypothesis, the zero hypothesis $H_0 : p \geq 0.85$ is to be tested at the significance level of 5%. For this purpose, the number of participants is determined for each of the four choral samples. The decision is to be made on the basis of the sum of these four figures. Determine the decision-making rule.

The item also mentioned that the Bernoulli model assumed may not be accurate, and asked the student to explain by a concrete example to show why. The answer keys showed that two kinds of answers are acceptable: (1) There may be couples who are mostly together, breaking the assumption of independence; (2) The attendance rate can be dependent on the age of the singer, violating the uniformity of the distribution.

This item is in a sense similar to the previous item, for they both involve the Bernoulli distribution. However, the defective rate of a microchip production is practically more consistent than the absent rate of a choir member, allowing the latter item to ask more about the assumption behind the model. Also, an interpretation is required when a student explains the *decision-making rule* in the hypothesis testing.

The German problems are very different in "style", since most of them are not multiple-choice questions, and many questions are grouped together in one context. This is, hypothetically, a way to ask more and deeper questions within the time limit, since understanding an item context takes quite a lot of time, and it will be impractical to have students comprehend so many different real-world contexts within one exam. We also see that it is not easy to have questions that involve the idealization step, but the item #4 is able to because the problem situation itself deviates more from the standard Bernoulli model. Finally, the Hauptschule items are less demanding, and the kind of questions is more regular since all of them are *classical* application tasks.

Bavarian Haupschule and Gymnasium, Germany, 2004-2012.

During the period 2004-2012, Germany's curriculum was affected much by international assessment programs such as the TIMSS and PISA. It is therefore expected to see more modeling and application requirements. In the unified standards of the 16 federal states, that were decided by *KMK*[10], six *general mathematical competencies* were indicated:

K1 Argue Mathematically
K2 Solve problems mathematically
K3 Model Mathematically
K4 Mathematical representations
K5 With symbolic, formal and technical elements of mathematics

[10] *Kultusministerkonferenz*, or the Standing Conference of the Ministers of Education and Cultural Affairs, is "a consortium of ministers responsible for education [...] and in this capacity formulates the joint interests and objectives of all 16 federal states" (www.kmk.org)

K6 communicate (KMK04Hau, p. 7)

We see that modeling becomes, at least nominally, a major component of mathematics standards. Note that it is not only K3 that relates to our sense of mathematical modeling, since some modeling steps overlap with the other competencies. It is also observed that the curriculum standards indicated three levels of mathematical problem solving – *reproduction*, *connection*, and *reflection*, the same as seen in PISA documents (OECD, 2000; 2003). In particular, for the modeling competency, these three levels are expressed as follows (translated):

```
Reproduce
- familiar and directly recognizable models
- assign mathematical objects to simple phenomena from the world of
experience
- Check results at context
Create connections
- Modeling that requires several steps
- Interpret the results of a modeling and test them at the initial
situation
- assign appropriate situations to a mathematical model
Generalize and reflect
- model complex or unfamiliar situations
- reflect and critically evaluate applied mathematical models (such
as formulas, equations, representations of assignments, drawings,
structured representations, flowcharts) (p. 13, KMK04Hau).
```

Given this emphasis of modeling in the *KMK*'s education standards, it is beneficial to review the specific task examples that are given within the standards document, *Bildungsstandards im Fach Mathematik für den Hauptschulabschluss* (Educational standards in mathematics for the Haupschule leaving certificate; Code: KMK04Hau). In the document, not only is modeling counted as one of the six competencies, there are also 15 modeling examples presented. After reviewing Bavarian's curriculum materials for *Hauptschule* and *Gymnasium*, the modeling examples will be reviewed together with the *Quali* and *Abitur* tasks.

In the 2004 curriculum plan of Bavarian *Hauptschule*, its social and economic efficiency ideology did not seem to change, and perhaps strengthened by more emphases on application. Moreover, "Application competence" (*Anwendungskompetenzen*) was specified among the five basic competencies throughout the school years; the other four include numeric competence, computational competence, algebraic competence, and geometric competence. This change of the curriculum is consistent to the *KMK* standards.

As for the curriculum plan of *Gymnasium*, the descriptions somehow introduce phrases that indicates social and economic efficiency, other than academic rationalism:

> BY04Gym, Fachprofile: Mathematik, p. 1
> Curriculum Ideology: Academic Rationaism & Social Economic Efficiency
>
> Die Mathematik hat sich über Jahrtausende als gemeinsame Kulturleistung der Menschheit entwickelt. Sie erfasst Aspekte der Wirklichkeit und erarbeitet Begriffe, Theorien, Strukturen und Modelle. Unter Wahrung ihrer Eigenständigkeit bietet sie Ideen und Methoden zur Lösung von Problemen aus unterschiedlichsten Disziplinen an; sie liefert als dynamische Wissenschaft wesentliche Beiträge zur Beschreibung und Gestaltung unserer Welt. Mathematik ist traditionell ein charakteristischer Teil der Sprache der Naturwissenschaften und der Technik. Aber auch in Wirtschaft und Politik sowie in den Sozialwissenschaften bilden mit mathematischen Methoden gewonnene Aussagen häufig die Grundlage für Entscheidungen von weitreichender Bedeutung.
>
> Die zentrale Aufgabe des Mathematikunterrichts am Gymnasium ist es daher, den Schülern neben konkreten mathematischen Kenntnissen und Arbeitsweisen auch allgemeinere Einsichten in Prozesse des Denkens und der Entscheidungs findung zu vermitteln, die für eine aktive und verantwortungsbewusste Mitgestaltung der Gesellschaft von Bedeutung sind. Dabei wird den jungen Menschen deutlich, dass Mathematik ein hilfreiches Werkzeug zur Analyse und zur Erkenntnisgewinnung sein kann, das letztlich auf menschlicher Kreativität beruht, und dass die Mathematik auch wegen ihrer ästhetischen Komponente einen Wert an sich darstellt.

> *Translation:* Mathematics has developed over thousands of years as a common cultural achievement of mankind. It captures aspects of reality and elaborates concepts, theories, structures and models. While maintaining its independence, it offers ideas and methods for solving problems from a wide range of disciplines; As a dynamic science, it makes a significant contribution to the description and design of our world. Mathematics is traditionally a characteristic part of the language of natural sciences and technology. But also in business and politics, as well as in social sciences, statements derived using mathematical methods often form the basis for decisions of far-reaching significance.
>
> The central task of mathematics teaching at the Gymnasium is therefore to teach pupils, in addition to concrete mathematical knowledge and methods of work, also more general insights into processes of thought and decision-making, which are important for an active and responsible participation in society. It shows that mathematics can be a helpful tool for analyzing and gaining knowledge, which is ultimately based on human creativity, and that mathematics also has a value in itself because of its aesthetic component.

In other words, the role of mathematics is not only a foundation of many disciplines, but also a *contributor to the world*. A student who masters mathematics acquires ways of thinking that support *participation in society*. In the following, we will review three groups of tasks, including: (1) Modeling examples offered in the *KMK* standards; (2) Bavaria's 2012 *Quali* items; (3) 2012 *Abitur* items.

The *KMK* modeling examples (Table 5.21) originally served as an attempt to offer possibilities to develop or assess the six competencies, but an analysis of their modeling spectra also help us find ways to assess modeling steps that were not involved often. For example, [3] *idealization* and [6] *interpretation* are not often observed, but we see five tasks that provide such opportunities; [7] *verification* was rare, but we also see four items here that involve this step. The item *Ferry Tickets* is an application to look for a pricing scheme, and it involves six modeling steps. The item description can be translated as:

Modeling Example: Ferry Ticket, Translation

```
On the landing stage of a large ferry you will find this price
list:
Single ticket, 1 person, EUR 50.00
Group ticket, 8 people, EUR 380.00
Group ticket, 20 people, EUR 900.00
(a) Calculate the best price for 16 people.
(b) What discount will Peter receive if he purchases the
corresponding group ticket instead of 20 individual tickets?
(c) For a group of 24 persons, Frank will charge a price of EUR
1140.00. Maike says the group can drive more cheaply. Is she right?
Justify.
(d) The ferry company wants to introduce a block card for 50
persons. What price would be appropriate? Justify.
```

The real-world question depends on the party who asks it: from the consumers' perspective, they would love to have the most discount for the same service; from the company's perspective, they want to see that the pricing scheme is attractive as well as appropriate. However, since the task writer did not define *appropriateness*, the student needs to *justify* why certain pricing scheme is such. For example, the new group ticket for 50 should be cheaper than the price for 50 people under the existing pricing. Of course student will also need to mathematize the situation and do the calculations, but instead of having student calculate everything, the task offered an answer

‘1140.00’ for a 24-people group. Student has to find out if it is possible to get more discount. Finally, student has to explain (justify) the calculations that were done.

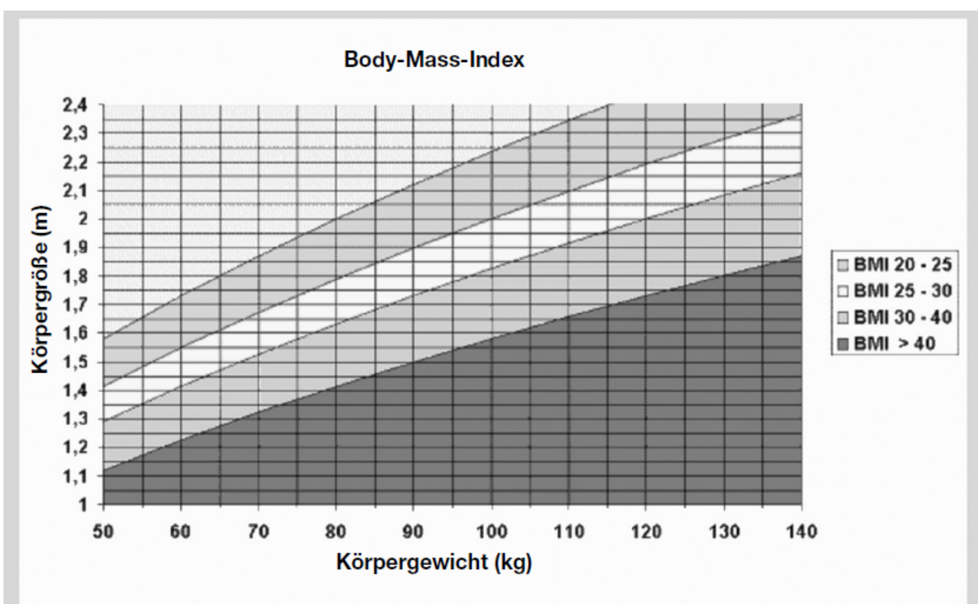
The item *Giant Barrel* showed one picture (Figure 1.2, Chapter I, p. 4) and only one question: “How much liquid fits into this barrel? Explain your answer.” Since nothing but the picture was provided, student has to estimate the dimensions using the typical height of the men, and probably need to approximate the barrel as a cylinder, therefore it is required to work on *idealization*. However, *verification* cannot be done unless the actual capacity is measured and/or told. On the contrary, another item, *Body Mass Index*, provided rich information: A link to a health service website, a table (not presented here) describing how different BMI ranges correspond to normal or abnormal weight, a sentence-definition of *BMI*, and a heat map chart that shows the relationship between BMI range, height, and weight. Consequently a question may be solved by multiple methods. For example, one question in the item assumes that Sarah’s father is 1.78m tall and weighs 103 kg, and asked for her father’s BMI value. Also, if her father is aiming at normal weight, how many kg must he lose? Then the student can use the BMI definition or the chart. If the student uses the chart, then it is a step of interpretation, since the mathematical result was given as the chart, and one can search on it for information in real-world terms (e.g. how many kilograms to lose).

From these two contrasting examples, it can be argued that a modeling problem does not directly depend on number of words or the given steps. The *Giant Barrel* task can test the ability of modeling because modeling is *possible* (as done by the problem writer), and the mathematics used in a model does not go beyond students’ expected mathematical knowledge. The problem writer may decide to give much more information, as is done in the *Body Mass Index* task, as a

Table 5.21: Modeling spectra of 15 examples in KMK 2004 Standards for Hauptschule leaving certificate, 2004.

#	Name	Application?	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
1	Terrace Slabs	Yes	110	110	000	001	001	001	000	000
2	Giant Barrel	Yes	110	110	001	001	001	000	000	000
3	Clearance Sale	Yes	000	110	000	001	001	000	001	000
4	Body Mass Index	Yes	000	110	000	110	111	001	000	000
5	Wheel of Fortune	Yes	000	110	000	001	001	000	000	000
6	Landscaping	Yes	000	100	000	001	001	000	000	000
7	Cube	Yes	000	110	000	110	001	000	000	000
8	Statistical Survey	Yes	110	111	001	001	000	000	000	000
9	Calculation Error	No								
10	Cellphone Cost	Yes	110	110	000	111	001	000	001	000
11	Ferry Tickets	Yes	110	111	000	001	111	001	001	000
12	Copying Paper	Yes	110	100	000	001	001	000	000	000
13	Cornflakes Pack	Yes	110	111	000	001	001	000	000	000
14	Investment	Yes	110	110	000	111	001	000	001	000
15	Planar Figures	No								

Figure 5.24: KMK04Hau, Item #4, Body Mass Index.



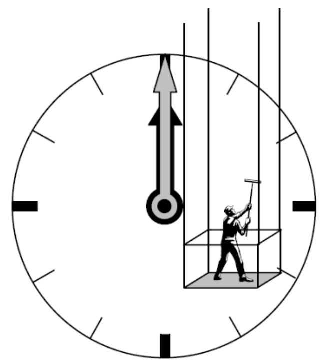
This item is reprinted from *Bildungsstandards im Fach Mathematik für den Hauptschulabschluss Beschluss vom 15.10.2004*, copyright 2004 by KMK Bonn und Berlin.

scaffolding network so that students can work within. In an extreme, a complete modeling may be tested by presenting a real situation which has not been specifically modeled by anyone, so that a modeler could begin the modeling process themselves by posing real-world question and/or doing conceptual modeling. Such scenario can be seen in modeling competitions or more realistic situations, but we will not study these cases in the current work because the tools are not ready for such complicated objects yet. Indeed, the current work is undergoing a conceptual modeling process, and an idealization is made by setting constraints within an environment of education.

