The Mathematics Educator
2018 Vol. 27, No. 2, 33–72

A Rubric Development Study for the Assessment of Modeling Skills

Ayşe Tekin-Dede and Esra Bukova-Güzel

The purpose of this study was to introduce a rubric named "Rubric for the Assessment of Modeling Skills" for assessing cognitive modeling skills and explaining the rubric's development process. The dimensions of the rubric are understanding the problem, simplifying, mathematizing, working mathematically, interpreting and validating. Twenty-three 6th grade and twenty 11th grade students took part in the study. During the ten weeks of the data collection process, we analyzed the participants' solutions on the modeling tasks and constructed the levels of the dimensions. The rubric offers a detailed assessment and scoring for different solutions that may arise in any modeling task implementation.

In the 21st century, students are expected to understand fundamental mathematical concepts, translate a new problem into a mathematical problem, make the problem amenable to mathematical treatment, identify relevant mathematical knowledge to solve the problem, and then evaluate the solution in the context of the original problem (Schleicher, 2012). In parallel with these expectations, it is also emphasized that students need to use mathematics while solving the emergent problems in everyday life, society and the workplace with both the Common Core State Standards for Mathematics (CCSSM; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) and the Principles and Standards for School Mathematics (National Council of Teachers of Mathematics [NCTM], 2000). To meet these expectations, international surveys such as the Programme for International Student Assessment (PISA) and Trends in

Ayşe Tekin-Dede *received her PhD degree in mathematics education at the University of Dokuz Eylül. Her research interests include mathematical modeling and collective argumentation in mathematics education.*

Esra Bukova-Güzel *is a full professor of mathematics education at the Faculty of Buca Education, University of Dokuz Eylül. Her research interests include teaching and learning of mathematical modelling, students' conceptual learning, and teacher professional development.*

International Mathematics and Science Study (TIMSS) measure students' mathematical literacy skills to assess their ability to use mathematics in solving real-life problems. Stacey (2010) argues that the concept of literacy is mostly related to mathematical modeling. Mathematical modeling demonstrates how mathematical knowledge can be applied to solve real-life problems (Lesh & Sriraman, 2005). Accordingly, it is important for teachers to take advantage of mathematical modeling, demonstrating how these subjects will still be relevant in life outside of school (Kaiser, Schwarz & Tiedemann, 2010; Lesh, Young & Fennewald, 2010). In this way, students will be competent in taking surveys such as the PISA and TIMSS. However, Maaß (2005) argues that, based on the PISA and TIMSS scores, students have difficulty applying mathematics to everyday life and propose that mathematical modeling should be integrated into school mathematics lessons to overcome these difficulties.

The fact that countries such as Singapore, who have top performances in mathematics (Organisation for Economic Co-operation and Development [OECD], 2016), place a particular importance on the development of modeling competencies from an early age supports the argument of the necessity of integrating modeling into the lessons. The students' ability to model real situations necessitates certain cognitive, metacognitive, affective and social competencies (Biccard & Wessels, 2011; Galbraith & Stillman, 2006; Kaiser, 2007; Kaiser et al., 2010; Lesh & Zawojewski, 2007; Maaß, 2006). These competencies undoubtedly have an impact on completing the modeling process. However, Maaβ (2006) points out that the content of the modeling process is directly related to cognitive modeling skills. Cognitive modeling skills require the conscious direction of modeling processes and approaches in performing the steps of the modeling process (Blomhøj & Kjeldsen, 2006).

For the reasons mentioned above, our purpose in the present study is to determine how to assess cognitive modeling skills. Some researchers assessed students' modeling competencies by using multiple-choice tests (e.g., Biccard & Wessels, 2011; Galbraith & Haines, 1997; Grünewald, 2012; Haines, Crouch & Davis, 2001; Maaβ, 2006; Maaβ & Mischo, 2011) or rubrics (e.g., Anhalt & Cortez, 2015; Berry & O'Shea, 1982; Chan, Ng, Widjaja & Seto, 2012; Galbraith & Clatworthy, 1990; Keck, 1996; Leong, 2012). For example, Haines et al. (2001) assessed students' model construction competency by using the following multiple-choice test question:

> A ferry has total deck space of area. It carries cars, each car taking up an amount C of deck space, and lorries, each lorry needing an amount of L deck space. Each car pays £p for the crossing and each lorry pays £q. the manager wants to know how many cars (x) and how many lorries (y) to take on board so as to obtain the maximum revenue. Which one of these options give the revenue subject to the restriction on deck space?
>
> A. $\mathrm{xp} + \mathrm{yq}$ subject to $yC + xL \leq A$
> B. $\mathrm{xp} + \mathrm{yq}$ subject to $xC + yL \leq A$
> C. $(\mathrm{x} + \mathrm{y})(\mathrm{p} + \mathrm{q})$ subject to $xC + yL \leq A$
> D. $\mathrm{xp} + \mathrm{yq}$ subject to $xC + yL = A$
> E. $(\mathrm{x} + \mathrm{y})(\mathrm{p} + \mathrm{q})$ subject to $(x + y)(C + L) \leq A$
> (p. 377)

The multiple-choice modeling tests may not give us realistic information, especially since the open-ended structure of the modeling tasks may reveal more than one model, depending on the assumptions. For example, the items of the multiple-choice test may not include a model constructed by a student. In this case, if the student cannot make any choice, one might mistake his/her inaction as a lack of model construction skills, even though s/he demonstrated a proper approach to model construction. Similarly, if a student selects the correct choice by chance, we might assume s/he has the skills of model construction. Thus, using open-ended tasks rather than multiple choice modeling tests may give us a better insight into students' thinking. Additionally, when a choice must be made from existing options, a student's cognitive process might differ from that which would lead him or her to the correct solution (Büchter & Leuders, 2005).

The purpose of this study was to develop an analytic scoring rubric to assess students' cognitive modeling skills as they work on modeling tasks. The research questions are as follows:

- How can we create a reliable and valid rubric that assesses students' cognitive modeling skills?
- How did the rubric change over time to accommodate the challenges encountered in applying the rubric to the students' modeling activities?

In the frame of the research questions, we concentrated on the fact that the rubric should fit multiple grade levels, move away from task-specific definitions toward task-independent definitions, and have a structure as detailed as possible to enable others to make independent measurements.

Theoretical Background

Some have defined modeling competencies as the skills and abilities necessary in purposively and properly completing the modeling process and an individual's willingness (Kaiser & Maaß, 2007; Kaiser & Schwarz, 2006; Maaß, 2006). In other definitions, modeling competencies are defined as the ability to conduct the modeling process in an independent way (Maaβ & Gurlitt, 2011). As can be seen from the above definitions, both completing the modeling process and willingness are in the process. These are two interactive dimensions which are cognitive and affective. These dimensions directly affect each other, that is, both cognitive competence and willingness are effective in solving any modeling task. However, these are separate structures because the former is associated with mental actions and the latter is associated with affective states. In the rubric we developed in this study, we did not address affective competences because we only focused on mental actions. We chose the modeling cycle under a cognitive perspective (Borromeo Ferri, 2006) as our theoretical framework to investigate the cognitive skills of the students. In the literature, different words such as competency/competence, skill and ability were used to explain the students' approaches to the modeling process. We preferred to use the term “modeling skills” to explain students’ solution processes in this study.

The researchers used different modeling processes to define modeling skills based on their perspectives in understanding and interpreting modeling and the structure of modeling tasks (e.g., Blomhøj & Højgaard Jensen, 2003; Blomhøj & Kjeldsen, 2006; Borromeo Ferri, 2006; Cabassut, 2010; Henning & Keune, 2007; Houston, 2007; Maaß, 2006; Stillman, Brown & Galbraith, 2010). Cognitive analyses are needed as the modeling process necessitates intensive mental actions of students. Since the content of the modeling process is related to the cognitive modeling skills (Borromeo Ferri, 2010; Maaβ, 2006), cognitive actions are required to complete this process effectively (Blum & Leiß, 2007). Analysis of students’ cognitive skills in the modeling process is required to describe, interpret and explain what is happening in students' minds (Blum, 2011).

