Universiteit Utrecht

Master Thesis

Secondary school students' meta-modeling knowledge of concept-process models in biology

by

Manou van Montfort

Science Education and Communication
Track: Research and Development
30 ECTS Research Project

m.j.e.vanmontfort@students.uu.nl
Student number: 3982386
Academic year 2018/2019

Supervisors: S. Jansen MSc & dr. M.C.P.J Knippels
Freudenthal Institute
Utrecht University

Abstract

Meta-modeling knowledge is an important part of scientific literacy. However, models in biology education are mostly used to teach about the content of science rather than to develop the students' meta-modeling knowledge. Before developing model-based learning techniques that approach this gap in teaching, it is important to know the students' meta-modeling knowledge without an explicit teaching approach. Previous research has focused on scale models (such as a skeleton). This study investigates the current understanding of biological concept-process models (such as blood sugar regulation) by secondary school students, in regard to the meta-modeling knowledge aspects *nature of models* and *multiple models*. The results show that students can think on a more advanced level for concept-process models than previous studies on scale models found. Students show more advanced meta-modeling knowledge for microscopic models than for macroscopic models. A possible explanation for both results is that the abstractness of the model influences the displayed meta-modeling understanding. Models that require more abstract understanding (concept-process models and microscopic models) can possibly trigger more advanced meta-modeling knowledge than models that require less abstract understanding (scale models and macroscopic models).

Table of Contents

1. Introduction

Models play a major role in scientific research. They provide conceptual frameworks and allow scientists to explore possibilities, investigate complex systems and generate predictions and explanations (Odenbaugh, 2005). However, modeling skills are not only useful for scientists, but also for non-scientists. Science and society are inherently intertwined, the results of scientific research often find their way into society (e.g. climate change or drug research). In order to understand and value the science that is needed in society, citizens should be scientifically literate (Schwarz et al. 2009). Scientific literacy encompasses the skills that are required for understanding science in everyday life and personal decision-making in socio-scientific issues, such as the vaccination debate in society (Burns, O'Connor & Stocklmayer, 2003). The models that citizens encounter are representations of complex phenomena that, similar to the models used by scientists, help us simplify, describe and communicate complex phenomena (Svoboda & Passmore, 2013). Knowledge about models, the creation of models and the use of models (i.e. the process of modelling; meta-modeling knowledge) are thus relevant to both scientists and non-scientists. For this reason, meta-modeling knowledge has been introduced into the new secondary school curriculum for biology in The Netherlands (College voor Toetsen en Examens, 2016).

In biology education, the models that are used in textbooks and teaching range from concrete scale models to more complex concept-process models. These models can all be used to describe and simplify the real-world situation. An example of a scale model is a torso showing the anatomy of human organs (see figure 1). A scale model reflects external characteristics of objects, such as the relative size of the organs in the torso. Scale models are usually static. In a concept-process model, the concept that is referred to is not an object (like an organ), but a process (Harrison & Treagust, 2000). An example of a concept-process model is a model that shows the interaction of hormones, organs and cells to regulate blood sugar levels. These models can not only be used to describe and simplify the real-world situation, but also to predict future events. Concept-process models are more complex than scale models, because they are more abstract; these models often include arrows which show interaction between elements at different times or different locations, thus including dynamics of time and movement (Harrison & Treagust, 2000).

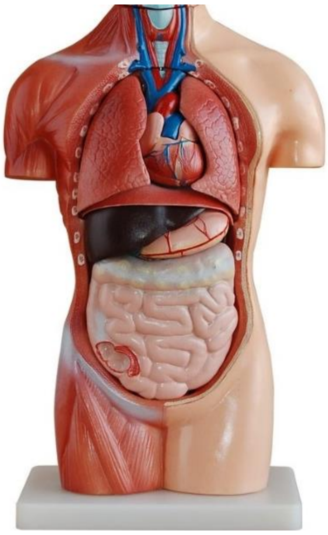


Figure 1 – A scale model for human anatomy.

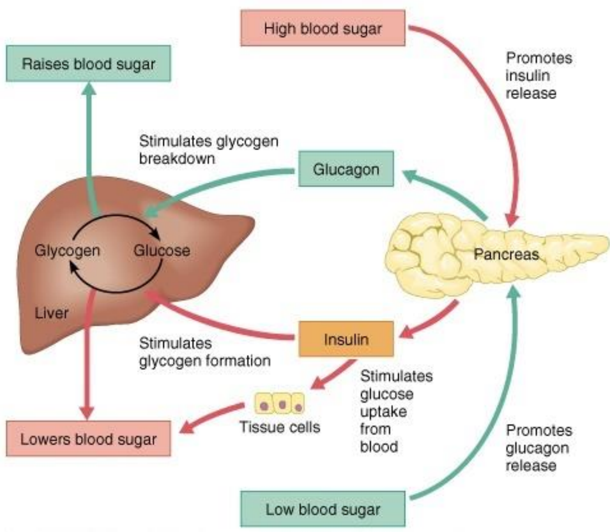


Figure 2 – A concept-process model for blood sugar level regulation.