Figure 5.25: BY12QH, Item A8, Size of a clock

8. Die Abbildung zeigt einen Arbeiter, der das Zifferblatt einer großen Turmuhr reinigt.

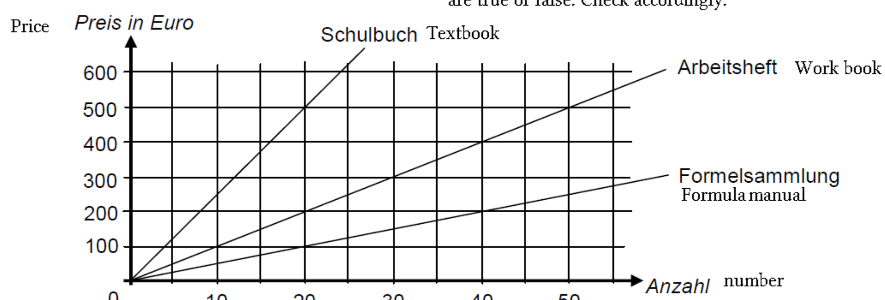
Welchen Umfang hat das Zifferblatt ungefähr? Begründe.



Translation:
8. The figure shows a worker who cleans the face of a large clock. What is the size of the face ? Justify.

Figure 5.26: BY12QH, Item A10, Cost of Textbooks

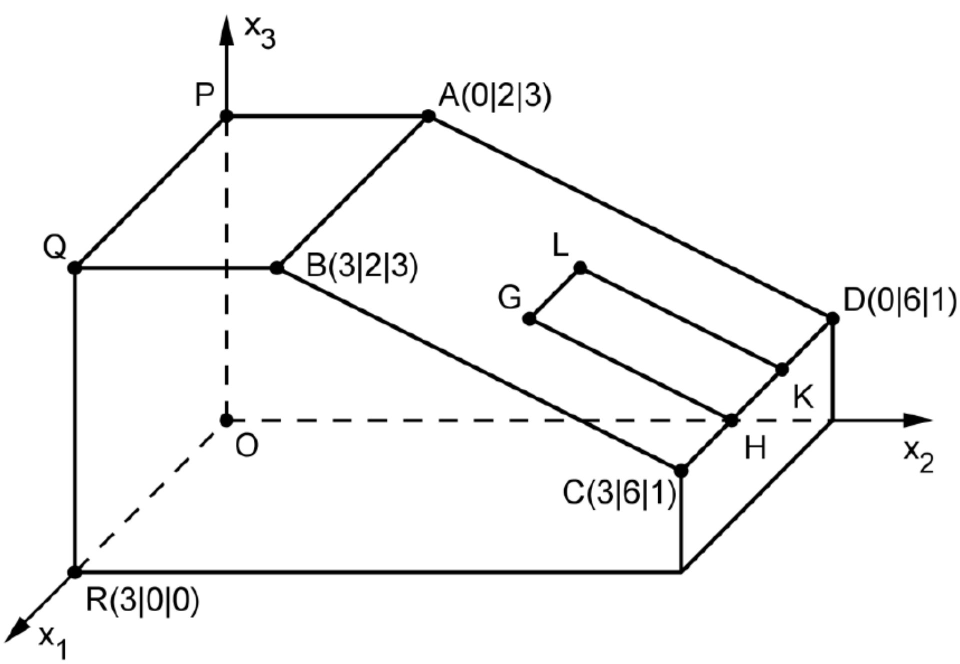
10. Entscheide mit Hilfe des Diagramms, ob die folgenden Aussagen wahr oder falsch sind. Kreuze entsprechend an. Tr.: Use the graph to determine whether the following statements are true or false. Check accordingly.



	true wahr	false falsch
40 formula manuals cost as much as 20 work books. a) 40 Formelsammlungen kosten so viel wie 20 Arbeitshefte.	☐	☐
20 textbooks cost four times as much as 20 formula manuals. b) 20 Schulbücher kosten viermal so viel wie 20 Formelsammlungen.	☐	☐
c) 2 Schulbücher kosten 50 €. 2 textbooks cost 50 Euros.	☐	☐

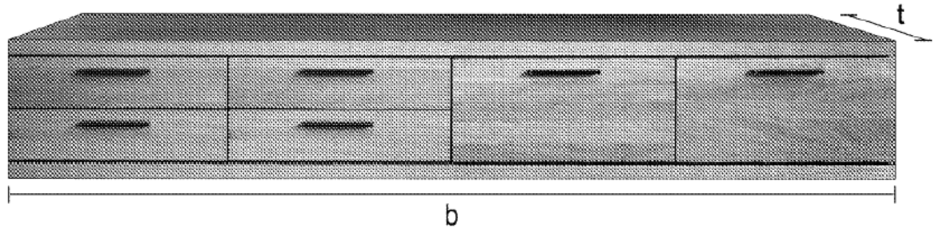
This item is reprinted from*Qualifizierender Hauptschulabschluss 2012, Besondere Leistungsfeststellung Mathematik (27. Juni 2012)*, copyright by Wolfgang Kuplent, 2012. Permission granted.

Figure 5.27: BY12Abi, Item G1, Figure of the roof-top room.



This item is reprinted from*Abiturprüfung 2012 Mathematik*, copyright 2012 by Wolfgang Kuplent. Permission granted.

Figure 5.28: BY12Abi, Item G1, Picture of the furniture.



This item is reprinted from*Abiturprüfung 2012 Mathematik*, copyright 2012 by Wolfgang Kuplent. Permission granted.

Finally, the analysis of the *Quali* and *Abitur* examinations in 2012 is presented briefly below, and can be compared with the 2001 versions. Both tests do not change a lot in the format of individual items, but the first part of *Quali* becomes a "Part A" which must be done without a calculator, and the other three parts are under "Part B" which can be done with a calculator. The *Abitur* provided more time, which has been increased to 240 minutes from 180 minutes. Items of *Quali* 2012 were still predominantly *classical* application tasks, and the overall proportion of application tasks did not change, but it started to have some items require interpretation steps (Figures 5.25 and 5.26, pages 173 to 174). *Abitur* paid much more emphasis on modeling compared to 2001, as can be seen as the number of items that are related to application. In 2001, only probability problems bore with application context (almost inevitably), but in 2012, two more items welcomed applications, with one analysis and one geometry. The Abitur items continue to be long, usually with one overall context and several related questions. More interpretation and verification steps can be assessed because of this style, but idealization step is still not tested. The *Abitur* problems are like "mini projects", and we take the geometric modeling item as an example. The complete problem is not presented, but the main scenario is like this: First, a roof-top room is modeled as a right prism as shown in Figure 5.27, where the floor and is in the x_1x_2 plane, and a wall lies in a plane through CD, and is parallel to the x_1x_3 plane . There is also a rectangular window

$GLKH$ that is rotatable about an axis passing through the centers of its sides GH and LK, with the lower edge of the window swings into the room; the swivel joint allows an end position so that the window surface can be perpendicular to the floor. One real-world problem to be solved is, in this design, one needs to confirm by calculation that the window cannot touch the floor when it is rotated. Now, one piece of furniture (see Figure 5.28, p. 175) is 40 cm high. In Figure 5.27 (p. 174), the length unit is 1m. So, another real-world problem is this: if the furniture is to stand with its back put against the wall (under the window), will the rotatable window bump into the piece of furniture? (It is given that the depth $t = 50$cm).

Analysis of Hong Kong curriculum materials

Public Schools in Hong Kong, period 1995-2003 and 2004-12.

Hong Kong is small enough to have a unified education system, though small state does not imply unity in education (Bray & Koo, 1999). Since 1935, secondary students in Hong Kong had been taking the HKCEE (Hong Kong Certificate of Education Examination) at the end of Form 5 (equivalent to America's 11th Grade), and would take the Advanced Level Examination throughout Form 6 and Form 7. In 2012, the HKDSE (Hong Kong Diploma of Secondary Education Examination) has replaced both exams, which pupils only take at the end of the sixth year of secondary school. All these exams are administered by the Hong Kong Examinations and Assessment Authority. The exam-oriented culture in Hong Kong is a worthy topic to discuss, but for this study, it should be taken as a convenience, that the assessed curriculum alone tells us much about the overall curriculum.

The mathematics syllabus in 1999 was quite concise, and straight to the point of social and economic efficiency:

HK99SSS (1999 Secondary Maths Syllabus).

Curriculum Ideology: Social and Economic Efficiency.

```
The secondary school mathematics curriculum continues the
development of the learning of mathematics in the primary school. To
enable students to cope confidently with the mathematics needed in
their future studies, workplaces or daily life in a technological
and information-rich society, the curriculum aims at [...] (p. 4)
```

Modeling was not found in the document, but application has its important role of promoting student learning, since it would be a more effective way to learn when students find mathematics relevant to their lives:

```
Students who find the study relevant to their experience will
be motivated to learn the subject. Daily life applications are
emphasized in the curriculum. Stories of historical development of
mathematics knowledge are included to enable students to understand
mathematics knowledge evolved from real-life problems and refined
after years. A new module "Further Applications" which includes
the application of mathematics in more complex real-life situations
requires students to integrate their knowledge and skills from
various disciplines to solve problems.
```

Hence, the HK99SSS specified real-life applications such as travel graphs, taxation and loans, stock market index, and so on. In the *2002 Mathematics Education Key Learning Area* (HK02KLA), application is seen as *numeracy skill*, which is one of the nine generic skills, including collaboration, communication, creativity, critical thinking, information technology, numeracy skills, problem-solving, self-management skills, and study skills. Numeracy skills include "the ability to perform basic computations, use mathematical concepts in practical situations, to make reasonable estimates, understand graphs, charts and numerical concepts in languages, to manage data, to handle money and do stock inventories." (HK02KLA, p. 17)

The newer *2007 Mathematics Education Key Learning Area* (HK07KLA) not only kept this ideology, but also added perspectives of *academic rationalism*, i.e. seeing mathematics as a fundamental subject that provides an understanding of other disciplines and of the world, and of *progressivism*, emphasizing students' intellectual development through mathematics, and the adaptiveness provided by this subject (see excerpt).

> HK07KLA.
> Curriculum Ideology.
>
> `Mathematics is a powerful means in a technology-oriented and information-rich society to help students acquire the ability to communicate, explore, conjecture, reason logically and solve problems using a variety of methods. (p. 1)`
>
> `Mathematics provides a means to acquire, organise and apply information, and plays an important role in communicating ideas through pictorial, graphical, symbolic, descriptive and analytical representations. Hence, mathematics at the senior secondary level helps to lay a strong foundation for students' lifelong learning, and provides a platform for the acquisition of new knowledge in this rapidly changing world. (p. 2)`
>
> `Many of the developments, plans and decisions made in modern society rely, to some extent, on the use of measures, structures, patterns, shapes and the analysis of quantitative information. Therefore, mathematical experiences acquired at the senior secondary level enable students to become mathematically literate citizens who are more able to cope with the demands of the workplace. (p. 2)`
>
> `Mathematics is a tool to help students enhance their understanding of the world. It provides a foundation for the study of other disciplines in the senior secondary and post-secondary education system. (p. 2)`
>
> `Mathematics is an intellectual endeavour through which students can develop their imagination, initiative, creativity and flexibility of mind, as well as their ability to appreciate the beauty of nature. Mathematics is a discipline which plays a central role in human culture. (p. 2)`

Having these in mind, we will review the mathematics exams in these two periods. The HKALE and HKCEE in 2002 will be reviewed for the period 1999-2003, and HKDSE in 2012 for the period 2004-2011.

Since the HKALE was equivalent to a college entrance qualification, it is expected that students should be more mathematically competent comparing to those who did not take it. The exam in 2002 was first divided into "Pure Maths" and "Applied Maths" exams. The Pure Maths exam does not have any application tasks, and the Applied Maths exam was further divided into *Paper 1*, which was mainly classical mechanics, and *Paper 2*, on probability and statistics. Classical mechanics can be seen as an overall context where movement of objects is understood by physi-

Figure 5.29: HKALE 2002, Paper 1, Item #8. *Record Breaking.*

(a)

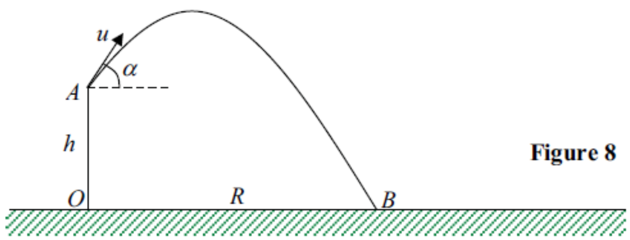


Figure 8

A particle is projected at an angle of elevation α, with speed u, from a point A at a height h above a point O on horizontal ground. The particle lands at a point B on the ground at a distance R from O (see Figure 8). Show that

(i) $gR^2 \tan^2 \alpha - 2u^2 R \tan \alpha + gR^2 - 2u^2 h = 0$,

(ii) $u^2 \geq g(\sqrt{h^2 + R^2} - h)$.

(8 marks)

(b) In a school's annual athletic meet, the past record in the men's shot-put was 18.00 m. Assume that the shot was projected from the shot-putter's height above the ground and that air resistance was negligible.

(i) A 1.70 m tall shot-putter participated in the event. Find the least speed of projection for him to equal the record. Find also the corresponding angle of projection. Give your answers correct to 3 significant figures.

(ii) If the fastest speed a certain shot-putter could project the shot was 12.3 m s^{-1}, do you think he would have a chance to break the record ? Explain your answer.

(7 marks)

Table 5.22: Modeling spectra of problems in HKALE 2002, Applied Maths Paper 1 &2.