The cognitive modeling skills consisted of understanding, simplifying, mathematizing, working mathematically, interpreting and validating the problem in line with the stages of Blum and Leiß’s (2007) modeling cycle. Students first tried to understand the given real situation while working on a modeling problem. They then made the required assumptions, simplified the problem, identified what was needed to solve the problem and constructed a real model of the problem (Blum & Borromeo Ferri, 2009; Borromeo Ferri, 2006; Maaß, 2006). Mathematizing skills require students to identify the key variables, associate the key variables with each other and construct mathematical models using the appropriate mathematical representations (Borromeo Ferri, 2006; Lesh, Galbraith, Haines & Hurford, 2007; Maaß, 2006). Since these models are constructed using relevant mathematical knowledge and skills, students find mathematical solutions to the given problems by working mathematically (Blum & Borromeo Ferri, 2009; Borromeo Ferri, 2006; Lesh et al., 2007; Maaß, 2006). In other words, students interpret the mathematical results in the context of a real-life situation, bringing about a solution to the real-life problem (Blum & Borromeo Ferri, 2009; Borromeo Ferri, 2006). In the validating phase, students evaluated the effectiveness of the mathematical models, the models’ solutions and the general modeling processes based on their real-life experiences and assumptions (Blum & Borromeo Ferri, 2009; Borromeo Ferri, 2006; Lesh et al., 2007; Maaß, 2006).

Blum (2011) emphasized the importance of teaching modeling skills by stating that one of the main aims in teaching mathematics is to develop a student’s modeling skills. Modeling implementations, such as Blomhøj and Højgaard Jensen’s (2003) holistic and atomistic approaches, help the development of the cognitive modeling skills (Blomhøj & Kjeldsen, 2006). In the holistic approach, students work through all the stages in the modeling process, while in the atomistic approach, they work in certain stages of the modeling process (Blomhøj & Højgaard Jensen, 2003). In order to reveal the development of students’ cognitive modeling skills, the question of how it can be assessed must come to the forefront.

The State of the Literature: The Rubric Assessment of the Cognitive Modeling Skills

While students are working on a task, the rubrics can be used for assessing their cognitive modeling skills. There are different examples of rubrics to measure students' modeling competencies in the literature. These rubrics are given in Table 1 specifying the dimensions of the rubrics, levels of the dimensions and brief notes about the rubrics.

When considering the rubrics in Table 1, Berry and O’Shea (1982), Galbraith and Clatworthy (1990), Keck (1996), and Leong (2012) identified the dimensions of the rubric in parallel with different modeling processes. Other researchers used mathematical categories as dimensions so they could not directly asses the modeling skills. Some rubrics had levels for each dimension but it was noteworthy that these levels were not explained in detail. In addition, assessment for modeling skills in some rubrics (Berry & O’Shea, 1982; Galbraith & Clatworthy, 1990) were overlapping. In light of these issues, the rubrics outlined in Table 1 provided a basis for us while we developed our rubric.

Table 1
Modeling Rubrics from the Literature

Researchers	Dimensions	Notes regarding Levels and Dimensions
Berry & O'Shea (1982)	▪ Abstract ▪ Formulation ▪ Initial model ▪ Data ▪ Revisions to the model ▪ Conclusions ▪ Presentation	No explained levels. Not exactly have a rubric structure. Used for marking students' cognitive actions.
Galbraith & Clatworthy (1990)	▪ Specify the problem clearly ▪ Formulate an appropriate model ▪ Solve the mathematical problem including interpretation, validation and evaluation/refinement ▪ Communicate results in a written and oral form	Three levels for each dimension. Overlapping dimensions.
Keck (1996)	▪ Identification of problem ▪ Formulation of assumptions ▪ Construction of model ▪ Solution and interpretation ▪ Validation ▪ Communication ▪ Mathematics	Different number of levels for each dimension. Overlapping dimensions.
Chan et al. (2012)	Dimensions are not related to modeling competencies even though mathematical competencies were given.	Three levels for each dimension: 1. Assumptions 2. Interpretation of task using real-world knowledge, 3. Mathematical reasoning and computation.
Leong (2012)	▪ Identifying variables ▪ Formulating a model ▪ Mathematical operations ▪ Interpreting the results ▪ Validating the conclusion ▪ Reporting on conclusions	Different number of levels for each dimension. Dimensions are grounded in the modeling process in the CCSSM. Scores from 0 to 4 and weights, but there were no details explaining them.

Researchers	Dimensions	Notes regarding Levels and Dimensions
Anhalt & Cortez (2015)	▪ Explanations ▪ Connections ▪ Working ▪ Reasoning ▪ Concepts ▪ Calculations	Six levels ranging between 0 and 5 for each dimension. The dimensions are considered under the categories named the modeling process, the constructed model, the solution, students' reflections in constructing mathematical models depending on the CCSSM.

It was also important for us to understand what the rubric means and what it contains. A rubric is defined as a scoring tool that articulates the expectations for a task by listing the criteria and describing levels of quality (Andrade, 2000; Stevens & Levi, 2013). A rubric has three essential features: evaluation criteria, quality definitions and scoring strategy (Popham, 1997). The dimensions of the rubrics correspond to the evaluation criteria and include the indicators to be used when determining the quality of a work. Quality definitions are associated with the levels of the dimensions and are considered as detailed explanation of what a student must do to meet the requirements of the so-called dimensions and correspond to a particular level of achievement. In terms of the modeling process, for example, while model construction is a dimension, the level of students' model construction approaches corresponds to the levels of these dimensions. In addition, the scoring strategy for a rubric should be explained in detail so that the final grade for the assessed skill can be included. When the rubrics in Table 1 are considered in terms of these three features described in the literature, it can be seen that each of them has certain deficiencies. For example, we could not see the detailed explanation of the dimensions of each rubric in the literature. It is necessary for the developed rubric to be comprehensive, easily interpreted by others, and for its levels to be clearly defined, as some rubrics' dimensions and skills overlap instead of being included separately. Teachers and researchers have had difficulty evaluating which dimension and level the modelers are in according to their solutions. In addition, Vos (2013) stated that there were no scoring guidelines that would enable different people to make similar assessments in evaluating modeling skills. Thus, an analytic scoring rubric was developed in this study to primarily fill the gaps in the literature and to present a detailed assessment instrument that both researchers and teachers can use to assess cognitive modeling skills. With the developed rubric, it is possible to determine students' cognitive modeling skills and to what extent they use these skills when solving modeling tasks. In addition, the rubric gives information about the skills required for any stage of the modeling process as well as the skills required for completing the whole modeling process.

Method

While developing the rubric, named "Rubric for the Assessment of Modeling Skills" (RAMS), we adopted the constructivist grounded theory approach (Charmaz, 2006). At the root of the constructivist grounded theory, researchers begin with a review of the literature to determine what has been done before, ask a particular question, and then construct the concepts (Evans, 2013).

In this context, we examined assessment approaches for modeling skills through literature review, identified strengths and weaknesses of these approaches, and tried to develop the rubric by conducting goal-directed implementations.

In the study, twenty-three 6th grade and twenty 11th grade students were included. Each teacher of the classroom together with the researchers implemented the modeling tasks by forming groups of four or five students. We then examined the students' solution approaches when they were solving the modeling tasks. Each group presented their solutions to the whole class. The solution and presentation processes were recorded by a video camera. The data we present in this study is from the video-recording transcripts, the students' written work and the researchers' observation notes. We conducted the implementation in a middle school first, and then in a high school over a one-week interval (see Table 2). We conducted the implementation during two-hour lessons in an elective course named Applications of Mathematics in the middle school and in a pre-calculus course in the high school. The groups solved the given tasks in the first lesson, and in the second lesson they presented their solutions and had in-class discussions. In the discussions, the students were requested to assess the other groups' solutions; the researchers provided scaffolding to enable the students in displaying more comprehensive approaches especially in studying some cognitive modeling skills such as interpreting and validating in which students discussed less.

The process of data collection was conducted in three stages, and the total data collection process took ten weeks. The stages and names of the implemented modeling tasks (see the Appendix) are given in Table 2.

Table 2
The Modeling Tasks

Stage	Week	Grade Level	Task
Preparation Stage	1	MS	Step Problem (Hıdıroğlu, Tekin & Bukova Güzel, 2011)
	2	HS	Step Problem (Hıdıroğlu, et al., 2011)
	3	MS	Bed Problem (Borromeo Ferri, 2014)
	4	HS	Bed Problem (Borromeo Ferri, 2014)
1st Stage	5	MS	Apple Pie Problem Problem (adapted from Schukajlow et al., 2012)
	6	HS	Obesity Problem (Tekin Dede & Bukova Güzel, 2013)
	7	MS	Invoice Problem (Tekin Dede, 2015)
2nd Stage	8	HS	Ancient Theatre Problem (Tekin, Hıdıroğlu & Bukova Güzel, 2010)
	9	MS	Time in School Problem (Maaß & Mischo, 2011)
	10	HS	Fuel Problem (Tekin, 2012)

Note. HS = high school, MS = middle school

In the preparation stage, two modeling tasks were implemented in both the middle and high schools, respectively. We analyzed the collected data, and the students' solution approaches were placed into the rubric's first dimension format. After the categorization process, we revised the first format, so we decided to move to the first stage. During three weeks, three different modeling tasks were implemented in the first stage. The researchers categorized the students' approaches by examining the deficiencies that emerged in the previous stage. After that, we decided to conduct the second stage by implementing three different modeling tasks to increase certain dimensions' levels. The whole categorization process, the development process of the rubric, and the validity and reliability studies will be discussed after the section that emphasizes the roles of the researchers.