Hodson (2003) identifies three main purposes of secondary school science education: (1) learning science, (2) learning about science and (3) doing science. Models can be used in each of these three purposes. In the first pace, models can be used as a teaching aid. The models can help to illustrate and communicate the theory. In the second place, models can be used to teach about the *nature of science*. Models can give students insight in the complexity of science. In the third place, models can be used to teach about *doing science*. Students who understand models and the process of modeling, also develop an understanding of how science investigates phenomena (Grosslight, Unger, Jay & Smith, 1991). In practice, most teachers use the models as an aid in teaching science content, neglecting the scientific process that creates, predicts and evaluates these models (Windschitl, Thompson, & Braaten, 2008). Therefore, students are not explicitly taught the required meta-modeling knowledge to develop an understanding of the models as they are used in science. This is a problem for the development of scientific literacy in students, since they do not learn the meta-modeling knowledge required for understanding the scientific use of models. A solution to this problem can be to incorporate suitable model-based learning techniques into the classroom. Before developing such model-based learning techniques, it is relevant to know the students' current scientific understanding of models.

To assess students' meta-modeling knowledge, Upmeier zu Belzen and Krüger (2010) developed a theoretical framework. This model was tested and used to assess the general understanding of models by students (Grünkorn, Upmeier zu Belzen & Krüger, 2014). Most of the models used in this study are scale models, leaving out the more complex concept-process models. To be able to investigate the students' understanding of concept-process models, Jansen, Knippels and Van Joolingen (2019) made adjustments to the theoretical framework by Upmeier zu Belzen and Krüger (2010). In order to obtain a thorough view of students' understanding of models, this study specifically aims to investigate the students' understanding of biological concept-process models. If the students' understanding of concept-process models in the current situation is known, it is possible to develop model-based learning techniques to specifically address the gaps in students' meta-modeling knowledge.

2. Theoretical Background

2.1 Different types of models used in biology education

In science education, many different models are used. Harrison and Treagust (2000) identify scale models as the most concrete models in science education, because they mostly display external characteristics. Somewhat more abstract models are symbolic models, such as chemical formulas. These models help students to build conceptual knowledge. According to Harrison and Treagust (2000), concept-process models are one of the most abstract models that are used in science education, since they display processes including multiple concepts and show the dynamics of time and movement.

In biology education, different types of models within the category of concept-process models can be identified. The field of biology consists of nested hierarchical organizational levels (Reece, 2011, p.), the organizational level of molecules is nested within the organizational level of cells, cells are nested within tissues, tissues are nested within organs, organs are nested within organ systems, organ systems are nested within organisms, organisms are nested within populations, and populations are nested within ecosystems. Concept-process models can occur at any of these organizational levels. Tsui and Treagust (2013) propose a classification of the models on different levels of representation in biology. The four levels that they distinguish are the (1) *macroscopic level*, (2) *microscopic level*, (3) *submicroscopic level*, and (4) *symbolic level*. The macroscopic level includes all that is visible to the naked eye. The microscopic level includes all that is visible using a light or electron microscope. The submicroscopic level includes biochemicals that can be visualized using laboratory techniques such as electrophoresis and chromatography. Last, the symbolic level is similar to the symbolic models that Harrison and Treagust (2000) describe: symbols, genotypes, phylogenetic trees and so on. Since models cannot be both a concept-process model and a symbolic model, the symbolic model is not addressed in this study.

The results of a previous study on students' meta-modeling knowledge of concept-process models (Jansen et al., 2019), suggested that there might be a difference between the students' meta-modeling knowledge for a model on a cellular organizational level and the students' meta-modeling knowledge for a model on an ecological organizational level, in such a way that students have a more advanced meta-modeling understanding for the model on a cellular level. These findings suggest that models on different organizational levels are different in such a way that they influence the students' meta-modeling knowledge. If the study by Jansen et al. (2019) is interpreted in the light of the classification by Tsui and Treagust (2013), a difference in students' meta-modeling knowledge is suggested between the macroscopic model (ecological level; visible to the naked eye) and the (sub)microscopic model (cellular level; invisible to the naked eye).