#	Item	Application?	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
	Paper 1									
	(11 items)	Yes	000	111	000	001	001	000	000	000
8	Record Breaking	Yes	110	111	110	001	001	111	000	000
	Paper 2									
1	Lateness	Yes	110	110	000	001	001	000	000	000
2	Dice	Yes	000	110	000	001	001	000	000	000
3	Bird Flu	Yes	000	110	000	111	001	000	000	000
	(5 items)	No								
9	AC Curcuit	Yes	000	110	000	001	001	000	000	000
10	Oil Wholesale	Yes	000	110	000	001	001	000	000	000
11	Machine Game	Yes	110	110	000	001	001	001#	000	000
12	IQ Survey	Yes	000	110	000	110	001	000	000	000

The # mark indicates an undesired possibility that although the item requires a certain step in order to be solved, students can solve it with other steps only.

cal concepts (conceptual model), and then modeled by physical quantities and equations. For all the 12 tasks of Paper 1, students have to work on the conceptual models in order to understand what quantities and equations are relevant. Almost all tasks have the same modeling spectrum: no real-world question was posed; only mechanical models needed to be worked on; some mathematization and mathematical solution is necessary. One distinctive task involving more modeling steps is presented in Figure 5.29 (p. 179). The projectile model was presented first in question (a), so that students may work on and solve the model. It followed by question (b) that applied the modeling directly. In sub-question (i), students were required to understand the model in (a) completely, so that they can substitute the corresponding quantities. In sub-question (ii), one either needs to realize that the expression in the right hand side of (a)(ii) is monotonically decreasing with h, or intuitively argue that higher h requires less initial speed u to break the record. Then, using the same equation, student may argue that even an extreme condition like $h = 2.5$ would require $u \geq 12.4$. A written explanation is also required. Table 5.22 shows the modeling steps involved in HKALE (Papers 1 & 2).

In Paper 2, about half of the 12 problems are application tasks related to probability and statistics. Out of the 7 application tasks, 5 of them are *standard mathematization* problems. One of them, the *Bird Flu*, has a quite relevant context, since Hong Kong suffered seriously from bird influenza in 1997.

> HKALE 2002, Paper 2, #3 Bird Flu
>
> A farm had a large number of chicken. At time $t = 0$ (in days), 1% of the chickens were found to have been infected with bird flu. After 2 days, it was found that 50% of the chickens had been infected. Denote by $p\%$ the percentage of chickens on the farm that had been infected with bird flu. It is known that the rate of increase of p with respect to t is proportional to $p(100 - p)$.
>
> Write down a differential equation for p. Hence express p in terms of t.

The model has two parameters: a proportionality coefficient, and an arbitrary constant due to integration. They have to be determined using the data provided.

In the rest of the two items, one of them is a classical application task, and another is an actual machine game. The latter is shown in Figure 5.30 (p. 182). If there were no (c), the task would only be one of the standard mathematization problem. Student had to say that the game was fair or not, and defend their own answer by conceptual and mathematical reasoning. In an "exam culture", it should be noted that "standard responses" for interpretation may be prevalent among students. In this case, "fair game" may be seen as a *synonym* to "zero expected net income". That is, it is no longer a real-world expression but a modeled concept, and therefore the question does not ask for interpretation, but simply understanding of the model. A "#" mark is attached to the modeling spectrum to indicate this possible exception, which will be further discussed in the Conclusion chapter, as an opportunity for future studies.

The 2002 HKCEE exam for Form 5 student had 17 extended response items in *Paper 1*, and 54 multiple choice in *Paper 2*. Out of these 71 items, there are 12 (17%) application tasks. 7 of them are standard mathematization tasks, and 5 of them are classical applied tasks. Most context is closely related to monetary concepts or situations, such as currency exchange, interest, expenditure of a

Figure 5.30: HKALE 2002, Paper 2, Item #11. *Machine Game.*

Figure 3

Figure 3 shows a machine game. There are four identical balls on a board carved with nine cups. Each cup can hold one ball only. In each play of the game, a player inserts a token into the machine and all the balls will then bounce up. Each ball eventually falls randomly on one of the nine cups. Tokens are awarded to the player if one of the following patterns is obtained :

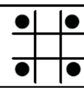
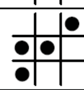
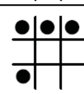
Prize	Pattern	Number of tokens awarded
First	The four balls fall on the four corner cups. (i.e. : )	6
Second	Any three balls fall on a diagonal of cups. (e.g. : )	4
Third	Any three balls fall on the same row or column of cups. (e.g. : )	2

Table 3

(a) Find the probability of winning a first prize in a play. (1 mark)

(b) Show that the probability of winning no prize in a play is $\frac{11}{18}$. (5 marks)

(c) Is this a fair game ? Explain your answer. (3 marks)

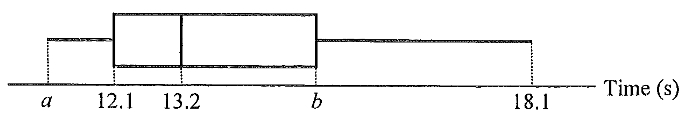
(d) Initially John had two tokens. If he then played the game twice, what would be the probability for each of the following events ?

(i) He would have no token left.
(ii) He would have at least four tokens left. (6 marks)

family, price of goods and profit, and maximization of profit under constraints (Paper 2-#12, 33, 14, 1-#17, and 2-#10, respectively; these items are not presented here). The items' modeling spectra was quite uniform, and modeling steps like idealization, interpretation, and verification are not assessed.

Figure 5.31: HKDSE 2012, Paper 1, Item #7. *Training Program.*

The box-and-whisker diagram below shows the distribution of the times taken by a large group of students of an athletic club to finish a 100 m race:



The inter-quartile range and the range of the distribution are 3.2 s and 6.8 s respectively.

(a) Find a and b .

(b) The students join a training program. It is found that the longest time taken by the students to finish a 100 m race after the training is 2.9 s less than that before the training. The trainer claims that at least 25% of the students show improvement in the time taken to finish a 100 m race after the training. Do you agree? Explain your answer.

(4 marks)

The 2012 HKDSE exam has a similar format to the 2002 HKCEE, having 19 extended responses items in *Paper 1*, but only 45 multiple choice in *Paper 2*. 13 out of the 64 items (20%) are application tasks. Except classical applied tasks and standard mathematization, we also observe some one task that require an interpretation step (Item 1-#7, Figure 5.31, p. 183). The box-and-whisker and the stem-and-leaf plots are introduced to the exam. There are also some issue in the item 1-#19 (whole problem not presented). It models an air cargo terminal's goods handling capacity by $A(n)$ tonnes in the nth year, where $A(n) = ab^{2n}$. Since it did not restrict the domain of n, it would mean that the handling capacity of the terminal grows exponentially all the time. This could be problematic because in reality, the goods handling capacity will be capped at some point. How realistic should a model be in a modeling task? It may depend on the available mathematical tools and the students' maturity, but it can be made clear that mathematical models have limitations, which can be seen a good opportunity to assess the interpretation and the verification steps that the tasks are lacking.

Summary of findings, and discussions

The five analyses done above showed a variety of modeling problems used in the intended and the assessed mathematics curricula, including the PISA studies 2003 and 2012, Older (1995-2003) and newer (2004-12) documents of USA (Florida and Massachusetts), Germany (Bavarian *Hauptschule* and *Gymnasium*), and Hong Kong. To narrow the scope to the education of modeling and application, we applied the framework offered in Chapter 4 to study (1) the curriculum ideology, (2) the role of modeling and/or application in the curriculum, and (3) the modeling steps involved in the assessment tasks. Note that we did not take the *holistic* view of modeling, due to the fact that each of the assessment items are predominantly expected to be done in a rather short time. A complete consideration of the modeling process may be more suitable for in-class activities or project assignments that are designed for team work. Instead, an *atomistic view* of modeling assessment is more convenient not only in the item design, but also in our analysis. To compensate for the disadvantage of the atomistic view, which considers the modeling steps as separated activities, we designed a tool called the *modeling spectrum*, which not only observes whether a task is related to a certain modeling step, but also how a step is involved: (1) A modeling step can be *described* by the task; (2) the description of a step may need to be *understood* by the student (or not); (3) the student may need to *perform* a particular modeling step.

After the systematic review of the curriculum documents from different education systems, including the curriculum framework and standards, and some public examination materials, some findings can be summarized as follows. Firstly, the contexts of an item follow the specifications of the curriculum documents, such as multicultural awareness, and the inclusion of contexts about daily activities, so that pupils may learn mathematics with connections to the society; however, the way that such a task relates to a modeling process can affect this curriculum goal. There are items in which the context is not related to the solution, such as the *context-detached* problems (as in some of the Florida's items) and the *extraneous context* (as in some of the Massachusetts' items). These may not help students stay connected, but rather strengthen the idea that mathematics is not so much relevant. Admittedly, we do not have to hold Pythagoras' motto, "All is number", but we can support students with these assessment items to see the relevance and the limitations of mathematics, namely, when it is useful *or not.*

Secondly, we found that all items fail to assess the first and last modeling steps. In other words, the tasks did not ask student to pose real-world relevant questions that can be solved by mathematical modeling, and also did not let students re-iterate a modeling process, if it needs improvement. On the contrary, the vast majority of application items assessed students to work mathematically within the model, and mathematize the situation. The two main classes of problems are the *classical application tasks* and the *standard mathematization.* The former presents the conceptual and/or the mathematical models, and students have to understand them in order to solve mathematically, but do not have to take part in searchng for a model. The latter requires students to look for appropriate mathematics to do, which means they need to figure out what operations, variables and equations can be used in order to solve the task.

Thirdly and lastly, it is found that, since the problems can be classified by their modeling spectrum (p. 107), one may use this tool to detect if more or less of a kind of problem is needed, which is beneficial to test design. After all, all the curriculum documents would require every aspect that seems good for mathematical learning, including logical arguments, calculation, problem solving, communication, and so on, but the *weights* of these aspects are not specified. Such weights can be revealed when one analyzes the modeling spectrum of the tasks.

Some discussions of issues and limitations are presented below.

(1) We applied this tool to all the studied curricula except the PISA items, since the tool was formed during a systematic analysis of Florida's items. Also, most of the PISA items are not released and therefore can only be studied by the item reponses but not the item descriptions.

(2) We did not analyze the whole range of modeling problems. Most of the analyzed items involve about three modeling steps (model conceptually, mathematize, and work mathematially) and only ask students to perform one or two of them. Students may still benefit from other opportunities like modeling competitions, modeling projects, and actual modeling tasks in the real-world to learn about the realistic usage of mathematics, but the *assessed curriculum* virtually overrides these opportunities. Curriculum and assessment designers should not shy away the more realistic, and perhaps more problematic, aspects of mathematical learning, and should allow *all* students to gain access to applications through their *regular* curriculum.

(3) The modeling spectrum was determined by the author only, and therefore the study of items is not considered conclusive, but exploratory. If more researchers work together on the classification, an *inter-rater reliability* may be calculated.

(4) Moreover, there are some classifications that are debatable. For some items, in the mathematization step, the author would need to decide whether "110" or "001" should be filled in: "110" means that the model is already provided in the task, so the student only needs to understand the model and do the calculations; "001" means that the model is not provi-

ded, but the student needs to decide on the mathematics to be done. For example, "If two fair dice are independently tossed, what is the probability that the sum of the two numbers faced up is a multiple of 5?" If one treats probability as a daily concept, then one needs to *decide* to count the number of possibilities that a sum is a multiple of 5, and then divide it by 36. However, if we treat probability as a result of mathematical modeling, then one can argue that the model is already given, and the student only need to use the operations that probability theories specify.

(5) Though the modeling steps involved are identified, the *difficulty* of each step is not considered. Admittedly, the mathematization of the total money that I have after spending some for lunch may be much easier than considering all the constraints for the maximization of profit. Also the item's real-world relevance is not considered. The context presented by the task may be very relevant to the real-world, but the modeling spectrum depends only on whether the student needs to understand it in order to solve the task successfully. However, the relevance of the context may affect how a student approaches the task.

CONCLUSION

The literature has stressed the importance of modeling assessment in mainly two ways. Firstly, in general, the assessed curriculum affects the whole curriculum a lot, and therefore any change in the (high-stake) assessments can eventually propagate to the whole system. Secondly, modeling tasks constitute the main materials for the teaching and learning of modeling, since the tasks are the interface between the abstract (mathematics) and the concrete (real-world). On one hand, one cannot expect students to construct their own model if they are only submerged into a real-world situation without instructions; on the other hand, students cannot learn how to apply mathematics only by drawing an abstract modeling cycle. Therefore, modeling tasks are at the heart of modeling education.

Because of this, many researchers had investigated in the problem and conceived many theories. The first part of the analysis was to review the theories systematically (in Chapter IV). The results are summarized as answers to research question 1:

Answers to Research Question 1

(a) On the ability to apply mathematical knowledge, to what extent do current theories explain existing empirical research?

(b) Do these theories constitute a larger framework that can analyze the ability to apply mathematical knowledge?

To answer RQ1(a), first we found that there are many groups of theoretical claims on various topics about modeling, listed near the end of Chapter 4. Seven of them are chosen Though all of these frameworks are elaborated well enough, the claims in these theoretical frameworks do not receive enough attention in empirical studies, and therefore they are neither supported nor refuted substantially according to Schoenfeld's standards (2000).

To answer RQ1(b), we see that the three perspectives cannot be seen as one theory in the classical sense, but we can use the idea of networking (Bikner-Ahsbahs & Prediger, 2010) to create a *dynamic* framework of ability to apply mathematical knowledge, which means that the solidness of theory is not yet achieved, as promoted by Kilpatrick (2010) and Niss (2007). The framework is a working tool to study the educational phenomena related to the ability to apply mathematical knowledge but allows for more theoretical elements to be added in, until some major conflicts occur, or some empirical studies that significantly support some elements over others. In order to apply the results to study the next question, three major frameworks for assessment are identified: (1) The atomistic, (2) The holistic, and (3) the Models and Modeling Perspective (Tables 5.1, 5.2, and 5.3)

Answers to Research Question 2

(a) To what extent are PISA 2003 and PISA 2012 designed to assess the ability to apply mathematics?