The Role of the Researchers

The researchers conducted the study based on their eight years of experience in implementing mathematical modeling at elementary, secondary and undergraduate levels. The researchers also developed several rubrics to evaluate the students' solutions to the particular modeling tasks applied in the implementations (Hıdıroğlu, Tekin Dede, Kula & Bukova Güzel, 2014; Hıdıroğlu, Tekin Dede, Kula Ünver & Bukova Güzel, 2017; Tekin Dede, Hıdıroğlu & Bukova Güzel, 2017). These studies formed a basis for what we could do systematically to develop a more comprehensive rubric. Along with the teacher of the class, we conducted implementations over a period of 10 weeks. Before conducting the implementations, we gave information to the middle and high school teachers about the content of the study. Then, in order to obtain information about the students, we observed some lessons, which we recorded on video to help the students feel relaxed in front of the camera. While the groups solved the modeling tasks, we took observation notes to reveal the students' cognitive modeling skills. We asked the students probing questions to help support their discussions about their solutions.

RAMS Development Process

Preparation Stage

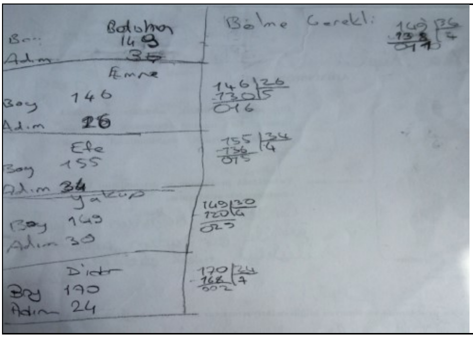
First, we determined a rubric format (see Table 3) according to the chosen theoretical framework. In this format, we selected three levels for each skill and gave examples from students' solutions. The reason why we chose three levels was that we predicted that the actions with the lowest and the highest levels would be considered for each dimension and that there would also be a middle group of actions. We also wrote our evaluations, which we compared amongst ourselves.

Table 3
The First Rubric Format

	Level			Samples from	Evaluation by
Skills	1	2	3	Students' Work	the Researcher
Understanding the problem					
Simplifying					
Mathematizing					
Working mathematically					
Interpreting					
Validating					

The reason why the levels did not have definitions is because the requirement for each skill's explicit definition was made based upon the opinions obtained after we examined the students' solutions. After we implemented the Step Problem, which required the construction of a mathematical model presenting the relationship between the distance of one's steps in walking and one's height, in the 6th grade class, we examined the transcript of the video-recording regarding the implementation process, the students' written work and our observation notes and then evaluated each group by using the first rubric format. After implementing the Bed Problem, we repeated the assessment process performed in the first implementation. As a result, we decided that the three levels given to each dimension were insufficient. Additionally, we noticed the differences in the middle and high school students' solution approaches. For instance, when we dealt with the middle schoolers' solutions in the simplifying stage, we determined that "not making assumptions" fit into Level 1, "making assumptions partially appropriate for real life" fit into

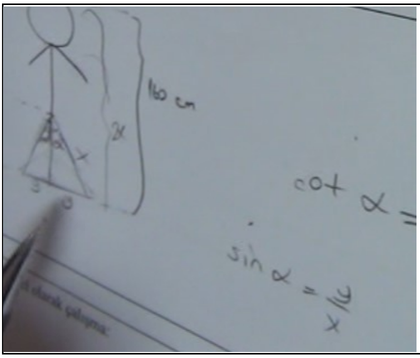
Level 2 and "making assumptions completely appropriate for real life" fit into Level 3 (see an example from Level 2 in Figure 1).



<u>Evaluation by the researcher:</u>
Because the group only benefitted from their measurements and did not consider relevancy to real-life, they made an assumption partially appropriate for real-life.

Figure 1. Example of Simplifying from the middle school students' solution to the Step Problem

Unlike the middle school students, the high school students considered the potential variables needed to construct a model as well as formulated the assumptions in the Step Problem (see an example from Level 2 in Figure 2).

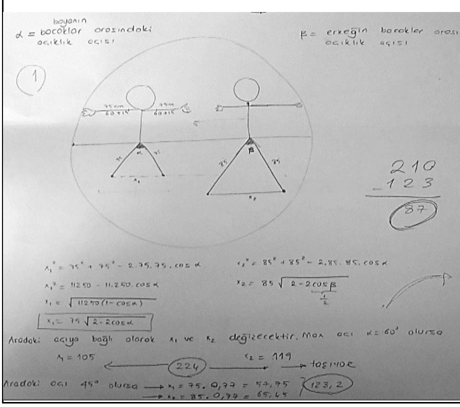


<u>Evaluation by the researcher:</u>
They made a partially appropriate assumption in noting that a human's length could be twice as much as his leg's length. Based on that assumption, they examined the angle between the legs. To find the distance between his steps, they determined variables based on their assumptions and constructed a model using these variables.

Figure 2. Example of Simplifying from the high school students' solution to the Step Problem

In the Bed Problem implementations, students' solutions in the simplifying stage brought about similar findings (see examples from Level 3 in Figure 3). The groups who did not apply a real-life situation to solve the problem used their own body lengths instead of their parents' body lengths to formulate a solution. The groups who made assumptions completely appropriate for real-life solved the problem by using their own opinions about the lengths of adult men and women. Similarly to the results of the Step Problem, the high school students also determined the variables needed to construct a model as well as expressed their assumptions verbally.

Evaluation by the researcher:
(Middle school) They made a partially appropriate assumption in noting that a human's length could be twice as much as his leg's length. Based on that assumption, they examined the angle between the legs. To find the distance between his steps, they determined variables based on their assumptions and constructed a model using these variables.



Evaluation by the researcher: (High school) The students made assumptions appropriate for real-life regarding the lengths and widths of their parents. Based on the assumptions and the angle between the legs, they identified the variables.

Figure 3. Examples of Simplifying from the middle and high school students' solution to the Bed Problem

At the end of the four weeks, we decided to revise the rubric's format and increase its levels. Even if the modeling tasks and student levels differed, it would show that the students' approaches contained similar patterns in each dimension. In other words, it would show how the task-specific definitions turned into task-independent definitions. More specifically, we firstly reviewed the rubric evaluations of each task. When we looked at the level explanations in these rubrics, we saw that some of the actions were similar if the task changed. For example, we coded students' actions as making incorrect assumptions, not distinguishing necessary/unnecessary variables, not simplifying the problem, constructing problem-specific models, etc. We then grouped the student actions in these encodings and discussed which ones could be used to explain the same level. Thus, we obtained statements independent of task. Four different levels for each dimension emerged from the solutions, and the researchers created the levels' definitions (see Table 4).

Table 4
The Rubric Format Including Four Levels for Each Dimension

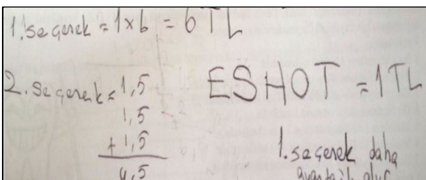
Level	Definition
Understanding the Problem	
Level 1	Includes the expressions showing that s/he did not understand the problem, did not determine the givens and goals, and did not form, or mistakenly formed, a relationship between them.
Level 2	Includes the expressions showing that s/he understood the problem to some extent, determined the givens and goals to some extent but did not form, or mistakenly formed, a relationship between them.
Level 3	Includes the expressions showing that s/he understood the problem completely, determined the givens and goals but did not form, or mistakenly formed, a relationship between them.
Level 4	Includes the expressions showing that s/he understood the problem completely, determined the givens and goals, and formed a relationship between them.
Simplifying	
Level 1	Not simplifying the problem, not determining the necessary/unnecessary variables, and making wrong assumptions.
Level 2	Simplifying the problem to some extent, and determining the necessary/unnecessary variables to some extent but making wrong assumptions.
Level 3	Simplifying the problem, determining the necessary/unnecessary variables, and making partly-acceptable assumptions.
Level 4	Simplifying the problem, determining the necessary/unnecessary variables, and making realistic assumptions.