2.2 A theoretical framework for assessing meta-modeling knowledge

In the past, various frameworks have been developed to assess students' meta-modeling understanding. Upmeier zu Belzen and Krüger (2010) combined insights from previous frameworks (Crawford & Cullin, 2005; Grosslight, Unger, Jay & Smith, 1991; Justi & Gilbert, 2003) into a new theoretical framework. Grünkorn et al. (2014) tested this theoretical framework using scale models. Jansen et al. (2019) adjusted this theoretical framework in order to make it suitable to assess students' meta-modeling knowledge for concept-process models next to scale models. The framework by Jansen et al. (2019) is the framework used in this study (table 1).

Table 1 – The theoretical framework for assessing students' understanding of models and their use in science (Jansen et al., 2019)

	Complexity		
Aspect	**Level I**	**Level II**	**Level III**
Nature of models	• Model as copy • Model with great similarity • Model represents a (non-) subjective conception of the original • Displays a process, its components and how they are related	• Parts of the model are a copy • Model as a possible variant • Model as focused representation	• Model as hypothetical representation
Multiple models	• Different model object properties	• Focus on different aspects	• Different assumptions
Purpose of models	• Model for showing the facts • Model for showing events	• Model to identify relationships	• Model to examine abstract/concrete ideas
Testing models	• Testing of material • Testing of basic requirements	• Comparison between original and model • Comparison and matching of original and model	• Testing hypotheses • Testing of hypotheses with research designs
Changing models	• Alterations to improve the model object • Alterations when there are errors in the model object • Alterations when basic • requirements are not met	• Alterations when model does not match the original • Alterations due to new findings about the origina	• Alterations due to findings from model experiments • Alterations when the focus of the model shifts to a different aspect of the process

The cognitive framework consists of five aspects: (1) *nature of models*, (2) *multiple models*, (3) *purpose of models*, (4) *testing models*, and (5) *changing models*. The first two aspects (nature of models and multiple models) describe the *understanding* of models, while the latter three aspects (purpose, of models, testing models and changing models) reflect the understanding of the *use* of models in science. For each of these aspects, three levels are defined. The levels represent how well the students understand that models are products of science as well as methods employed by science. For the basic level, level I, students do not link the model with an existing original, it is solely an object that can be used (*model object*, Mahr, 2009). If students understand that the model represents the original, is created by people and can be used to communicate about the original, then the student has reached level II (*model of something*; Mahr, 2009). The most advanced level of model understanding, level III, represents the understanding that a model can be used as a method in science. A model can be tested and it can be used to generate observations and conclusions about the original that the model represents (*model for something;* Mahr, 2009). Next to the three levels of complexity, for some modeling-aspects an initial level of understanding is defined. Student responses fall in the category of initial level understanding when they refuse to agree with any level I, II or III statements. An initial level of understanding can be seen as lower than level I understanding.

The first aspect, *nature of models*, is about the comparison of the model to the real world original. The level of the students' understanding is indicated by how comparable they think the model and the original are. The second aspect, *multiple models*, is about the understanding of how different models can represent the same original. If students refuse the existance of multiple models for the same original, they have an initial level of understanding. The third aspect, *purpose of models*, is about how the model is used. The fourth aspect, *testing models*, is about testing the correctness of the model. If students think that no testing is needed or that authorities provide the absolute thruth, they only have an initital level of understanding. The fifth aspect, *changing models*, is about the possible reasons for changing a certain model. If students think there is no reason to make alterations, they again have an initial level of understanding.

According to Grünkorn et al. (2014), the aspects nature of models and multiple models are about the ontological and epistemological *understanding* of models, while the other three aspects (testing models, changing models and purpose of models) are about reflecting on how models are *used*. Furthermore, level I understanding of both nature of models and multiple models is an incorrect understanding of modeling. There are no circumstances in which a model is a copy of the original (level I understanding of nature of models) or in which multiple models exist because there are multiple versions of the original (level I understanding of multiple models). Therefore it is important, especially for the aspects *nature of models* and *multiple models* that students reach at least a level II understanding. In our study we will focus on these two modeling-aspects.