(b) How were the test items related to the modeling process?

Using the framework generated from RQ1(b), an analytical tool for assessments was presented in Table 5.4, and was applied to PISA's assessment frameworks and items. To answer RQ2(a), keeping in mind that we conceptualized the ability to apply mathematics as *modeling competence*, we had investigated what role modeling plays in PISA's assessment framework. It was found that although the main focus of PISA is *mathematical literacy*, the actual process that it tries to assess is the "mathematising process", which is essentially a variant of the *modeling cycle*. It was stated as a crucial part of the PISA tasks.

To answer RQ2(b), we take the atomistic approach and inspect how each PISA item is related to each modeling step. Comments are also made to explain why an item is related to the steps specified. It was as expected that most PISA items involve only one or two modeling steps. The first and last parts of the modeling process is less tested, but are found to be more than other public assessments, as we will see in the results of RQ3. Nevertheless, it was found that some items seemingly included real-world context, but it is possible that student solves the problem without taking the context into account. In that case, we call such items *camouflage* word problem.

Answers to Research Question 3

(a) Between 1995 and 2003, what was the role of modeling and application in the curriculum documents of the United States, Germany, and Hong Kong?

(b) In this period, how were their public assessments related to the modeling process?

(c) Between 2004 and 2012, what was the role of modeling and application in the curriculum documents of the United States, Germany, and Hong Kong?

(d) In this period, how were their public assessments related to the modeling process?

As in the answers to research question 2, the analytical tool for assessment is used (i.e. Table 5.4) combining with some guiding questions of curriculum studies (adopted from Adamson & Morris, 2014)). This has been helpful for the study of curriculum and assessment of different education systems. During the analysis, it was found that the holistic framework benefited the most for obtaining insights. Furthermore, the construct of *modeling spectrum* is based on this framework. The tool helped the author to analyze many test items systematically, and can be concisely recorded in a table (such as 5.22). Before reporting the answers to research questions, it is more convenient to review briefly about the modeling spectrum (cf. Table 5.6), which is a table that records how each modeling step is covered by an item. The record is not simply "covered or not covered", but based on the answering to the following questions:

(i) Does the task description contain or imply the modeling step?

(ii) If so, is it necessary to understand in order to solve the task?

(iii) Does the problem solver need to perform the modeling step in order to solve the task?

Items can be roughly classified according to various modes of responses to these questions. The most typical ones are:

- *Classical applied mathematics*: Student is required to understand some of the modeling steps presented or implied by the problem. That is, the answer to (i) and (ii) must be *Yes* in some steps such as [2] model conceptually and [3] mathematize.
- *Standard mathematization task*: Student needs to apply appropriate mathematical model / operation that is not presented or implied by the problem. The answer to (iii) must be *Yes.*

This helped the author to singularize the other less typical items, which can either be worth mentioning in its way of assessing some modeling steps, or particularly problematic otherwise.

The answers to RQ3 can be summarized as follows:

(1) Between 1995 and 2003, the idea of modeling was not elaborated in Florida and Massachusetts' curriculum, though the phrases "real-world" were read many times in some of the documents. The items showed that the interest of modeling steps is rarely more than a mathematization

step, where students need to figure out what mathematics is to be done before doing any calculations; however, the ways to figure it out does not need much thinking of the real-world context. Between 2004 and 2012, both Florida and Massachusetts were putting forward the Common Core State Standards initiatives (CCSS), and therefore necessarily to accommodate the principles of modeling, which is one of the *Standards for Mathematical Practice* in CCSS. However, it was found that the public assessments in this period still had not involved more modeling. Massachusetts' tasks contain even less modeling and application compared to the previous period.

(2) Between 1995 and 2003, in both *Gymnasium* and *Hauptschule* of Bavaria, applications of mathematics were listed very specifically, though the emphases are quite different. *Hauptschule* curriculum paid attention to recognizing the power and limits of mathematization of everyday phenomena, and specified that students need to carefully interpret statistics, and also know common financial concepts. *Gymnasium* placed mathematics as the fundamental subject, therefore listed many applications that can be done to virtually every other discipline. Though the difference of cognitive demand between *Arbitur* and *Quali* seems to be huge, both of them allow more time for students to think in terms of one context, since the questions are usually grouped together around the same real-world scenario. However, the *Quali* at that time mainly focused on usual word problems, and the *Arbitur* did not involve application except for its *stochastics* section.

After 2003, the "PISA shock" to Germany had a huge impact on their mathematics curriculum: many elements the PISA's assessment framework were actually written into the curriculum documents. Modeling was introduced systematically, with extended, standard examples of modeling tasks offered to all federal states. With the ideals of modeling, the tasks of *Quali* had been revised to include more mathematizations, and all sections of the *Arbitur* (*analysis*, *stochastics*, and *geometry*) have modeling contents.

(3) Hong Kong's curriculum documents from 1999 to 2003 did not mention modeling, but did

require “numeracy” as a basic skill for all pupils, so as to deal with the mathematics required by the general activities of the society. Other than that, not many applications were included. The public assessments in this period did not involve much about application either. The HKCEE, served as a high school exit examination, had less than 20% application problems, all of which are the standard mathematizaiton and applied mathematical tasks. That is, none of the important modeling steps like *interpret* and *verify* were touched. The HKALE, which was designed for college entrance, included the “Pure Maths” and the “Applied Maths”. However, the “Applied Maths” part is simply mechanics problems and probability problems. Admittedly, both mechanics and probability are good mathematical *models*, the modeling process was systematized already, giving no room for students to create, adjust, or revise the models. Hong Kong’s stance toward modeling did not change much from 2004 to 2012, as can be seen from the examination items.

Discussions

The findings described at the end of Chapter 5 can be seen as an application of the theoretical tool presented, but also contribute to the thesis *per se*. As we saw, *most* test items only tested little of modeling steps other than the calculations or writing a formula that represents the quantitative relationships. The assessment of posing real-world question, idealization, interpretation, verification, and re-iteration of the modeling process is rare, if exists. This furthers an earlier result (Germain-Williams, 2014) that certain textbooks, though claiming the importance of modeling, almost ignored the first and last steps of modeling. It is found that public assessments are even more limited: most test items that are supposed to test students’ knowledge of application involve only one or two steps of modeling.

It was also found how important it is to specify the assessment requirements in detail in the curriculum documents. After all, if the documents only mention the words of *contexts*, *applications*, or *modeling*, then the test items will also have them in a level that may not be deep enough. It can be suggested that, if modeling and application is to be included in the curriculum, one can follow the idea of Niss (2001). The consideration of three dimensions is not only helpful for designing modeling tasks, but also for introducing modeling into a curriculum. Those three dimensions with respect to modeling competence were introduced in Chapter 2 before, but it is convenient to state them again here:

- *Degree of coverage*: The parts of a modeling process that someone can (or should) work with, and the level of reflections involved;
- *Radius of action*: A collection of situations or contexts in which someone can (or should) perform mathematical modeling activities;
- *Technical level*: The kinds of mathematical knowledge that someone can (or should) use and how flexible they are in their use of such knowledge. (Blomhøj & Jensen 2007, p. 55; cf. 2)

The work done in Chapter 5 dealt solely with *degree of coverage*, and is exploratory in nature. The *technical level* dimension has been done extensively under the brand called *Applied Mathematics*. Consequently, more exploratory studies are needed for a collection and classification of *contexts*, so that all the three dimensions can be ready for the curriculum or assessment design. This suggests some possible further studies. First, since the current work is exploratory, it is possible set up some confirmatory studies as extensions: (i) show that the classification by modeling spectrum is statistically reliable, meaning that it can be operationally defined, as well as replicable. (ii) When the context of a task is found *extraneous* (cf. 5, MCAS12 items #34, #22, and #42), the test taker either skips it or is distracted by it; (iii) If a test is not creative enough, interpretation steps may be reduced to the understanding of a model, if some group of student had learned a direct translation of the interpreting question (cf. 5, HKALE02, #2-11). Much more generally, it is to say that the actual modeling steps involved can be coupled with the prior knowledge that a student has.

Moreover, some more creative work should be done on the missing modeling steps of the items. In particular, there is a call for as many items as possible that assess the following modeling steps —- Step [1]: posing a relevant real-world question that can possibly solved by mathematical modeling; Step [3]: making idealizations of the conceptual model, including assumptions or simplifications, so that the mathematics involved is within a comfortable level; Step [7]: checking if the mathematical results make sense when translated back to the real-world context, and if not, criticize one or more modeling steps that were done before the calculations were done; Step [8]: knowing ways to change an unsatisfactory model by making alternative assumptions, choosing different kinds of mathematics to work with, or revising the conceptual model. Inspired by many of the curriculum materials, one possible project to create more modeling problems is to start with a real modeling process, making simplifications or providing scaffoldings, so that the task is suitable for the level of the students.

Finale

Mathematics had been treated as "mental gymnastics" that trains human's mind. As Stanic (1986) describes, mathematics has been viewed by 19th century educators as a subject that has "extraordinary potential for improving certain faculties of the mind, especially that of reason." (p. 41) *Faculties of the mind* is a construct used in the mental discipline theory, which is generally agreed to be a historical, 'outdated' term, therefore it is no longer valid, in the academia, to say that mathematics is a gymnastics of the mind. However, deep down in the heart is the belief that a mind trained in mathematics is so powerful that it may (attempt) to attack problems in many fields in the "outside world". How do the failure of the theory and the intuition reconcile? Perhaps this is a way out: Yes, the school subject of mathematics is still trains the mind; it is not the kind of gymnastics that seems unrelated to other sports, but rather a *gymnasium* that allows various activities to be done in. The following analogy can be treated as a hypothesis that can be tested: a person who plays one kind of sport everyday may not be good at any other ones, but if he is generally well in many kinds of sports, then it is likely that he can take on one more pretty quickly. Similarly, if one only does mathematical problems in, say, indoor design, he may not be good at modeling other stuff. However, if he is exposed to various lively applications in the school already, modeling is no longer alien to him, and he can start to model any situation with mathematics. In this ideal sense, his mental faculty is really trained successfully.

Bibliography

Adamson, B. & Morris, P. (2014). Comparing curricula. In M. M. Mark Bray Bob Adamson (Ed.), *Comparative education research* (Chap. 11, pp. 309–332). Springer.

Antonius, S. (2007). Modelling based project examination. In W. Blum, P. L. Galbraith, H.-W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education* (pp. 409–416). New ICMI Study Series. Springer US.

Ärlebäck, J. B. (2010). *Mathematical modelling in upper secondary mathematics education in sweden* (Thesis, Linköping University).

Assude, T., Boero, P., Herbst, P., Lerman, S., & Radford, L. (2008). The notions and roles of theory in mathematics education research. In *10th international congress on mathematical education, monterrey, mexico, july* (pp. 6–13).

Bernstein, B. (1990). *Class, codes and control: Applied studies towards a sociology of language*. Rouledge.

Bernstein, B. (2003). *Class, codes and control: Applied studies towards a sociology of language*. Psychology Press.

Berry, J. & Le Masurier, D. (1984). Open university students do it by themselves. *Teaching and applying mathematical modelling*, 48–85.

Bikner-Ahsbahs, A. & Prediger, S. (2010). Networking of theories—an approach for exploiting the diversity of theoretical approaches. In *Theories of mathematics education* (pp. 483–506). Springer.

Bikner-Ahsbahs, A. & Prediger, S. (2014). *Networking of theories as a research practice in mathematics education*. Springer.

Bikner-Ahsbahs, A., Prediger, S., Artigue, M., Arzarello, F., Bosch, M., Dreyfus, T., … Kidron, I. (2014). Starting points for dealing with the diversity of theories. In *Networking of theories as a research practice in mathematics education* (pp. 3–12). Springer.

Blomhoj, M. & Højgaard Jensen, T. (2003). Developing mathematical modelling competence: Conceptual clarification and educational planning. *Teaching mathematics and its applications*, *22*(3), 123–139.

Blomhøj, M. & Jensen, T. H. (2007). What's all the fuss about competencies? In W. Blum, P. L. Galbraith, H. W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education* (pp. 45–56). New York: Springer.

Blum, W. (1995). Applications and modelling in mathematics teaching and mathematics education: Some important aspects of practice and of research. In C. Sloyer, W. Blum, & I. Huntley (Eds.), *Advances and perspectives in the teaching of mathematical modeling and applications* (pp. 1–20). Yorklyn, DE: Water Street Mathematics.

Blum, W., Galbraith, P. L., Henn, H.-W., & Niss, M. (2007). *Modelling and applications in mathematics education.* Springer New York.

Blum, W. & Kaiser, G. (1997). *Vergleichende empirische untersuchungen zu mathematischen anwendungsfähigkeiten von englischen und deutschen lernenden.* Unpublished manuscript.

Blum, W. & Leiß, D. (2007). How do students and teachers deal with modelling problems? In C. Haines, P. Galbraith, W. Blum, & S. Khan (Eds.), *Mathematical modelling (ICTMA 12): Education, engineering and economics* (pp. 222–231). Chichester: Horwood.

Blum, W. & Niss, M. (1991). Applied mathematical problem solving, modelling, applications, and links to other subjects—state, trends and issues in mathematics instruction. *Educational Studies in Mathematics*, *22*(1), 37–68.

Borromeo Ferri, R. (2007). Modelling problems from a cognitive perspective. *Mathematical Modelling (ICTMA 12): Education, Engineering and Economics, Chichester, Horwood Publishing*, 260–270.

Bray, M. [Mark], Adamson, B., & Mason, M. (Eds.). (2014). *Comparative education research.* Springer.

Bray, M. [Mark] & Thomas, R. M. (1995). Levels of comparison in educational studies: Different insights from different literatures and the value of multilevel analyses. *Harvard educational review*, *65*(3), 472–491.

Brown, S. I. (2001). *Reconstructing school mathematics: Problems with problems and the real world.* Peter Lang Pub Incorporated.

Bryman, A. (2012). *Social research methods (4th ed.)* Oxford University Press.