Level	Definition
Mathematizing	
Level 1	Not constructing, or mistakenly constructing, mathematical model/s.
Level 2	Constructing correct mathematical model/s based on partly-acceptable assumptions.
Level 3	Constructing incomplete/wrong mathematical model/s based on realistic assumptions and relating them to one another.
Level 4	Correctly constructing the needed mathematical model/s according to realistic assumptions, explaining model/s, and relating them to one another.
Working Mathematically	
Level 1	Not presenting a mathematical solution, wrongly solving the constructed models or trying to solve the wrong mathematical model.
Level 2	Correctly solving the mathematical models that were incompletely/wrongly constructed.
Level 3	Including deficiencies/mistakes in the solution of the correctly constructed mathematical models.
Level 4	Achieving the correct mathematical solution by solving the correctly constructed mathematical models.
Interpreting	
Level 1	Misinterpreting, or not interpreting, the obtained mathematical solution in a real-life context.
Level 2	Correctly interpreting the erroneous/incomplete mathematical solution in a real-life context.
Level 3	Incompletely interpreting the obtained correct mathematical solution in a real-life context.
Level 4	Correctly interpreting the obtained correct mathematical solution in a real-life context.
Validating	
Level 1	Not validating or making an invalid validation.
Level 2	Validating completely and not correcting the determined mistakes.
Level 3	Validating completely and correcting the determined mistakes to some extent.
Level 4	Validating completely and correcting the determined mistakes.

Both the middle school and high school students had difficulty in the interpreting and validating stages. Even if the students tried to interpret or validate the obtained solution, the interpretation and validation were either incomplete or they treated the validation as checking the calculation. In the next implementations, we decided to use different modeling tasks at the middle and high school levels to observe the different solution approaches and varying rubric levels.

The First Stage

In the first stage of the RAMS development process, we implemented three different modeling tasks in the different levels. After that, we revised the dimensions of mathematizing and working mathematically by including five levels (see Table 5). The reason for the revision was that when the students made partially acceptable assumptions in the simplifying stage, they constructed both incomplete/wrong and correctly constructed mathematical models. An excerpt from the solution paper of the Apple Pie Problem is given in Figure 4.



Evaluation by the researcher: The students made a partially acceptable assumption that the ESHOT* fee is 1 TL (ESHOT fee is 0.90 TL). It was determined that they constructed incomplete mathematical models because they had forgotten to include the so-called fee in constructing their models. So, they constructed an incomplete mathematical model based on a partially-acceptable assumption.

*ESHOT is the public transport bus service in Izmir, Turkey.

Figure 4. Example of Mathematizing from the middle school students' solution to the Apple Pie Problem

After we defined the mathematizing levels, we realized that there were some cases that could also fit into the evaluation in working mathematically stage. The second level of solving mathematical models was split into two levels, including deficiencies/mistakes in solving the models and solving the models accurately. Thus, the skill of working mathematically had five defined levels (see Table 5).

During the modeling task implementations in the first stage, the groups were prompted to present their solutions and their solution approaches were discussed as a whole class. We decided to examine whether the students should have displayed deeper approaches in the interpretation and validation stages, since they had gained experience in both working on the modeling tasks and engaging in class discussions. In the discussions, the students' validation approaches emphasized that validation meant not only checking the calculations and correcting mistakes but also validating the assumptions, models, solution of models and the whole process. Thus, we decided to implement an additional three modeling tasks to determine whether the interpretation and validation skill levels should be changed.

The Second Stage

When the Ancient Theatre Problem was implemented, we observed that one group interpreted the problem related to their real-life experiences, although they reached incorrect results. The students found the height of the Aspendos Ancient Theatre to be 15m, whereas it is 36m in real-life. They then discussed their results in terms of their real-life experiences and thought that the height of one step could be greater than they initially determined (see a section from the students' expressions in Figure 5).

We revised the interpreting dimension by including five levels (see Table 5). When the students got an incorrect/incomplete solution, they could interpret the results in an incomplete way, depending on the real life. Similarly, different validating approaches were also elicited. The students who checked only their calculations in the first stage came to validate their assumptions, models and solutions. They tried to correct their mistakes based on their experiences gained by the discussion of in-class presentations and the researchers' emphases on the necessity of validation. In the solution of the Time in School Problem, one group first considered the number of school days to find the time spent in school. While validating their solution, they thought that solving the problem by depending on hours instead of days was more reasonable and handled the hours spent at school during a day (see a section from the students' expressions in Figure 6).

Oğuz:	The height of a step is 40 cm.
Seval:	55. The height of a step is 55 cm.
Volkan:	Just here of a step is 60cm (showing the height of one step).
Oğuz:	Isn't 60 cm so high?
Veli:	Steps in an ancient theatre are a little bit higher (means that he had been to an ancient theatre before).
Volkan:	A step is 60 cm in height.

<u>Evaluation by the researcher:</u>
While the students were questioning if their solution was logical or not in real life, they discussed the assumptions as to the height of a step by referring back to the solution process. In this context, they made interpretations that the steps in ancient theatres should be higher depending on their real life experiences.

Figure 5. Example of Interpretation from the high school students' solution to the Ancient Theatre Problem

We extended the validating dimension and divided into two main categories: validating completely and validating partially. We evaluated the approaches of the students who had validated all their assumptions, models and solutions as validating completely, while the approaches of the students who had considered one or a few of these were evaluated as validating partially. Some groups tried to correct their mistakes by validating, but others completed the solution process without correcting their mistakes. Therefore, whether or not the condition of correcting the mistakes was involved in validation, so the validating skill was defined at seven levels (see Table 5). Finally, we constructed the RAMS with five levels for understanding the problem, four levels for simplifying, five levels for mathematizing, five levels for working mathematically, five levels for interpreting and seven levels for validating.

Nuran:	[First writes down on the paper, "We think that we do not spend most of the year in school."] Now, we will prove this thought. We'll make a calculation. How many hours of a day do we spend in school?
Ulaş:	6.
Nuran:	The lessons start at 07.30 and finish at 13.30.
Pırıl:	6 hours.
Ulaş:	But, we will count the days; look, it doesn't say the hours. There are 52 weeks in a year. We'll multiply 52 by 50 and subtract that result from 365. [The students make calculation on the number of days.]
…	
Nuran:	First, we'll find how many hours we spend in school, then subtract the result from 24, then multiply it by 5.
Erdem:	Why do we multiply by 5?
Nuran:	We go to school five days. Then, to find our free time, we'll multiply the result by 5. [The students decide to make the calculation depending on the hours spent in school; afterwards, they consider the holidays as hours.]

<u>Evaluation by the researcher:</u>
The students first constructed and solved a model depending on the number of days spent in school. Then, in validating their solution, they decided to consider the hours because they did not spend the whole day in school. Then they corrected all assumptions about the time spent in school and on holiday. Afterwards, they were observed to correct their model and solution based on the so-called assumptions.

Figure 6. Example of Validating from the middle school students' solution to the Time in School Problem

As we explained earlier, we aimed to evaluate all possible solution approaches in problem solving. When we discussed understanding the problem, there were three different states including not understanding the problem, partly understanding, and completely understanding. In the meantime, since the level of relationship between the givens and goals also affected the understanding of the problem, we structured the levels of understanding the problem in the rubric so as to include problem understanding and association approaches developmentally. Similarly, simplifying approaches included actions to simplify the problem, identify necessary/unnecessary variables, and make wrong, partially acceptable and realistic assumptions, respectively. When we came to mathematizing, we found that student approaches were affected by simplifying. Therefore, the constructed models had to contain expressions based on previously created assumptions to explain the levels related to mathematizing. Since the partially accepted assumptions were not fully appropriate to the problem, even if the correct model was constructed based on those assumptions, there would be no suitable models for solving the problem. But while the process based on realistic assumptions continued, it was more valuable for us to construct an incomplete or wrong model based on these assumptions. Therefore, we decided to order the expressions in Level 3 and Level 4 as seen in Table 5. Similarly, the skills of working mathematically were also influenced by mathematical models constructed in the previous step. Therefore, even if the student constructed a wrong model, the correct or incorrect progress in solving this model was valuable in terms of working mathematically. In other words, even if the model was wrong, the students' progress in the solution process was important in terms of fulfilling the modeling skills. When we consider the interpretation and validation skills, we found that students encountered the complexity and intertwining of the modeling process. In fact, students made the interpretation of mathematical solutions that they performed/obtained. Additionally, we found that students made interpretation as well as validation in many steps throughout the whole modeling process. For example, both during simplifying the problem and making assumptions, or after obtaining mathematical results, students made the interpretation and validation. Therefore, as we aimed to assess students' modeling skills, we expressed all the cases in which interpretation and validation may occur in the rubric. In addition, when considering the whole modeling process, we expected the students to solve the problem in an appropriate way and to make an interpretation accordingly. Therefore, the interpretation made in cases where the model or solution was complete was more valuable than the interpretation made in cases where the model or solution was incomplete or erroneous.

Table 5
The Final Form of the RAMS

Level	Definition
Understanding the Problem	
Level 1	Not understanding the problem, not determining the givens and goals, and not forming or mistakenly forming a relationship between them.
Level 2	Understanding the problem partly, determining the givens and goals to some extent but not forming or mistakenly forming a relationship between them.
Level 3	Understanding the problem completely, determining the givens and goals, not forming or mistakenly forming a relationship between them.
Level 4	Understanding the problem completely, making unimportant mistakes in determining the givens and goals, not forming a relationship between them.
Level 5	Understanding the problem completely, determining the givens and goals, and forming a relationship between them.
Simplifying	
Level 1	Not simplifying the problem, not determining the necessary/unnecessary variables, and making wrong assumptions.
Level 2	Simplifying the problem partly, determining the necessary/unnecessary variables to some extent, and making wrong assumptions.
Level 3	Simplifying the problem, determining the necessary/unnecessary variables, and making partially acceptable assumptions.
Level 4	Simplifying the problem, determining the necessary/unnecessary variables, and making realistic assumptions.
Mathematizing	
Level 1	Not constructing or mistakenly constructing mathematical model(s).
Level 2	Constructing incomplete/wrong mathematical model(s) based on partially acceptable assumptions.
Level 3	Constructing correct mathematical model(s) based on partially acceptable assumptions.
Level 4	Constructing incomplete/wrong mathematical model(s) based on realistic assumptions and relating them to one another.
Level 5	Correctly constructing the needed mathematical model(s) according to realistic assumptions, explaining model(s) and relating them to one another.

Level	Definition
Working Mathematically	
Level 1	Not presenting a mathematical solution, solving the constructed models wrongly, or trying to solve the wrong mathematical model.
Level 2	Including deficiencies/mistakes in the solution of the mathematical models constructed incompletely/wrongly.
Level 3	Solving correctly the mathematical models constructed incompletely/wrongly.
Level 4	Including deficiencies/mistakes in the solution of the correctly constructed mathematical models.
Level 5	Achieving the correct mathematical solution by solving the correctly constructed mathematical models.
Interpreting	
Level 1	Misinterpreting or not interpreting the obtained mathematical solution in a real-life context.
Level 2	Incompletely interpreting the erroneous/incomplete mathematical solution in a real-life context.
Level 3	Correctly interpreting the erroneous/incomplete mathematical solution in a real-life context.
Level 4	Incompletely interpreting the obtained correct mathematical solution in a real-life context.
Level 5	Correctly interpreting the obtained correct mathematical solution in a real-life context.
Validating	
Level 1	Not validating or making a wrong validation.
Level 2	Validating partially, not correcting the determined mistakes.
Level 3	Validating partially, correcting the determined mistakes to some extent.
Level 4	Validating partially, correcting the determined mistakes.
Level 5	Validating completely, not correcting the determined mistakes.
Level 6	Validating completely, correcting the determined mistakes to some extent.
Level 7	Validating completely, correcting the determined mistakes.

Studies of Validity and Reliability

Creswell (2013) stated that researchers should spend a long time in the field to increase the validity of qualitative research. We confirmed the validity through our experiences with modeling, implementation with students, a ten-week data collection, constant comparative analysis, and an eight-month period in which the data were examined again. In data collection and analysis, we used triangulation (Creswell, 2013) based on validity strategies by utilizing the transcriptions and the researchers' observations. We defined the RAMS development process with all stages and tried to increase the reliability of the study. While the levels were explained, the reliability was provided by stating excerpts of students' expressions and solutions.

Data analysis made based on a theoretical framework increases the reliability of a study (Yıldırım & Şimşek, 2008). Defining the dimensions and levels based on the framework was one of the aspects of increasing the reliability of the study. Opinions from experts (i.e., mathematics educators studying modeling) were considered for the RAMS, and we asked them to make the needed corrections to the statements by reading the definitions of the levels. We aimed for the rubric to be understood in the same way by every user. Thus, the definitions for all levels were structured and standardized, and the statements were made to be more obvious. In addition, the validity and reliability were increased by implementing several modeling tasks with more than one group at different grades.

Patton (2014) stated that validity and reliability could be enabled with multiple coders. We increased validity and reliability by handling the evaluations of two expert mathematics educators who used the RAMS. The experts, studying modeling independently, evaluated the students' solutions on modeling tasks. We gave them the transcripts of different groups' solutions to each task. Then, we compared their evaluations , and calculated the percentage of agreement (Miles & Huberman, 1994) by considering which skill was handled at which level. The agreed-upon evaluation number was 22 out of 30. The percentage of the agreement proved to be 73.3%. Four tasks were coded as T1, T2, T3 and T4 in Table 6, and the levels specified by the two experts were given as from L1 to L7.

Table 6
Percentage of Agreement in Evaluations by Experts with the RAMS

	First Specialist				Second Specialist			
Skills	T1	T2	T3	T4	T1	T2	T3	T4
Understanding the Problem	L5	L5	L5	L5	L5	L5	L5	L5
Simplifying	L4	L4	L4	L4	L4	L4	L4	L4
Mathematizing	L5	L4	L5	**L5**	L5	L4	L5	**L4**
Working mathematically	L5	L2	L5	**L5**	L5	L2	L5	**L3**
Interpreting	**L5**	L2	L4	L3	**L4**	L2	L4	L3
Validating	L7	**L1**	L7	L6	L7	**L3**	L7	L6
Agreement ratio	5/6	5/6	6/6	4/6				
Total agreement ratio	22/30 = 73.3%							

Note. Four tasks were coded as “T1,” “T2,” “T3,” and “T4.” The discrepencies between coders are shown in bold.

According to the stability method, which is used to ensure the reliability of the data analysis, a certain period after the first analysis of the data, a second analysis should be carried out by the same person (Krippendorff, 1980; Weber, 1985). To that end, the same researcher evaluated the data again obtained from three modeling tasks applied at the second stage with the RAMS about six months later. For this purpose, the researcher chose the solution of three groups and used the rubric for evaluating their solution. We used the formula of Miles & Huberman (1994) to calculate the correspondence percentage. In the comparison, the number of all cases is 36, while the number of matching scores is 32. Thus, the correspondence percentage was 88.8%.

Scoring Suggestions for Assessments with the RAMS

The purpose of rubrics is not only to assess the students’ performance, but also to inform the students about their own development based on the results of the rubric (Black & William, 2009). It is necessary to give feedback to the students in the consequence of the assessments carried out with the RAMS. After developing the RAMS, we studied how the students’ solutions would be scored. Although we started out with the holistic approach when we started to design the rubric, we saw that during the designing phase, the RAMS can be used with any atomistic approach provided that we divide the rubric into its dimensions. Therefore, the RAMS is a feasible rubric with both atomistic and holistic approaches. We decided on five different scoring options related to the RAMS through the opinions of field experts working on modeling and mathematics teachers. Level 1 was scored with a 0 in each skill dimension in all different scoring options.

In the first approach, according to the atomistic approach, when a modeling task included using only one specific modeling skill, solutions were scored by considering the dimensions and levels of this skill.

The second approach is to weight each of the dimensions equally. When a modeling task was solved by the students according to the holistic approach, the researcher/teacher primarily determined the maximum score that s/he could give when it was time to assess the students' solutions. The number of points to assign each dimension are determined by dividing the maximum possible score by 6. After the score of each dimension is determined equally to one another, the points within each dimension are divided according to the number of levels. In this approach, the students' solutions are assessed and equiponderant scoring is carried out for each skill. Because the number of levels in each dimension is different, difficulty can arise in scoring because the scores obtained at the end of the divisions are decimal expansions.

Thirdly, when a task is solved according to the holistic approach, the researcher/teacher may want to assess each level by starting from 0, in which case the maximum score to be taken with the RAMS is 25. In calculation of scores with this assessment, trouble may occur. Because the numbers of the levels in each dimension are not equal, there can be a bias in favor of a greater number of levels such as the validation skill.