Burghes, D. N. (1984). Prologue, in teaching and applying mathematical modeling. In J. S. Berry, D. N. Burghes, I. D. Huntley, D. J. G. James, & A. O. Moscardini (Eds.), *Teaching and applying mathematical modeling* (pp. xi–xiv). Ellis Horwood.

Burkhardt, H. (2001). World class assessment: Principles, practice and problem solving. In C. Richardson (Ed.), *Assessing gifted and talented children*. London: Qualifications and Curriculum Authority.

Burkhardt, H. (2006). Modelling in mathematics classrooms: Reflections on past developments and the future. *ZDM*, *38*(2), 178–195.

Campbell, S. R. (2006). Educational neuroscience: New horizons for research in mathematics education. *Online Submission*.

Cockcroft, W. H. (1982). Report of the committee of inquiry into the teaching of mathematics in schools. London: HMSO.

Cook, D. J., Mulrow, C. D., & Haynes, R. B. (1997). Systematic reviews: Synthesis of best evidence for clinical decisions. *Annals of internal medicine*, *126*(5), 376–380.

Damlamian, A. & Sträßer, R. (2009). Icmi study 20: Educational interfaces between mathematics and industry. *ZDM*, *41*(4), 525–533.

Depaepe, F., De Corte, E., & Verschaffel, L. (2014). Students' non-realistic mathematical modeling as a drawback of teachers' beliefs about and approaches to word problem solving. In B. Pepin & B. Roesken-Winter (Eds.), (pp. 137–156). Springer.

Devlin, K. (1999). Supermarket math. american association of america online.

Dooley, W. H. (1915). *Vocational mathematics*. DC Heath & Company.

Drakes, C. I. (2012). *Mathematical modelling: From novice to expert* (Doctoral dissertation, Simon Fraser University).

Eisenhart, M. (1991). Conceptual frameworks for research circa 1991: Ideas from a cultural anthropologist; implications for mathematics education researchers. In R. G. Underhill (Ed.), *Proceedings of the thirteen annual meeting, north american chapter of the international group for the psychology of mathematics education, (blacksburg, virginia, october 16-19, 1991)* (pp. 202–211).

Ekol, G. (2011). Understanding and promoting mathematical modelling competencies: An applied perspective. In *Trends in teaching and learning of mathematical modelling* (pp. 57–64). Springer.

English, L. D. (2010). Young children's early modelling with data. *Mathematics Education Research Journal*, *22*, 24–47. doi:10.1007/BF03217564

Ernest, P. (2010). Reflections on theories of learning. In *Theories of mathematics education* (pp. 39–47). Springer.

Evans, J. (2002). *Adults' mathematical thinking and emotions: A study of numerate practice*. Routledge.

Ferri, R. B. (2011). Effective mathematical modelling without blockages: A commentary. In G. Kaiser, W. Blum, R. B. Ferri, & G. Stillman (Eds.), *Trends in teaching and learning of mathematical modelling* (pp. 181–185). International Perspectives on the Teaching and Learning of Mathematical Modelling. Springer Netherlands.

Fisher, I. (1930). The application of mathematics to the social sciences. *Bulletin of the American Mathematical Society*, *36*(4), 225–243.

Flinders, D. J., Noddings, N., & Thornton, S. J. (1986). The null curriculum: Its theoretical basis and practical implications. *Curriculum Inquiry*, *16*(1), 33–42.

Florida Department of Education. (2012). Florida 2012 geometry end-of-course assessment test item specifications version 2. Tallahassee: Florida Department of Education.

Frejd, P. (2013). Modes of modelling assessment—a literature review. *Educational Studies in Mathematics*, *84*(3), 413–438.

Freudenthal, H. (1968). Why to teach mathematics so as to be useful. *Educational studies in mathematics*, *1*(1), 3–8.

Galbraith, P. & Clatworthy, N. (1990). Beyond standard models—meeting the challenge of modelling. *Educational Studies in Mathematics*, *21*(2), 137–163.

Garfunkel, S. A. & Pollak, H. (1998). *Mathematics: Modeling our world*. WH Freedman.

Germain-Williams, T. (2014). *Mathematical modeling in algebra textbooks at the onset of the common core state standards* (Doctoral dissertation, Columbia University). Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2016-03-09. Retrieved from http://eduproxy.tc-library.org/?url=/docview/1537949683?accountid=14258

Gerofsky, S. G. (1997). A conjecture on the history of mathematical word problems:were word problems ever practical? In Y. M. Pothier (Ed.), *Canadian mathematics education study group proceedings of the annual meeting (lakehead university, thunder bay, ontario, canada, may 23-27, 1997)* (Chap. Ad Hoc Session 6). Halifax, Nova Scotia: Mount St. Vincent University Press.

Gerofsky, S. G. (2001). *The word problem as genre in mathematics education*. National Library of Canada= Bibliothèque nationale du Canada.

Goodlad, J. I. & Richter, M. N. (1966). The development of a conceptual system for dealing with problems of curriculum and instruction. ERIC. Journal Article. ERIC.

Gould, H. (2013). *Teachers' conceptions of mathematical modeling* (Thesis, Columbia University).

Gravemeijer, K. & Doorman, M. (1999). Context problems in realistic mathematics education: A calculus course as an example. *Educational studies in mathematics*, *39*(1-3), 111–129.

Greer, B. (1997). Modelling reality in mathematics classrooms: The case of word problems. *Learning and instruction*, *7*(4), 293–307.

Greer, B. & Verschaffel, L. (2007). Modelling competencies—overview. In W. Blum, P. L. Galbraith, H. W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education* (pp. 219–224). New York: Springer.

Haines, C., Crouch, R., & Davis, J. (2000). *Mathematical modelling skills: A research instrument* (Report No. 55). University of Hertfordshire, Department of Mathematics.

Haines, C. & Crouch, R. (2007). Mathematical modelling and applications: Ability and competence frameworks. In W. Blum, P. L. Galbraith, H.-W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education* (pp. 417–424). New ICMI Study Series. Springer US.

Haines, C. & Houston, K. (2001). Assessing student project work. In D. Holton (Ed.), *The teaching and learning of mathematics at university level: An icmi study* (Vol. 7, pp. 431–442). Springer Science & Business Media.

Haines, C. & Izard, J. (1995). Assessment in context for mathematical modelling. In C. Sloyer, W. Blum, & I. Huntley (Eds.), *Advances and perspectives in the teaching of mathematical modelling and applications* (pp. 131–150). Yorklyn, DE: Water Street Mathematics.

Haines, C. R. & Crouch, R. (2010). Remarks on a modeling cycle and interpreting behaviours. In P. L. Galbraith, C. R. Haines, A. Hurford, & R. A. Lesh (Eds.), *Modeling students' mathematical modeling competencies* (pp. 145–154). Springer.

Hall, G. (1984). The assessment of modelling projects. *Teaching and applying mathematical modelling*, 143–148.

Henning, H. & Keune, M. (2006). Levels of modeling competence. In M. Bosch (Ed.), *Cerme 4. proceedings of the fourth congress on the european society for research in mathematics education* (pp. 1666–1674). European Society for Research in Mathematics Education.

Højgaard Jensen, T. (2007). Assessing mathematical modelling competency. In C. Haines, P. Galbraith, W. Blum, & S. Khan (Eds.), *Mathematical modeling (ICTMA 12): Education, engineering and economics* (pp. 141–148). Chichester: Ellis Horwood.

Houston, K. (2007). Assessing the "phases" of mathematical modelling. In W. Blum, P. L. Galbraith, H. W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education* (pp. 249–256). New York: Springer.

Inoue, N. (2005). The realistic reasons behind unrealistic solutions: The role of interpretive activity in word problem solving. *Learning and Instruction*.

Jurdak, M. (2016). *Learning and teaching real world problem solving in school mathematics*. Springer.

Kaiser, G. (2007). Modelling and modelling competencies in school. In C. Haines, P. Galbraith, W. Blum, & S. Khan (Eds.), *Mathematical modelling (ICTMA 12): Education, engineering and economics* (pp. 110–119). Chichester: Ellis Horwood.

Kaiser, G., Blum, W., Ferri, R. B., & Stillman, G. (2011). *Trends in teaching and learning of mathematical modelling: ICTMA14*. Springer Science & Business Media.

Kaiser, G. & Brand, S. (2015). Modelling competencies: Past development and further perspectives. In G. A. Stillman, W. Blum, & M. S. Biembengut (Eds.), *Mathematical modelling in education research and practice* (pp. 129–149). Springer.

Kaiser-Messmer, G. (1986). *Anwendungen im mathematikunterricht. 2. empirische untersuchungen*. Bad Salzdetfurth: Franzbecker.

Kilpatrick, J. (2010). Preface to part i. In B. Sriraman & L. English (Eds.), *Theories of mathematics education* (pp. 3–6). Springer.

Klamkin, M. S. (1968). On the teaching of mathematics so as to be useful. *Educational Studies in Mathematics*, *1*(1), 126–160.

Klein, M. F., Tye, K. A., & Wright, J. E. (1979). A study of schooling: Curriculum. *The Phi Delta Kappan*, *61*(4), 244–248.

Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge University Press.

Lesh, R. A. & Lamon, S. J. (Eds.). (1992). *Assessment of authentic performance in school mathematics*. Rouledge.

Lesh, R. & Sriraman, B. (2010). Re-conceptualizing mathematics education as a design science. In *Theories of mathematics education* (pp. 123–146). Springer.

Lesh, R. & Yoon, C. (2007). What is distinctive in (our views about) models & modelling perspectives on mathematics problem solving, learning, and teaching? In W. Blum, P. L. Galbraith, H.-W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education* (pp. 161–170). Springer.

Lester, F. K. (2005). On the theoretical, conceptual, and philosophical foundations for research in mathematics education. *ZDM*, *37*(6), 457–467.

Lester, F. K. (2010). On the theoretical, conceptual, and philosophical foundations for research in mathematics education. *Theories of mathematics education*, 67–85.

Lingefjärd, T. & Holmquist, M. (2005). To assess students' attitudes, skills and competencies in mathematical modeling. *Teaching Mathematics Applications*, *24*(2-3), 123–133. doi:https://doi.org/10.1093/teamat/hri021

Maaß, K. (2006). What are modelling competencies? *ZDM*, *38*(2), 113–142.

Maaß, K. (2007). Modelling in class: What do we want the students to learn? In P. Galbraith, W. Blum, & S. Khan (Eds.), *Mathematical modelling: Education, engineering and economics (ICTMA 12)* (pp. 63–78). Woodhead Publishing. doi:http://dx.doi.org/10.1533/9780857099419.2.63

Maasz, J. & O'Donoghue, J. (2011). *Real-world problems for secondary school mathematics students*. Springer Science & Business Media.

Maddern, S. & Crust, R. (1989). *Applications: Where there's life, there's maths*. Extended Tasks for GCSE Mathematics. Shell Centre for Mathematical Education.

Marsh, C. J. (2009). *Key concepts for understanding curriculum*. Routledge.

Mason, J. & Waywood, A. (1996). The role of theory in mathematics education and research. In *International handbook of mathematics education* (pp. 1055–1089). Springer.

Matos, J. & Carreira, S. (1995). Cognitive processes and representations involved in applied problem solving. *Advances and perspectives in the teaching of mathematical modelling and applications*, 71–82.

McClelland, D. C. (1973). Testing for competence rather than for" intelligence." *American psychologist*, *28*(1), 1.

Miles, M. B. & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook.* SAGE.

Murphy, J. & Adams, J. E. J. (1998). Reforming america's schools 1980-2000. *Journal of Educational Administration*, *36*(5), 426–444.

Nesher, P. (2015). On the diversity and multiplicity of theories in mathematics education. In *Pursuing excellence in mathematics education* (pp. 137–148). Springer.

New, P., Britton, S., Sharma, M., & Brew, A. (2012). Researching the transferability of mathematical skills. In *Proceedings of the australian conference on science and mathematics education (formerly uniserve science conference)* (Vol. 7).

Niss, M. (1993). Assessment of modelling applications and modelling in mathematics teaching. In J. de Lange (Ed.), *Innovation in maths education by modelling and applications* (Chap. 4, pp. 41–51). Ellis Horwood.

Niss, M. (2003). Mathematical competencies and the learning of mathematics: The danish kom project. In *3rd mediterranean conference on mathematical education* (pp. 115–124).

Niss, M. (2007). The concept and role of theory in mathematics education. *Relating practice and research in mathematics education. Proceedings of Norma*, *5*, 97–110.

Niss, M. (2013). Modeling a crucial aspect of students' mathematical modeling. In R. Lesh, P. L. Galbraith, C. R. Haines, & A. Hurford (Eds.), *Modeling students' mathematical modeling competencies* (pp. 43–59). International Perspectives on the Teaching and Learning of Mathematical Modelling. Springer Netherlands.

Niss, M. (2015a). Mathematical competencies and pisa. In *Assessing mathematical literacy* (pp. 35–55). Springer.

Niss, M. (2015b). Prescriptive modelling–challenges and opportunities. In *Mathematical modelling in education research and practice* (pp. 67–79). Springer.

Niss, M. & Højgaard Jensen, T. (2011). Competencies and mathematical learning. *Ideas and inspiration for the development of mathematics teaching and learning in Denmark.*

Nunes, T., Schliemann, A. D., & Carraher, D. W. (1993). *Street mathematics and school mathematics.* Cambridge University Press.

Organisation for Economic Co-operation and Development. (2004). *The pisa 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills.* OECD Publishing.

Organisation for Economic Co-operation and Development. (2006). *Pisa released items - mathematics*. Copyright 2006 by OECD. OECD Publishing. Retrieved from http://www.oecd.org/pisa/38709418.pdf

Organisation for Economic Co-operation and Development. (2011). Germany: Once weak international standing prompts strong nationwide reforms for rapid improvement. In *Successful reformers in education: Lessons from pisa for the united states*. Organisation for Economic Co-operation and Development.