When a modeling task is solved according to the holistic approach, the researcher/teacher may carry out an assessment by numerically weighting the score of whichever skill s/he wants to emphasize in the fourth approach. Thus, if s/he prefers to focus on what skill s/he has initially determined, s/he will have the chance to make an assessment by keeping the score of related dimension more.

Alternatively, when a task is solved with the holistic approach, the scoring of the skills apart from the understanding-the-task skill can be sufficient according to the fifth approach. In modeling applications, students start the solution after they have understood the problem. If the students do not understand the problem, they may ask for help from the teacher, and the teacher should enable them to understand correctly with scaffolding. So, the scoring of the understanding skill can be ignored in the assessment. An assessment could be made in which the rest of the five skills get equal scores (each 20 points) if the total point is set to 100. Thus, 20 points of each skill are equally shared according to the levels.

Conclusion

In this study, the RAMS, which is a rubric to assess students' cognitive modeling skills, and its development process were introduced. The RAMS can be used for assessing solutions in different modeling tasks because it was formulated from the implementation of several modeling tasks. Many researchers and teachers who want to determine their students' cognitive modeling skills can take advantage of the RAMS as an assessment tool because skill levels are defined clearly.

The RAMS discloses the strengths and weaknesses of students in terms of modeling skills and provides the opportunity to make both quantitative and qualitative assessment. Determining the students' modeling skills through quantitative assessment, teachers may express these values as points. Which approaches students display for modeling skills can be revealed in a clear and detailed way through qualitative assessment. Thus, students' difficulties with skills may be clearly observed and effective modeling implementations may be planned to overcome these difficulties.

Discussion

Based on the literature that we reviewed in terms of the rubrics developed for assessing cognitive modeling skills, we can assert that the RAMS has many aspects different from those rubrics and it was constructed in a more detailed way than those rubrics. The main reasons for being more detailed were that the dimensions were appropriate to each modeling skill and the level definitions for each dimension include all possible actions and were explained in detail. In addition, the implementation of many tasks, not a single task, and the fact that the results of the implementations shaped the rubric by using ongoing analyses made the RAMS more detailed. The rubric can apply to both group's modeling activities and individuals' modeling activities based on the purpose of the evaluator.

In the rubric of Berry and O'Shea (1982), even though the dimensions were handled as cognitive modeling skills, explanations about the dimensions' levels were not included. There were no explanations of how students' solutions would be assessed. Berry and O'Shea stated that assessment should be made with a maximum of five points considering the assumptions, simplifications and features determined to be important. However, the points to be taken regarding the existing level of students' assumptions or to what extent they make simplifications were not included. This may raise dilemmas for evaluators and could prevent objectivity among different evaluators. Besides, they pointed out that presentations of solutions could be assessed in terms of clarity and layout. On the other hand, in the RAMS the social skills of students were not assessed in the presentation; instead, the students' approaches are cognitively dealt with in the presentations because the rubric was aimed to evaluate cognitive modeling skills.

In the rubric of Galbraith and Clatworthy (1990), each dimension assesses more than one skill at the same time. For example, the second dimension contained assumptions and eliciting models, and the third dimension contained interpretations and validation as well as performing the mathematical solution. Even though the levels' definitions of the dimensions were available, the evaluator may have difficulty in deciding the students' level. For instance, the first level of the third dimension includes find a solution for the task with no assistance, the second level includes finding a solution with little or no assistance and no correction of the model, and the third level includes the interpretation and validation of the model by solving the task independently. A problem may appear in determining students' level when they have solved the task with little assistance, and interpreted and validated a constructed model. Different from the RAMS, the rubric of Galbraith and Clatworthy (1990) aims to assess cognitive and social skills of students based on their written and verbal solutions. However, the RAMS may provide a more detailed assessment because it focuses only on cognitive skills.

Keck's (1996) rubric assesses use of mathematics and communication skills as well as cognitive modeling skills. When considering cognitive modeling skills, we note that the dimensions of the rubric have similarities with the RAMS. It only differs from the RAMS because of its fourth dimension where mathematical solution and interpretation were assessed together. Considering the definitions of the levels, interpretation of the solution may be assessed without working mathematically because it only assessed the compatibility of the mathematical solution. There was no dimension or level definition about the assessment of the solution of the model. Because of the absence of the definitions, the rubric of Keck may be evaluated as lacking in assessing cognitive modeling skills. He explained that the dimension of identification of the problem is assessed in three levels and all other dimensions in five levels. The RAMS is more comprehensive than Keck's rubric because in the validation skill, the seven levels were handled as corrections of assumptions, models and solutions.

The rubric of Chan et al. (2012) contained an assessment in three dimensions as assumptions, interpretation of task and solutions, and mathematical reasoning and computation. Each dimension was assessed using three levels; however, the levels remained restricted while explaining dimensions. For instance, the level of mathematical reasoning and computation dimension evaluated the number of discussed variables, the compatibility of the used mathematics, the reasonability and explicitness of mathematical reasoning and computation at the same time. The fact that the levels contained more than one skill at the same time revealed the question of how the assessment would be carried out when one or some skills were not observed. Therefore, it was thought that the evaluators who use the rubric could have difficulty in classifying the students' solutions under appropriate categories and making an assessment.

Leong (2012) used a modeling process in his rubric, similar to that in the RAMS. Yet, Leong's rubric was thought not to provide an assessment clear enough for the researchers/teachers who wanted to assess cognitive modeling skills because of the fact that dimensions in the rubric measured more than one skill, definitions of levels about the dimensions were not made, and the reason for the difference in the weights of dimensions was not explained.

Anhalt and Cortez (2015), in their rubric, considered the students' skills in the solution of the modeling tasks but did not specifically assess cognitive modeling skills. The levels of the rubric assessed the student explanations in the solution, connections among concepts, student work, their reasoning, use of representations and mathematical concepts and calculations. Although there were some criteria for the cognitive modeling skills in levels' definitions, they were observed to differ from the RAMS because an assessment parallel to the skills was not specifically carried out.

Further Implementation

The RAMS provides an opportunity for the assessment of cognitive modeling skills, and presents scoring options for the researchers/teachers who wish to combine their assessments with quantitative or qualitative techniques depending on their purposes. Therefore, the RAMS may address different audiences and purposes. The strengths and weaknesses of the RAMS may be examined by teachers who use the RAMS in their modeling implementations. Besides the assessment of cognitive modeling skills, the RAMS may be improved by adding different dimensions such as assessing the meta-cognitive, affective and social skills used in the modeling process. By this means, different skills of students working in modeling tasks will be addressed. Studies may be designed how to assess students' solutions while working on a modeling task, so it may be possible to provide different solution approaches in each dimension of the RAMS. Moreover, it is suggested that researchers conduct design-based studies to enable students to develop their particular modeling skills that are evaluated as deficient by using the RAMS in future studies.

References

Andrade, H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership, 57*(5), 13–18.

Anhalt, C., & Cortez, R. (2015). Mathematical modeling: A structured process. *Mathematics Teacher*, *108*(6), 446–452.

Berry, J., & O'Shea. T. (1982). Assessing mathematical modelling. *International Journal of Mathematical Education in Science and Technology*, *13*(6), 715-724.

Biccard, P., & Wessels, D. (2011). Development of affective modelling competencies in primary school learners. *Pythagoras*, *32*(1), 1–9.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment Evaluation and Accountability*, *21*, 5–31.

Blomhøj, 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., & Kjeldsen, T. N. (2006). Teaching mathematical modelling through project work. *Zentralblatt für Didaktik der Mathematik-ZDM*, *38*(2), 163-177.

Blum, W. (2011). Can modelling be taught and learnt? Some answers from empirical research. In G. Kaiser, W. Blum, R. Borromeo Ferri & G. Stillman (Eds.), *Trends in Teaching and Learning of Mathematical Modelling* (pp. 15–30). New York, NY: Springer.

Blum, W., & Borromeo Ferri, R. (2009). Mathematical modelling: Can it be thought or learned? *Journal of Mathematical Modelling and Application*, *1*(1), 45–58.

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 – Proceedings from the Twelfth International*

Conference on the Teaching of Mathematical Modelling and Applications (pp. 222–231). Chichester, UK: Horwood Publishing.

Borromeo Ferri, R. (2006). Theoretical and empirical differentiations of phases in the modelling process. *Zentralblatt für Didaktik der Mathematik-ZDM*, *38*(2), 86–95.

Borromeo Ferri, R. (2010). On the influence of mathematical thinking styles on learners' modelling behaviour. *Journal für Mathematikdidaktik, 31*(1), 99–118.

Borromeo Ferri, R. (2014, April). *Mathematisches Modellieren Lernen und Lehren [Mathematical Modelling for Teachers and Learners].* Paper session presented at the 3rd Mathematics Education Applications Mathematical Modelling Activities Workshop, İzmir, Turkey.

Büchter, A., & Leuders, T. (2005). From students' achievement to the development of teaching: Requirements for feedback in comparative tests. *Zentralblatt für Didaktik der Mathematik-ZDM*, *37*(4), 324–334.

Cabassut, R. (2010). The double transposition in mathematisation at primary school. In V. Durand-Guerrier, S. Soury-Lavergne & F. Arzarello (Eds.), *Proceedings of the 6th Congress of the European Society for Research in Mathematics Education CERME 6* (pp. 2156–2165). Lyon, France: Service des publications, INRP.

Chan, C. M. E., Ng, K. E. D., Widjaja, W., & Seto, C. (2012). Assessment of primary 5 students' mathematical modelling competencies. *Journal of Science and Mathematics Education in Southeast Asia, 35*(2), 146–178.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis.* London: Sage.

Creswell, J. W. (2013). *Nitel, Nicel ve Karma Yöntem Yaklaşımları Araştırma Deseni*. [Qualitative, Quantitative, and Mixed Methods Approaches.] (S. B. Demir, Trans.) Ankara, Turkey: Eğiten Kitap.

Evans, G. (2013). A novice researcher's first walk through the maze of grounded theory: Rationalization for classical grounded theory. *The Grounded Theory Review, 12*(1), 37–55.

Galbraith, P. L., & Haines, C. R. (1997). Some mathematical characteristics of students entering applied mathematics courses. In S. K. Houston, W. Blum, I. Huntley & N. Neill (Eds.), *Teaching and Learning Mathematical Modelling* (pp. 77–92). Chichester, UK: Albion Publishing.

Galbraith, P. L., & Stillman, G. (2006). A framework for identifying student blockages during transitions in the modelling process. *Zentralblatt für Didaktik der Mathematik-ZDM*, *38*(2), 143–162.

Galbraith, P.L., & Clatworthy, N.J. (1990). Beyond standard models: meeting the challenge of modelling. *Educational Studies in Mathematics, 21*(2) 137–163.

Glaser, B. G. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. Mill Valley, CA: Sociology Press.

Grünewald, S. (2012, July). *Acquirement of modelling competencies: First results of an empirical comparison of the effectiveness of two approaches to the development of (metacognitive) modelling competencies of students*. Paper presented at the 12th International Congress on Mathematical Education, Seoul, Korea.

Haines, C., Crouch, R., & Davis, J. (2001). Recognizing students' modelling skills. In J. F. Matos, W. Blum, S. K. Houston, & S. P. Carreira (Eds.), *Modelling and mathematics education - ICTMA 9: Application in science and technology* (pp. 366–380). Chichester, UK: Horwood.

Henning, H., & Keune, M. (2007). Levels of modelling competencies. In W. Blum, P. L. Galbraith, H. W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education: The 14th ICMI study* (pp. 225–232). New York, NY: Springer.

Hıdıroğlu, Ç. N., Tekin Dede, A., Kula Ünver, S. & Bukova Güzel, E. (2017). Mathematics student teachers' modelling approaches while solving the designed Eşme rug problem. *EURASIA Journal of Mathematics, Science and Technology Education, 13*(3), 873–892.

Hıdıroğlu, Ç. N., Tekin Dede, A., Kula, S., & Bukova Güzel, E. (2014). Öğrencilerin kuyruklu yıldız problemine ilişkin çözüm yaklaşımlarının matematiksel modelleme süreci çerçevesinde incelenmesi [Examining students' solutions regarding the comet problem in the frame of mathematical modeling process]. *E-Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi, 31*, 1–17.

Tekin, A., Hıdıroğlu, Ç. N., & Bukova Güzel, E. (2010, October). *Öğrenciler matematiksel modellemede birlikte çalıştıklarında hangi yaklaşımları sergiliyorlar? [Which approaches do students display while working collaboratively in modeling?]*. Paper presented at the 9th Mathematics Symposium, Trabzon, Turkey.

Hıdıroğlu, Ç. N., Tekin, A., & Bukova Güzel, E. (2011). *The analysis of student teachers' thought processes in the mathematical modelling through a designed fermi problem.* Paper presented at the 35th Conference of the International Group for the Psychology of Mathematics Education (PME 35), Ankara, Turkey.

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: The 14th ICMI study* (pp. 249–256). New York, NY: 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 – Proceedings from the Twelfth International Conference on the Teaching of Mathematical Modelling and Applications* (pp. 110–119). Chichester, UK: Horwood.

Kaiser, G., & Maaß, K. (2007). Modelling in lower secondary mathematics classroom: Problems and opportunities. In W. Blum, P. L. Galbraith, H. W. Henn & M. Niss (Eds.), *Modelling and applications in mathematics education: The 14th ICMI study* (pp. 99–108). New York, NY: Springer.

Kaiser, G., & Schwarz, B. (2006). Mathematical modelling as bridge between school and university. *Zentralblatt für Didaktik der Mathematik-ZDM*, *38*, 196–208.

Kaiser, G., Schwarz, B., & Tiedemann, S. (2010). Future teachers' professional knowledge on modeling. In R. Lesh, P. L. Galbraith, C. R. Haines & A. Hurford (Eds.), *Modeling students' mathematical modeling competencies* (pp. 433–444). New York, NY: Springer.

Keck, H. L. (1996). *The development of an analytic scoring scale to assess mathematical modelling projects.* (Doctoral dissertation). Retrieved from Scholar works University of Montana.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills, CA: Sage Publications.

Leong, K.E. (2012). Assessment of mathematical modeling. *Journal of Mathematics Education at Teachers College*, *3*(1), 61–65.

Lesh, R., & Sriraman, B. (2005). Mathematics education as design science. *Zentralblatt für Didaktik der Mathematik-ZDM, 37*(6), 490–505.

Lesh, R., & Zawojewski, J. (2007). Problem solving and modeling. In F. Lester (Ed.), *Handbook for Research on Mathematics Education* (pp. 763–804). Charlotte, NC: Information Age Publishing.

Lesh, R., Galbraith, P., Haines, C., & Hurford, A. (Eds.). (2007). *Modeling students' mathematical modelling competencies: Proceedings of the 13th International Conference on the Teaching of Mathematical Modelling and Applications*. New York: Springer.

Lesh, R., Young, R., & Fennewald, T. (2010). Modeling in K-16 mathematics classrooms and beyond. In R. Lesh, P. L. Galbraith, C. R. Haines & A. Hurford (Eds.), *Modeling students' mathematical modeling competencies* (pp. 275–283). New York, NY: Springer.

Maaß, K. (2005). Barriers and opportunities for the integration of modelling in mathematics classes: Results of an empirical study. *Teaching Mathematics and Its Applications. 24*(2-3), 61–74.

Maaß, K. (2006). What are modelling competencies?. *Zentralblatt für Didaktik der Mathematik-ZDM*, *38*(2), 113–142.

Maaß, K., & Gurlitt, J. (2011). Designing a teacher questionnaire to evaluate professional development in modelling. In V. Durand-Guerrier, S. Soury-Lavergne & F. Arzarello (Eds.), *Proceedings of the 6th Congress of the European Society for Research in Mathematics Education CERME 6* (pp. 2056–2065). France, Lyon: Service des publications, INRP.

Maaß, K., & Mischo, C. (2011). Implementing modelling into day-to-day teaching practice: The project stratum and its framework. *Journal Für Mathematik-Didaktik*, *32*(1), 103–131.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis.* Thousand Oaks, CA: Sage Publications.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics.* Washington, DC: Author. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf.

Organisation for Economic Co-operation and Development. (2016). PISA 2015 results: Excellence and Equity in Education (Vol. I). Paris: OECD.

Patton, M. Q. (2014). *Nitel araştırma ve değerlendirme yöntemleri.* [Qualitative research & evaluation methods]. Ankara, Turkey: Pegem Akademi.

Popham, J. W. (1997). What's wrong and what's right with rubrics. *Educational Leadership, 55*(2), 72–75.

Schleicher, A. (2012). (Ed.). Preparing teachers and developing school leaders for the 21st century: Lessons from around the world. Paris, France: OECD.