Organisation for Economic Co-operation and Development. (2013a). *Pisa 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. OECD Publishing. doi:10.1787/9789264190511-en

Organisation for Economic Co-operation and Development. (2013b). *Pisa 2012 released mathematics items*. Copyright 2013 by OECD. OECD Publishing. Retrieved from https://www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf

Pollak, H. (1969). How can we teach applications of mathematics? *Educational studies in mathematics*, *2*(2), 393–404.

Pollak, H. (1997). *Solving problems in the real world*. Why Nymbers Count: Quantitative Literacy for Tomorrow's America, New York: The College Board.

Pollak, H. (2003). A history of the teaching of modeling. In G. M. A. Stanic & J. Kilpatrick (Eds.), *A history of school mathematics* (Vol. 1, pp. 647–672). National Council of Teachers of Mathematics.

Popper, K. R. (1972). *Objective knowledge: An evolutionary approach*. Clarendon Press.

Potgieter, M., Harding, A., & Engelbrecht, J. (2008). Transfer of algebraic and graphical thinking between mathematics and chemistry. *Journal of Research in Science Teaching*, *45*(2), 197–218.

Prediger, S., Arzarello, F., Bosch, M., & Lenfant, A. (2008). Comparing, combining, coordinating-networking strategies for connecting theoretical approaches: Editorial for zdm-issue 39 (2008) 2. *ZDM*, *40*(2), 163–164.

Prediger, S., Bikner-Ahsbahs, A., & Arzarello, F. (2008). Networking strategies and methods for connecting theoretical approaches: First steps towards a conceptual framework. *ZDM*, *40*(2), 165–178.

Punch, K. (2009). *Introduction to research methods in education*. SAGE Publications. Retrieved from https://books.google.com/books?id=lBvMqiaN5EgC

Quinn, M. & Kridel, C. (2010). Null curriculum. *Encyclopedia of curriculum studies*, *2*, 613–614.

Reusser, K. & Stebler, R. (1997). Every word problem has a solution—the social rationality of mathematical modeling in schools. *Learning and instruction*, *7*(4), 309–327.

Reuters, T. (2013). Endnote X7 (7.5 bld 9325) [software]. *Thomson Reuters: Philadelphia, PA.*

Roorda, G., Vos, P., & Goedhart, M. (2007). The concept of the derivative in modelling and applications. In P. Galbraith, W. Blum, & S. Khan (Eds.), *Mathematical modelling: Education, engineering and economics (ICTMA 12)* (pp. 288–293). Woodhead Publishing. doi:http://dx.doi.org/10.1533/9780857099419.5.288

Roth, W.-M. (2014). Rules of bending, bending the rules: The geometry of electrical conduit bending in college and workplace. *Educational Studies in Mathematics*, *86*(2), 177–192.

Schoenfeld, A. H. (1991). On mathematics as sense-making: An informal attack on the unfortunate divorce of formal and informal mathematics. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education* (Chap. 16, p. 311). Routledge.

Schoenfeld, A. H. (2000). Purposes and methods of research in mathematics education. *Notices of the AMS*, *47*(6), 641–649.

Schoenfeld, A. H. (2001). Purposes and methods of research in mathematics education. *The Teaching and Learning of Mathematics at University Level*, 221–236.

Schubert, W. H. (2008). Curriculum inquiry. In *The sage handbook of curriculum and instruction*. SAGE Publication.

Shavelson, R. J. (2010). On the measurement of competency. *Empirical research in vocational education and training*, *2*(1), 41–63.

Sriraman, B. & English, L. (2010a). Surveying theories and philosophies of mathematics education. In *Theories of mathematics education* (pp. 7–32). Springer.

Sriraman, B. & English, L. (2010b). *Theories of mathematics education: Seeking new frontiers.* Springer Science & Business Media.

Sriraman, B. & Kaiser, G. (2006). Theory usage and theoretical trends in europe: A survey and preliminary analysis of cerme4 research reports. *Zentralblatt für Didaktik der Mathematik*, *38*(1), 22–51.

Stacey, K. & Turner, R. (2015). The evolution and key concepts of the pisa mathematics frameworks. In *Assessing mathematical literacy* (pp. 5–33). Springer.

Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. (2005). *Beschlüsse der kultusministerkonferenz: Bildungsst andards im fach mathematik für den hauptschulabschluss (jahrgangsstufe 9).* München, Neuwied: Luchterhand, Wolters Kluwer.

Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland Sekretariat & Institut zur Qualitätsentwicklung im Bildungswesen. (2015). *Bildungsstandards im fach mathematik für die allgemeine hochschulreife (beschluss der kultusministerkonferenz vom 18.10.2012).* Copyright 2015 by KMK Bonn und Berlin 2012. Köln: Wolters Kluwer.

Steen, L. A. (1999). Theories that gyre and gimble in the wabe. JSTOR.

Steiner, H.-G. (1985). Theory of mathematics education (tme): An introduction. *For the learning of mathematics*, *5*(2), 11–17.

Stillman, G. A. & Galbraith, P. L. [Peter L]. (1998). Applying mathematics with real world connections: Metacognitive characteristics of secondary students. *Educational Studies in Mathematics*, *36*(2), 157–194.

Stolz, M. (2002). The history of applied mathematics and the history of society. *Synthese*, *133*(1), 43–57.

Sträßer, R. & Thiering, J. (1986). Action group 7: Mathematics in adult, technical and vocational education. In M. Carss (Ed.), *Proceedings of the fifth international congress on mathematical education* (pp. 124–132). Boston, MA: Birkhäuser Boston. doi:10.1007/978-1-4757-4238-1_10

Tang, K. & Bray, M. (2000). Colonial models and the evolution of education systems centralization and decentralization in hong kong and macau. *Journal of Educational Administration.*

Tanner, H. & Jones, S. (1994). Using peer and self-assessment to develop modelling skills with students aged 11 to 16: A socio-constructive view. *Educational Studies in Mathematics*, *27*(4), 413–431.

Tanner, H. & Jones, S. (1995). Developing metacognitive skills in mathematical modelling–a socio-contructivist interpretation. *Advances and perspectives in the teaching of mathematical modelling and applications*, 61–70.

Teppo, A. R. (2015). Grounded theory methods. In *Approaches to qualitative research in mathematics education* (pp. 3–21). Springer.

Tian, X. (2014). *Mathematical modeling in the people's republic of china—indicators of participation and performance on comap's modeling contest* (Thesis, Columbia University).

Treilibs, V. (1979). *Formulation processes im mathematical modelling* (M. Phil. Thesis, University of Nottingham).

Tsatsaroni, A., Lerman, S., & Xu, G.-R. (2003). A sociological description of changes in the intellectual field of mathematics education research: Implications for the identities of academics. In *The annual meeting of the american educational research association (chicago, il, april 21-25, 2003).* ERIC.

Van den Akker, J. (2004). Curriculum perspectives: An introduction. In *Curriculum landscapes and trends* (pp. 1–10). Springer.

Van den Heuvel-Panhuizen, M. (2003). The didactical use of models in realistic mathematics education: An example from a longitudinal trajectory on percentage. *Educational studies in Mathematics*, *54*(1), 9–35.

Van den Heuvel-Panhuizen, M. & Drijvers, P. (2014). Realistic mathematics education. In *Encyclopedia of mathematics education* (pp. 521–525). Springer.

Van Hecke, T. (2011). Release the prisoners game. *Teaching Mathematics & its Applications*, *30*(1), 37–42. Retrieved from http://ezproxy.cul.columbia.edu/login?qurl=http%3a%2f%2fsearch.ebscohost.com%2flogin.aspx%3fdirect%3dtrue%26db%3dehh%26AN%3d58768801%26site%3dehost-live%26scope%3dsite

Verschaffel, L., de Corte, E., & Lasure, S. (1994). Realistic considerations in mathematical modeling of school arithmetic word problems. *Learning and Instruction*, *4*(4), 273–294.

Verschaffel, L., Greer, B., & de Corte, E. (2000). *Making sense of word problems.* CRC Press.

Vos, P. (2007). Assessment of applied mathematics and modelling: Using a laboratory-like environment. In W. Blum, P. L. Galbraith, H.-W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education* (pp. 441–448). New ICMI Study Series. Springer US.

Wake, G. & Hardy, G. (2007). Mathematics in the physical sciences: Multiple perspectives. In P. Galbraith, W. Blum, & S. Khan (Eds.), *Mathematical modelling: Education, engineering and economics (ICTMA 12)* (pp. 250–259). Woodhead Publishing. doi:http://dx.doi.org/10.1533/9780857099419.5.250

Wake, G. & Yasukawa, K. (2015). Mathematics education in and for work. In *The proceedings of the 12th international congress on mathematical education* (pp. 403–407). Springer.

Watt, H. M. (2005). Attitudes to the use of alternative assessment methods in mathematics: A study with secondary mathematics teachers in sydney, australia. *Educational studies in mathematics*, *58*(1), 21–44.

Westen, D. & Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. *Journal of personality and social psychology*, *84*(3), 608.

Wigner, E. P. (1960). The unreasonable effectiveness of mathematics in the natural sciences. richard courant lecture in mathematical sciences delivered at new york university, may 11, 1959. *Communications on pure and applied mathematics*, *13*(1), 1–14.

Williams, J. & Wake, G. (2007). Black boxes in workplace mathematics. *Educational Studies in Mathematics*, *64*(3), 317–343.

Appendix A. Dataset of Modeling Papers

A1. List of Keyword Searches

Referring to Chapter III, p. 66, the purpose of the first series of keyword searches is to narrow down the targeted articles by sorting out the T4 articles in advance: A article is categorized as T4 if it is not related to modeling and/or application. In the following, we present the steps of the keyword search. Each step specifies a subgroup that is restricted by keyword occurrences[11]. We mainly searched for keywords “model*” or “appl*”, but also did searches of a list of alternative keywords, including “vocation*”, “realistic”, “functional”, “transfer”, “literacy”, “word problem”, “real world”, “in work”, “for work”, and “workplace”[12]. The number of references that were found is also recorded.

1. All non-specialized publications (no restriction): 2,997
    - Title contains “model*” or “appl*”: 141
    - Title does not contain “model*” or “appl*” but Abstract does: 486
      Subtotal: 627
    - Not belong to the subtotal above, but contain alternative keywords: 144
    - All else: 2,220 (74.2 %)
2. All specialized publications (no restriction): 527
    - Title contains “model*” or “appl*”: 238
    - Title does not contain “model*” or “appl*” but Abstract does: 244
      Subtotal: 482
    - Not belong to the subtotal, but contain alternative keywords: 14
    - All else: 31 (5.9%)

[11]A technical note: In *EndNote*, we create Smart Groups that is able to perform a search of logical conjunction or disjunction, specifying whether a keyword should be found in a data entry such as “title” or “Abstract”. After that, a new Group Create From Groups is created so that the condition in the Smart Group can be applied to a predefined list of references.

[12]It is not preferred to search a single word such as “work” because it is expected to be overly inclusive.

"Model*" and "appl*" are the main keyword strings that are looked for, targeting the words "model", "modeling", "modelling", "modeler", "application", "apply", etc. Note that possible "false positives" would mostly occur with "model" or "apply", since "model" may not refer to a mathematical model, and the verb "apply" has many other meanings than using mathematical knowledge. Ideally, these two cases should be singled out for examination, but this was not done in the first analysis. Possible "false negatives" occur when neither "model*" nor "appl*" are used in the title and the abstract, but other words or phrases convey similar meaning. A list of alternative keywords were included to reduce false negatives.

The second series of searches is referred to when the articles are actually being classified as T1 or T2, see Chapter IV, p. 80. The following keyword searches are used for the decision of T1 and T2 codes, in decreasing order of priority. If an article's title and abstract do not contain "model*" or "appl*" but only alternative keywords, then it is not included in the keyword searches below.

1. All non-specialized publications related to modeling and/or application, not including alternative keywords: 627
    - Title or abstract contains T1 keywords, such as:
"assess": 65
"test" or "tests": 52
"difficult*": 61
"blockage": 0
"barrier": 0
"competen*": 44
"ability" of "abilities": 109
"skill*": 41
Not containing any T1 keywords: 362
    - Other articles that contain T2 keywords, such as
"teach*" or "learn*" or "develop*": 280
    - All else: 82
2. All specialized publications related to modeling and/or application, not including alternative keywords: 482
    - Title or abstract contains T1 keywords, such as:
"assess": 61
"test" or "tests": 35
"difficult*": 47
"blockage": 6
"barrier": 7
"competen*": 49
"ability" of "abilities": 73
"skill*": 49

- Not containing any T1 keywords: 262
- Other articles that contain T2 keywords, such as "teach*" or "learn*" or "develop*": 219
  All else: 43

A2. Code List for Article Classification

T1 is the most relevant type of article for the study of RQ1. The codes indicate a reason why an article is categorized as T1. There could be more than one reasons, but only one will be recorded as the code, in decreasing order of priority. T1 has three codes only. The codes and the related keywords are listed below, preceded by the definition of Type 1:

> Type 1, Ability to Apply: The purpose of the article is to study individual's ability to apply mathematical knowledge, and the testing of such ability.
> T1 test ~ [13]: The article proposes or studies a test on the ability to apply mathematical knowledge. Such test is a measurement or evaluation of the test takers' ability to apply mathematical knowledge. (Note: A general assessment activity, such as a set of problems only, will not be included in this category if it does not provide a clear procedure to infer about students' ability level.) Related keyword searches: "assess", "test" or "tests".
> T1 study ~ : The article assumes some form of ability to apply mathematical knowledge, and contains an empirical study of human subjects' possession of such ability. Related keyword searches: "difficult*", "blockage", "barrier", "competen*", "ability" of "abilities", "skill*".
> T1 describe ~ : The article contains a theoretical description of the ability to apply mathematical knowledge (or part of such ability). The article does not focus on empirical study, which may refute or support the description. Related keyword searches: "competen*", "ability" of "abilities", "skill*".