Schukajlow, S., Leiss, D., Pekrun, R., Blum, W., Müller, M., & Messner, R. (2012). Teaching methods for modelling problems and students' task-specific enjoyment, value, interest and self-efficacy expectations. *Educational Studies in Mathematics, 79*, 215–237.

Stacey, K. (2010). Mathematical and scientific literacy around the world. *Journal of Science and Mathematics Education in Southeast Asia, 33*(1), 1–16.

Stevens, D. D., & Levi, A. J. (2013). *Introduction to rubrics: An assessment tool to save grading time, convey effective feedback, and promote student learning*. Sterling, VA: Stylus.

Stillman, G., Brown, J., & Galbraith, P. (2010). Identifying challenges within transition phases of mathematical modelling activities at year 9. In R. Lesh, P. L. Galbraith, C. R. Haines & A. Hurford (Eds.), *Modeling students' mathematical modelling competencies: ICTMA 13* (pp. 385–398). New York, NY: Springer.

Tekin, A. (2012). *Matematik öğretmenlerinin model oluşturma etkinliği tasarım süreçleri ve etkinliklere yönelik görüşleri [Mathematics teachers' views concerning model eliciting activities, developmental process and the activities themselves]* (Unpublished master's thesis). Dokuz Eylül University, İzmir, Turkey.

Tekin Dede, A. (2015). *Matematik derslerinde öğrencilerin modelleme yeterliklerinin geliştirilmesi: Bir eylem araştırması [Developing students' modelling competencies in mathematics lessons: An action research]* (Unpublished doctoral dissertation). Dokuz Eylül University, İzmir, Turkey.

Tekin Dede, A., & Bukova Güzel, E. (2013). Examining the mathematics teachers' design process of the model eliciting activity: Obesity problem. *Elementary Education Online, 12*(4), 1100–1119.

Tekin Dede, A., Hıdıroğlu, Ç. N., & Bukova Güzel, E. (2017). Examining of model eliciting activities developed by mathematics student teachers. *Journal on Mathematics Education, 8*(2), 223–242.

Vos, P. (2013). Assessment of modelling in mathematics examination papers: Ready-made models and reproductive mathematising. G.A. Stillman, G. Kaiser, W. Blum & J. P. Brown (Eds.), *Teaching mathematical modelling: Connecting to research and practice* (pp. 479–488). New York, NY: Springer.

Weber, R. P. (1985). *Basic Content Analysis*. Beverly Hills, CA: Sage Publications.

Yıldırım, A., & Şimşek, H. (2008). *Sosyal bilimlerde nitel araştırma yöntemleri* [Qualitative research methods in social sciences]. Ankara, Turkey: Seçkin Yayıncılık.

Appendix
Modeling Tasks Implemented in Data Collection

Step Problem (Hıdıroğlu et al., 2011)
Construct a model mathematically stating the relationship between the distance of one's steps in walking and one's height (Figure A1).

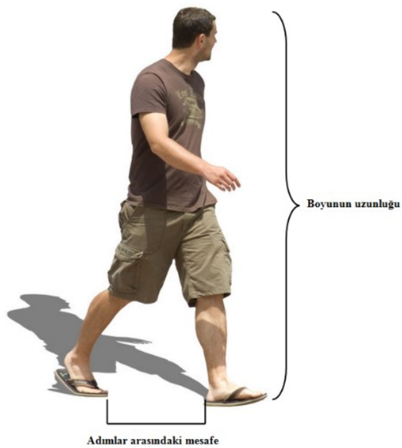


Figure A1.

Bed Problem (Borromeo Ferri, 2014)
When Deniz's parents looked at the catalogue of a furniture store, they liked a circular bed (diameter 210 cm) given in the above picture and decided to buy it. But they could not be sure whether they would feel comfortable when they slept on the bed. How much space will be left between them when Deniz's parents are sleeping on this bed so that none of their arms and legs are outside? (Figure A2).

Figure A2.

Apple Pie Problem (adapted from Schukajlow et al., 2012)
Sevinç, who invited to her home several friends for the weekend, wants her mother to make her famous apple pie for her friends. Her mother noticed that there are no apples in the house, and she wants to buy 3 kg apples from Sevinç. She has two options for buying apples:

- Option 1: Half a kilo of apples in the grocery store next door to her home is 1 Turkish Lira.
- Option 2: 1 kg of apple in the market slightly away from home is 1.5 Turkish Lira. But since the market is far away, it is absolutely necessary to get on the ESHOT bus.

Which option makes sense for Sevinç? Explain your thoughts with reasons.

Obesity Problem (Tekin Dede & Bukova Güzel, 2013)
Can, whose body mass index is 31.2 and is therefore considered obese, needs to lose weight. He will perform an exercise to select the table on the side, which will be 3 days a week, 20 minutes a day, and he will continue the same exercise by increasing 5 minutes compared to the previous week. He will only make a choice as an exercise from the table and not change this selection. (9 calories for 1 gram of fat should be spent.) Develop a model that will help him find how many weeks later he will reach the target weight. (Figure A3).

Lifting barbell	30 minutes	150 Cal
Skating	15 minutes	15 Cal
Climbing up stairs	15 minutes	15 Cal
Dancing	30 minutes	75 Cal
Biking	30 minutes	300 Cal
Playing tennis	30 minutes	120 Cal
Hiking	20 minutes	60 Cal
Playing basketball	30 minutes	300 Cal
Swimming	30 minutes	300 Cal
Playing volleyball	1 hour	180 Cal

Figure A3.

Invoice Problem (Tekin Dede, 2015)
In the table below, the price list of four different mobile phone operators is given per month. In addition to the monthly conversation and monthly messaging amounts for each item on the price list, there is also how much to pay for speech and messaging in case of package

overrun. How would you choose the most appropriate tariff under these conditions? Describe your choice mathematically. (Figure A4).

Name of the List		Monthly speech (every direction)	Monthly messaging (every direction)	Monthly price
Multum in parvo	Package overrun	100 minutes 50 krş per minute	300 SMS 40 krş per SMS	20 TL
SMS-full	Package overrun	350 minutes 50 krş per minute	750 SMS 30 krş per SMS	30 TL
Five hundred	Package overrun	500 minutes 45 krş per minute	500 SMS 35krş per SMS	45 TL
One thousand	Package overrun	1000 minutes 20 krş per minute	1000 SMS 15 krş per SMS	60 TL

Figure A4.

Ancient Theatre Problem (Tekin et al., 2010)
A group of tourists went to Aspendos Antique Theatre during their Antalya vacation. During this excursion, they took the photograph you see here, in the figure. (Figure A5).

A) What is the actual distance between marked people?
B) What is the actual height of the ancient theatre?

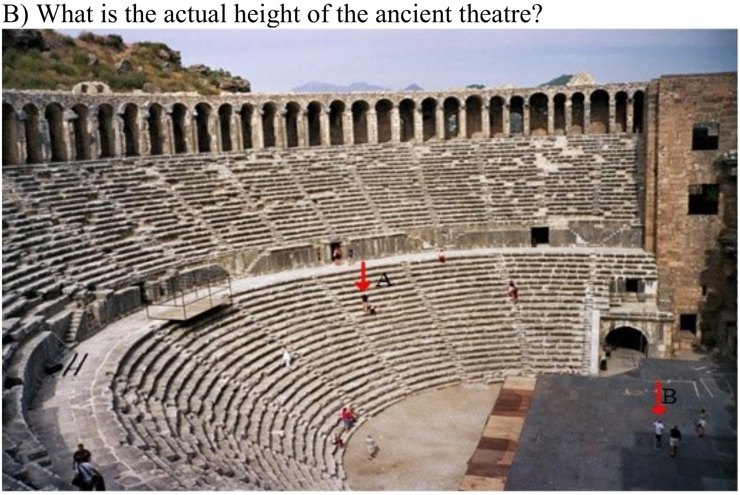


Figure A5.

Time in School Problem (Maaß & Mischo, 2011)
Deniz thinks that he has spent a lot of time in school, and he says, “I do not understand how time passes in school! I spend most of the year at school.” What do you think about this situation? Do you also think that you spend most of the year at school? Please do the necessary calculations to determine whether he is right or not.

Fuel Problem (Tekin, 2012)
Ali, who is visiting the country with a vehicle with a malfunctioning fuel tank, wonders if he could calculate the amount of fuel left in the tank by looking at the wet part of a stick put into the tank. For this, he wants you to develop a model that can calculate the amount of fuel remaining in the tank relative to the wet part of the stick.