Some articles do not directly address the ability to apply mathematical knowledge or the testing of such ability, but may still contribute to the topic with relevant claims and evidences. Type 2 articles serve as a source of such possibility. In the study of the teaching and learning about application, we may see what teachers and students actually think about application, how they actually apply their mathematical knowledge, and the difficulties they are facing to. The study of these issues can imply some understanding of the ability to apply mathematical knowledge. The codes in T2 are:

> Type 2, Teaching and Learning: The article is not classified as Type 1, but it studies the teaching and learning about application.

[13] The symbol, "~", stands for the ability to apply mathematical knowledge.

T2 assessment analysis: The article analyzes an existing assessment of mathematics that may contain modeling and/or application, or offers some principles of such analysis.
T2 assessment design: The article discusses specific design, or principles of designing an assessment of mathematics that may contain modeling and/or application.
T2 conception of modeling: The article studies how individuals or groups think about modeling related concepts.
T2 course development: The article describes a modeling course's development, or the principles related to such development.
T2 curriculum development: The article describes the development of curriculum materials on modeling and/or application, or the principles related to such development.
T2 develop ~ : The article describes how the ability to apply mathematical knowledge can be developed for individuals or groups.
T2 difficulty: The article studies how individuals find it difficult to solve modeling and/or applications, but does not posits an "ability" as in "T1 study ~".
T2 example: The article presents examples of a modeling problem that could be used in classes.
T2 learning process: The article analyzes the process of learning modeling and/or application.
T2 modeling process: The article analyzes the process of doing modeling and/or application.
T2 nature of modeling: The article analyzes the nature of mathematical modeling and related activities.
T2 pedagogy: The article presents how modeling can be taught.
T2 teacher development: The article describes the development of teachers on teaching modeling and/or application.
T2 teacher evaluation: The article analyzes teachers' ability to teach modeling and/or application.
T2 teacher knowledge: The article analyzes the related knowledge that a teacher has in order to teach modeling and/or application.
T2 teaching process: The article analyzes the process of teaching modeling and/or application.
T2 tool: The article presents tools, software, or other aids that helps doing, teaching, or learning modeling and/or application.

Many articles use, study, and discuss mathematical applications, but do not direct readers to the issue of individual's ability to apply. Thus, Type 3 articles will not be included in the analysis step, even though there are some valuable discussions on, for example, how modeling helps the teaching of regular topics in mathematics (T3: modeling as vehicle). The codes in T3 are:

Type 3, Application: The article is neither classified as type 1 or 2; it still studies application of mathematics.
T3 commentary: The article is a commentary on other articles related to modeling and/or application, but not presenting new constructs.
T3 curriculum study: The article studies some aspects of modeling, such as its status, in school curricula.
T3 general: The article is merely a general description of how modeling and/or application is related to mathematics education.
T3 general competence: The article describes a competence much more general than the ability to apply.
T3 general promotion: The article is a general promotion on how and why modeling is important in education.
T3 government decision: The article analyzes how government or other stakeholders decide on the content of curriculum that relates to modeling and/or application.
T3 modeling as vehicle: Modeling is not the objective for the article, but a way to achieve other objectives.

T3 modeling for education research: The article uses mathematical modeling to study topics in mathematics education, or education research in general.
T3 non education: The article is not directly related to education.
T3 overview: The article overviews things about modeling and its education.
T3 problematization: The article raised problems on the current practices or theories about the ability to apply.
T3 social development: The article emphasizes the social function of modeling education.

The first analysis showed that some articles showed an instance of a nice mathematical model, but did not specify how such model could help student to develop their ability to apply (T3: non education). Some other articles addresses mathematical modeling and application as a general introduction or promotion (T3: general).

Finally, articles that are simply not related to the teaching and learning of application will be classified as type 4. Some of them use the word "model" and thus are included through the keyword searches, but it is well-known that the word "model" may not refer to a mathematical one (T4: model not mathematical). Another main source of false positives come from the journal Teaching Mathematics & its Applications, with many articles that are only related to teaching and learning mathematics in general, but nothing on application. They passed the keyword search simply because all articles from this journal contains a common paragraph at the end that has the exact phrase "Teaching Mathematics & its Applications".

An important exception is T4: non-English, which may contain relevant articles but are excluded simply because they are not in English. The codes in T4 are:

Type 4, Non-application: none of Type1, 2, or 3.
T4 editorial: The reference does not refer to a full article, such as an editorial.
T4 model not mathematical: The article uses the word "model" but it does not refer to mathematical modeling.
T4 non ~ : The article is not at all related to the ability to apply mathematical knowledge.
T4 non English: The article is not in English.
T4 wrong ref: The reference isn't pointing to an article, or it points to a duplicated one.

A3: Selected theories

Seven theories (groups of theoretical elements) were selected according to their *falsifiability* (or *f*, operationally defined as the number of testable predictions) and popularity (or *p*, defined as the number of authors that had a publication on it). We include indices in the form of "i.j" to show that the theoretical elements came from the j-th theory of the i-th article; the articles are numbered within the 75 articles obtained from all T1 articles and the screened T2 articles (cf. p. 85). The scope of the theory is presented as letters A to F, referring to the framework specified in Chapter IV, p. 91. We also listed the main constructs and some testable claims of these theories.

Theory 1, "normative" vs individual modeling cycle process (f=7,p=3)

Theory Indices: [12.1, 16.1, 17.1, 35.1]
Scope: B. Modeling Process & C. Nature of Modeling competence
Main Constructs:

Modeling Cycle (Stages and Transitions)
Modeling Cycle Chart, Holistic vs Subprocesses (atomistic)
Continual referencing back to the real world
Modeling behavior descriptors: Ideal behavior and Individual route, Adequate Modeler (takes into account the relationship between the real world and the model), Inadequate modeler (thinks about the relationship and tries to model but lack of knowledge to have it done), Non-modeler (not taking into account the relationship; think from only one side)
Continuum of Expertise
Personal Factors

Some Testable Claims:

12.1(1) modeling process is mostly cyclic, though it may happen in reversed direction, and may involve multiple cycles
16.1(1) same modeling problem can be a higher level modeling stage for one education level, and a lower level modeling stage for a higher education level.
16.1(2) The (collective) improvement of modeling over education level is gradual, not sudden (variance higher) somewhat supported by 30.1
17.1(1) novice modelers take various "modeling routes" (more than experts do)
17.1(2) there is an ideal behavior, so as modelers become expert, the variety of modeling routes become much smaller
35.1(1) Some sub-competencies are for specific step, some others are needed throughout the whole process

35.1(2) Cognitive abilities (intelligence) or knowledge, motivational determinants, domain specific and cross-domain competences, are all the personal factors that affect modeling competence

Theory 2, Competence according to modeling step. (f=6,p=3)

Theory Numbers: [22.1, 32.3, 4.1, 19.1, 35.3]
Scope: C. Nature of Modeling competence & D. Measurement of Modeling competence
Main Constructs: Modeling Steps (Stages and Transitions), Modeling Competence, Three dimensions of assessment, Modeling Cycle as assessment framework
Some Testable Claims:

22.1(1) if one of the three dimensions is not well established (being able to activate the modeling steps, being able to understand the context, being able to proceed the mathematical techniques), then student is not able to carry out the modeling process well.
22.1(2) actual modeling process usually does not go along the ideal steps, but still need all the ste
4.1(1) A less specific definition of modeling competence hinders the development of it
19.1(1) any assessment for modeling can be fit into the 7 phases framework
35.3(1) Personal factors like reading competence (cross-domain competence), fluid intelligence (general ability), beliefs (motivation).
35.3(2) step specific factors – Step 1 (set up situation model) – Word comprehension belongs to so called crystallized intelligence (general ability),
Step 2 (set up a real model) – general knowledge.
Step 3 to 4 (set up a mathematical model and finding a solution) – mathematical competence (domain specific competence),
Step 5 (interpreting the mathematical solution) – mathematical competence and word comprehension,
Step 6 (validation) – general knowledge.

Theory 3, Meta-cognition. (f=6,p=3)

Theory Numbers: [11.1, 11.2, 32.2, 40.1]
Scope: F. More general attributes and their development
Main Constructs:

(1)Meta-cognition of one's own knowledge, control or self-regulation, beliefs and intuition
(2)Declarative metacognition, Procedural metacognition, Motivational metacognition (Sjuts, 2003, p18),
(3)Model of metacognition and cognitive monitoring: (a) metacognitive knowledge, (b) metacognitive experiences, (c) goals (or tasks), (d) actions (or strategies) (Flavell, 1979)
Meta-meta-cognitive process; Productive metacognitive acts; Red Flag

Some Testable Claims:

11.2(1) (2.2) If meta-cognitive strategies are taught in earlier ages, modeling development will be easier later
11.2(2) (2.3) A teacher needs to have meta-meta-thinking in order to successfully teach meta-cognitive strategies
11.2(3) (2.4) How pupils work together on modeling tasks is a great factor of the developing meta-cognitive skills, including task analysis, planning, Monitoring, checking, and reflections
40.1(1) if one does not have good meta-cognitive skills, it is not likely that he/she performs well in modeling
40.1(2) some meta-cognitive activities are not productive
40.1(3) if a teacher successfully teaches meta-cognition, a "meta-meta-cognitive" activity must be in place

Theory 4, Early experiences of data modeling (f=6, p=2)

Theory Numbers: [10.1, 29.1]
Scope: B. Modeling Process & E. Development of Modeling competence
Main Constructs:

Model-eliciting activities
Data modeling (as an example of MEA)
Creation, analysis, and revision of data classification models (Lehrer & Schauble, 2000);
selecting attributes and classifying items according to these attributes;
decide what is worthy of attention (Hanner, James, & Rohlfing, 2002).
Structuring and displaying data – creating their own form of inscription;
overlooking important information;
including redundant information

Some Testable Claims:

10.1(1) It is possible that young children (before the 3rd grade) to learn data modeling (including: developing their own data models, inscriptions, and sense-making systems to deal with complex situations), just as they learn other subjects
10.1(2) If a child is lack of or has low Inscriptional capacities, then he cannot do well on mathematical activities.
10.1(3) If a child has better meta-representational knowledge, then his mathematical thinking and communication skills will also be better.
29.1(1) if the learning does not engage the process "express $\rightarrow$ test $\rightarrow$ revise (reject)" the current ways of thinking, it is not possible to have significant conceptual changes
29.1(2) Activities that are developed according to MEA principles and with models that are involve mathematically significant concepts, then model development tends to involve significant forms of concept development
29.1(3) achievers in MEA activities can often be students who have been classified average or below-average under traditional tests

Theory 5, Modeling competence description (p=2,f=1)

Theory Numbers: [23.1, 23.2, 24.1, 31.1, 31.4, 32.1]
Scope: Nature of Modeling competence
Main Constructs: Kaiser and Maass' Modeling Sub-competencies framework
Some Testable Claims:

> 31.1(1) if one is consider not competent in modeling, he/she must be lack of one of the subcompetency

Theory 6, CHAT and situated cognition (f=4,p=1)

Theory Numbers: [43.1, 43.5]
Scope: A. Nature of Modeling and Application, C. Nature of Modeling competence, & E. Development of Modeling competence
Main Constructs:

> Situated Cognition
> Action mediated by instruments (artifacts, tools, concepts, language genres); distinction between physical and semiotic tools
> Community, division of labour, rules

Some testable claims:

> 43.1(1) (1.3) If one learned some mathematical knowledge in a specific context, he does better in a similar context than in a very different context
> 43.5(1) Other things being equal, the more a final assessment emphasizes application, a higher percentage is the students who are able to apply
> 43.5(2) If proportions are not taught or applied in an activity within a specific context, students cannot make sense of it.
> 43.5(3) If the boundary between mathematics and other subjects is blurred, students will understand more about how mathematics is applied"

Theory 7, Project examination (f=3,p=1)

Theory Numbers: [1.1]
Scope: D. Measurement of Modeling competence
Main Constructs: Traditional written examination, oral examination, project examination, oral defense
Some Testable Claims:

> 1.1(1) Effort explains much variations of project performance
> 1.1(2) A 30-min presentation is enough to determine whether a student cheated; a 10-min presentation is not sufficient
> 1.1(3) A students' modeling ability is well measured by project examination

A4: Partial treatment of PDF files

Within the list of publications described in 4, there are some of them whose indexes are not available in the on-line databases such as ProQuest and Google Scholar. Such publications are:

- ICME11 (The 11th International Congress of Mathematics Education, Mexico, 2008)
- PMEs (Conference proceedings of the International Group for the Psychology of Mathematics Education, 2006-2015)
- PMENAs (Conference proceedings of the North American Chapter of the International Group for the Psychology of Mathematics Education, 2006-2015)

In this thesis, these publications are not analyzed since it will be much more convenient if electronic indexes are available. Such indexes are possibly available to the conference managers internally. When access to these indexes are available, the systematic review on these publications can be done efficiently. Nonetheless, this appendix section presents some possible ways to perform the review even without access to the electronic citation indexes to such publications.

ICME-11 (Mexico) did not publish the proceedings as a whole. Their official website has a list of all article titles and abstracts, and so they are downloaded for our analysis. One problem is that they are all saved as PDF files but not Endnote references. It is solved by running a Python script to extract the first page of each file. Another problem is that not all papers are available on the website. Fortunately, they provided full access to the two topic study groups (TSG 9 and 21) that are related to application.

There are 10 PME meetings from 2006 to 2015, but only 2006-09, 13, 14 proceedings are available. Furthermore, the PDF files are large (24 files, with a few hundred pages each). The Python script used in ICME-11 is not helpful because the articles are not in separate PDF files. Thus, in order to extract the first page of each article, the PDF action wizard is utilized instead. It can be done in two different ways: (1) the pages that contains an article's title usually shared some feature, and one may search for such feature to identify all these pages. This was done in the 2006, 07, and 09 files. (2) The table of contents can be turned into a text file, so that each line of the file contains one title, then a search is applied to the PDF file with this list. This was done in the 2008, 13, and 14 files.

The 2008 and 2014 PME-NA proceedings were printed with the PME proceedings. The 8 other proceedings were managed in a similar way as done to PME.

For these publications, PDF files are available from the official website, which are sufficient for the purpose of reviewing one or a few articles. However, since each year's proceeding contains more than 1,000 PDF pages, a systematic review is not possible without a mechanism to include only the relevant articles. A partial mechanism is developed for two different kinds of PDF collections. In the first kind of collection, each article is stored as an individual PDF file. This is how ICME11 provided their files: an official proceeding is not available for that year, but when an author submitted a publishable article after the conference, the article is stored and published at the conference website. In the second kind of collection, many articles are stored as a single PDF file. This is how the PMEs and PMENAs did: volumes of conference proceedings are published, each of which contain hundreds of articles. The partial mechanism is described as follows:

(1) *Extract only the first page of each article.* The first page of each article is supposed to contain its citation information necessary for a systematic review, including the title, author(s), and abstract. In the following, the first page of each article is also called a *title page*.

For ICME11, since it is known that there is one article per PDF file, a Python script (presented below) can read the PDF files, extract the first page, then combine all these title pages into one PDF file.

```
from PyPDF2 import PdfFileWriter, PdfFileReader
import os, sys
directory_name = 'output'
```

```
for filename in os.listdir('.'):
    print('name: %s' % filename)

output_file = PdfFileWriter()
     input_handle = open(filename, 'rb')
    input_file = PdfFileReader(input_handle)

output_file.addPage(input_file.getPage(0))
output_stream = file('firstof_'+filename,'wb')
output_file.write(output_stream)

output_stream.close()
input_handle.close()
```

For PME and PMENA, a PDF Action Wizard is used instead of a Python script, because one needs some extra criteria to distinguish the title pages from other pages. By inspection, it was found in many PME proceedings that the title pages exclusively contain a bottom line like this:

```
[Year. Author.] Proceedings xx-th Conference of the International
Group for the Psychology of Mathematics Education, [...]
```

In other words, these title pages contain the name of the conference exactly, and most other pages do not. Therefore, it was decided to do a keyword search of "Conference of the International Group for the Psychology of Mathematics Education'' in the PME proceedings, make these keywords highlighted, then extract all pages that has something highlighted. These steps can be done automatically using PDF Action Wizard scripts[14]. However, some of the proceedings (e.g., PME 2008, 2013, 2014) cannot be done this way, because either the title pages do not contain the conference's name at the bottom, or the conference's name is included in *every* page. In either case, a way to extract the title pages is to highlight some words in each title page *manually*, then

[14]All the PDF Action Wizard scripts used are made available by the author at the following link:

perform the rest of the steps using PDF Action Wizard scripts. The PMENA proceedings can be managed by the same methods.

This step generated 723 title pages for ICME11, 2,752 pages for PME, and 2,647 pages for PMENA, as indicated in Table 3.1 (p. 67), not counted towards the total.

(2) *Highlighting keywords.* To do a keyword search of “model*’’ and “appl*’’, the PDF Action Wizard scripts are again useful, since they provided ways to highlight all the matched occurrences of a list of keywords, and extract all the pages with highlighted words. This step narrowed the collection to 321 pages for ICME11, 1,309 pages for PME, and 1,206 pages for PMENA.

It is observed that the number pages was cut down greatly but there were still about three thousand PDF pages. In order to check how many of these articles are useful, a partial analysis is done in the PME 2006 proceeding, Volume 1. Out of its 57 pages, 4 of them can be categorized as *T1* or *T2* articles (cf. Appendix A1), which is less than 10%. However, the ICME11 has more relevant articles because there are two special Topic Study Groups (TSG): *TSG 9, Mathematics Education In and For Work* and *TSG 21, Mathematical Applications and Modelling in the Teaching and Learning of Mathematics.* These two study groups contributed 47 articles, which are all supposed to be relevant. Therefore, we estimate that the loss of articles due to the exclusion of ICME11, PME, and PMENA to be $47 + (321 - 47 + 1309 + 1206) \cdot 4/57 \approx 250$ (articles). This is not a small number compared to our analysis in Chapter 4, which was done upon 527 articles. If time permits in the future, the excluded articles can be analyzed by the help of electronic indexes, possibly provided by the conference managers for ICME, PME, and PMENA.

Appendix B. Sample analysis records of public assessment items

Figures B1 & B2 show the comments and records in the analysis of PISA items. Figure B3 illustrates a full analysis on Florida's FCAT 2004 using the "modeling spectrum"; only records for items 1 through 30 are shown. Figure B4 shows how subsequent analysis were recorded more concisely using binary codes.

Unit Item Code	Unit Name	Modeling Tasks	Inference Methods	Modeling subcompetencies	[1] real model	[2] math model	[3] math work	[4] interpret	[5] verify	no model
PM00FQ01	"MATH - P2012 Apartment Purchase Q1"									
M124Q01	P2000 Walking	model is given as algebraic equation, need to understand, and solve within the model using simple arithmetic	with partial credit for the subs step				yes			
M124Q03T	P2000 Walking	similar to 1, need to interpret between pace and speed, but also need to convert units, with not "nice" numbers	with partial credit for correct first step			yes	yes			
M145Q01T	P2000 Cubes	model is implicit. Only need to follow instruction, and the arithmetic is very simple					yes			
M148Q02T	Continent Area	although map itself is a model, student has to "remodel" it to more familiar geometrical figure and realize that the area is approximated. No further interpretation needed. The question did not specify how good the approximation should be, and did not require student to care about it.				yes	yes			
M150Q01	P2000 Growing Up	question not directly depend on the picture. Only need to understand the words and do subtraction of decimals					yes			
M150Q02T	P2000 Growing Up	not only need to understand how to find the time when females are taller than males but also need to express answer correctly (as an interval in math or in daily language)						yes		
M150Q03T	P2000 Growing Up	explanation is needed, and the steepness of the curve is the key. Not required to explain in detail WHY the slope represents the average growth rate. This is standard kind of mathematization.						yes		
M179Q01T	P2000 Robberies	read bar graphs, interpret "huge increase". Need sharp explanations							yes	
M266Q01T	P2000 Carpenter	the geometric models are made, and the solutions does not involve much mathematisation. Student only needs to see that the border is the perimeter. The main work is a calculation technique of perimeter。 . It is also possible that a student misinterpret the border and still get it perfectly right	only base on students' correct answers				yes			
M302Q01T	Car Drive	Need to know how to read graphs. Need to translate the question to "find highest points in the graph"						yes		
M302Q02	Car Drive	need to model the "slammed brake" as a sharp decrease of speed, then need to read the graph accordingly				yes		yes		
M302Q03	Car Drive	need to explain how it is possible to compare the two distances using the graph, and do the calculations correctly					yes	yes		
M421Q01	Height	student needs to explain the definition of average								yes

Figure B1: Analysis of PISA released items (Part 1/2)

Unit Item Code	Unit Name	Modeling Tasks	Inference Methods		Modeling subcompetencies	[1] real model	[2] math model	[3] math work	[4] interpret	[5] verify	no model
M421Q02T	Height	it is assumed that student leaned "mode" and "median". However, since argument is not needed, it is well possible that student just think that all are false because they are not mentioned in the definition. This task detects "misunderstanding".						yes			
M421Q03	Height	it is either higher algebraic skill or a deeper understanding of average						yes			
M535Q01	Twisted Building Q1	need to make assumption of the height of each floor, and create a mathematical model. The math work is simply multiplication (and perhaps rounding). Explanation is needed				yes	yes	yes			
M535Q02	Twisted Building Q2	need spacial understanding, and work among different perspectives					yes	yes			
M535Q03	Twisted Building Q3	similar as question 2 but is somewhat harder					yes	yes			
M535Q04	Twisted Building Q4	need to model "the 10th floor" to the existing model					yes	yes			
M537Q01	Heart Beat Q1	model is given. Only need to understand the question mathematically						yes			
M537Q02	Heart Beat Q2	model is also given. Only need to express the model mathematically.						yes			
M552Q01	Rock Concert	has to model the space occupied by standing fans in a rock concert. The math work is easy.				yes	yes	yes			
M598Q01	Making A Booklet	if treat geometrical shapes and transformations as models then the student has to actually do the "calculation"					yes	yes			
M710Q01	Forecast of Rain	no need to create a model but need to understand probability models and statements so as to interpret correctly							yes		
M810Q01T	Bicycles	model is given by the table. Only need to understand the model and do the work (easy subtraction)						yes	yes		
M810Q02T	Bicycles	again no modeling is needed; also, it is possible to do the problem by simple proportions						yes	yes		
M810Q03T	Bicycles	model is given (gears as ratio)						yes	yes		
M833Q01T	Seeing the tower	first it is necessary to simplify the situation – no need to consider the "shadowing" or where the bird is. Then need to use a common model for seeing – the straight line path of light or "eyesight", making a mathematical problem: which faces can be hit by a ray from the location of observation, without hitting other faces first? Then need to work mathematically.				yes	yes	yes			

Figure B2: Analysis of PISA released items (Part 2/2)

Item number	[0] application	[1] pose r-w question	[2] conceptual modeling	[3] idealization	[4] mathematiz ation	[5] math work	[6] interpret	[7] verify	[8] re-iterate
1	Yes	No NA No	No NA No	No NA No	Yes Yes No	No NA Yes	No NA No	No NA No	No NA No
2	Yes	No NA No	Yes Yes No	No NA No	Yes Yes No	No NA Yes	No NA No	No NA No	No NA No
3	Yes	No NA No	No NA No	No NA No	No NA No	No NA Yes	No NA No	No NA No	No NA No
4	Yes	No NA No	Yes Yes No	No NA No	No NA Yes	No NA Yes	No NA No	No NA No	No NA No
5	No								
6	Yes	No NA No	No NA No	No NA No	Yes Yes No	No NA Yes	No NA No	No NA No	No NA No
7	Yes	No NA No	Yes Yes No	No NA No	Yes Yes Yes	No NA Yes	No NA No	No NA No	No NA No
8	Yes	No NA No	Yes Yes No	No NA No	Yes No No	No NA Yes	No NA No	No NA No	No NA No
9	Yes	No NA No	Yes Yes No	No NA No	Yes Yes No	No NA Yes	No NA No	No NA No	No NA No
10	Yes	No NA No	Yes Yes No	No NA No	No NA Yes	No NA Yes	No NA No	No NA No	No NA No
11	Yes	No NA No	Yes Yes No	No NA No	Yes Yes No	No NA Yes	No NA No	No NA No	No NA No
12	Yes	No NA No	No NA No	No NA No	No NA No	No NA Yes	No NA No	No NA No	No NA No
13	Yes	No NA No	Yes Yes No	No NA No	No NA No	No NA Yes	No NA No	No NA No	No NA No
14	Yes	No NA No	No NA No	No NA No	Yes Yes No	No NA Yes	No NA No	No NA No	No NA No
15	Yes	No NA No	No NA No	No NA No	Yes Yes No	No NA Yes	No NA No	No NA No	No NA No
16	Yes	No NA No	Yes Yes No	No NA No	Yes Yes No	No NA Yes	No NA No	No NA No	No NA No
17	Yes	No NA No	Yes Yes No	No NA No	Yes Yes No	No NA Yes	No NA No	No NA No	No NA No
18	Yes	No NA No	Yes No No	No NA No	Yes Yes No	No NA Yes	No NA No	No NA No	No NA No
19	Yes	No NA No	Yes No No	No NA No	Yes Yes No	No NA Yes	No NA No	No NA No	No NA No
20	Yes	No NA No	Yes Yes No	No NA No	Yes Yes No	No NA Yes	No NA No	No NA No	No NA No
21	Yes	No NA No	Yes Yes No	No NA No	Yes Yes No	No NA Yes	No NA No	No NA No	No NA No
22	Yes	No NA No	No NA No	No NA No	Yes Yes No	No NA Yes	No NA No	No NA No	No NA No
23	Yes	No NA No	Yes Yes No	No NA No	Yes Yes Yes	No NA No	No NA No	No NA No	No NA No
24	Yes	No NA No	Yes Yes No	No NA No	Yes Yes No	No NA Yes	No NA No	No NA No	No NA No
25	Yes	No NA No	Yes Yes No	Yes Yes No	Yes Yes No	No NA Yes	No NA No	No NA No	No NA No
26	Yes	No NA No	Yes Yes No	No NA No	Yes Yes No	No NA Yes	No NA No	No NA No	No NA No
27	Yes	No NA No	Yes Yes No	No NA No	Yes Yes No	No NA Yes	No NA No	No NA No	No NA No
28	Yes	No NA No	Yes Yes No	No NA No	No NA No	No NA Yes	No NA No	No NA No	No NA No
29	Yes	No NA No	Yes No No	No NA No	Yes Yes No	No NA Yes	No NA No	No NA No	No NA No
30	Yes	No NA No	Yes No No	No NA No	Yes No No	No NA Yes	No NA No	No NA No	No NA No

Figure B3: Sample of analysis of public assessment items (Florida's FCAT 2004, items 1-30)

Item number	[0] application	[1] pose r-w question	[2] model conceptually	[3] idealize	[4] mathematize	[5] solve mathematically	[6] interpret	[7] verify	[8] re-iterate
Abitur 2001									
GM1. Infinitesimalrechnung (calculus)									
I									
II									
GM2. Wahrscheinlichkeitsrechnung/Statistik (Probability theory / Statistics)									
III	1	110	110	000	111	001	000	000	000
IV	1	110	110	111	111	001	001	000	000
GM3. Analytische Geometrie									
V									
VI									
	2 out of 6								
Analysis (I)									
Teil 1	0								
Teil 2	1	110	110	000	111	001	001	001	000
analysis (II)									
Teil 1	0								
Teil 2									
1	1	000	000	000	111	001	000	000	000
2	1	000	110	000	110	001	001	000	000
Stochastik(I)	1	000	110	000	111	001	000	001	000
Stochastik(II)	1	110	110	000	001	001	001	001	000
Geometrie(I)	1	110	110	000	111	001	111	000	000
Geometrie(II)	0								
	4 out of 6								

Figure B4: Sample of analysis of public assessment items (Bavaria, Germany's Abitur 2012)