2.3 Experimental findings on meta-modeling knowledge

Grünkorn et al. (2014) used the framework by Upmeier zu Belzen and Krüger (2010) to assess the meta-modeling knowledge of secondary school students. In their study, they only used scale models (e.g. the skull of a Neanderthal man). The results showed that for each aspect, there were more level I and level II responses than level III responses. This indicates that most students understand that a model is a representation *of* something (level II), but not a hypothetical construct to use *for* something (level III). Gogolin and Krüger (2016) conducted another study using the same framework. They investigated the meta-modeling knowledge of upper secondary school students with a forced choice task. Similar to the previous study, no distinction between scale models and concept-process models is made. The models that are used are mostly scale models, but they also use one concept-process model. Their results confirmed that students showed more level I and II responses than level III responses. Specifically, they found that the majority of the students was on level II for nature of models and on level I for purpose of models. Another important finding in their study was that the responses for the nature of models differed greatly for different models. Since these studies focussed on the students' meta-modeling understanding of scale models, our study will focus on concept-process models.

Krell (2019) proposes that the context in which models are given to students is much more important than has been assumed in previous research. He argues that it is not relevant to measure a general level of model understanding, but that researchers should look at the consistency between the students' understanding and the modeling-purpose. The aim of his research is to find how the students' meta-modeling knowledge differs depending on the given modeling-purpose. He uses three different modeling-purposes: aesthetic purpose (corresponding to level I understanding), explanatory purpose (corresponding to level II understanding) and a research tool purpose (corresponding to level III understanding). The models that are used are scale models. Krell (2019) found that in general most students agreed with statements corresponding to level II meta-modeling knowledge. However, if different modeling-purposes are compared, students agree more often with a level III statement when it is presented in a level III modeling-purpose task and that students agree more often with a level I statement when it is presented in a level I modeling-purpose task. This indicates that the meta-modeling knowledge displayed by students depends on the modeling-purpose of the presented model. This implies that when a model is presented without a clear modeling-purpose, the student can create their own modeling-purpose, which can promote an answer that represents a lower meta-modeling knowledge (level I or II) while the student is capable of level III understanding. Providing a modeling-purpose is thus essential to find out whether students are capable of level III understanding. Therefore, in our study we will take the modeling-purpose into account to assess students' meta-modeling knowledge of concept-process models.

Previous research has focused on students' meta-modeling understanding of scale models. Jansen et al. (2019) have used semi-structured interviews to specifically gain a deeper insight in the understanding of concept-process models by students. The results show two types of differences between students' understanding. In the first place, similar to Gogolin and Krüger (2016), they found a difference between different models. In this study, the difference between the models is the organizational level that they represent. Students' show a higher level meta-modeling understanding for the (sub)microscopic model than for the macroscopic model. In the second place, they find that students from pre-university education show a higher level of understanding than students from higher general secondary education. In our study we consider a possible difference between macroscopic and (sub)microscopic models and we will use a more homogenous student group.

2.4 Research questions

The aim of the present study is to describe the current understanding of biological concept-process models by secondary school students, in regard to the meta-modeling knowledge aspects *nature of models* and *multiple models*. Previous research has shown that an explicit modeling-purpose influences the students' displayed meta-modeling knowledge. However, in models that students encounter at school or in society, an explicit modeling-purpose is not always present. Therefore, we will first investigate the students' meta-modeling knowledge of concept-process models without an explicit modeling-purpose.

RQ1: What is students' meta-modeling knowledge of concept-process models for the aspects *nature of models* and *multiple models* if no explicit modeling-purpose is present?

Without a modeling-purpose present, students cannot be expected to display their highest possible meta-modeling knowledge. This, however, can be triggered by providing a modeling-purpose. Previous research has shown that a modeling-purpose can trigger students to display meta-modeling knowledge that matches the modeling-purpose. Therefore, we will investigate how an explicit modeling-purpose influences the students' meta-modeling knowledge scores.

RQ2: How does students' meta-modeling knowledge of concept-process models for the aspects *nature of models* and *multiple models* differ in the presence of an explicit aesthetic, explanatory or research tool modeling-purpose?

Previous research has also suggested that the students' meta-modeling knowledge of concept-process models can differ, depending on the biological level that the model belongs to. We will investigate whether there is an actual difference between the students' meta-modeling knowledge regarding models that are visible to the naked eye (macroscopic level) and models that are invisible to the naked eye (microscopic/submicroscopic level).

RQ3: How does students' meta-modeling knowledge of concept-process models for the aspects *nature of models* and *multiple models* differ for models on the macroscopic level or on the (sub)microscopic level?

3. Methods

In order to answer the research questions, a quantitative study using an assignment with closed question tasks was conducted. For each closed question task, there are three variables:

1. **Model type**
   macroscopic, or (sub)microscopic.
2. **Modeling-purpose**
   no explicit purpose, aesthetic purpose (level I), explanatory purpose (level II), or research tool purpose (level III).
3. **Modeling-aspect of meta-modeling knowledge**
   nature of models, or multiple models.

3.1 Participants

430 Dutch secondary school students in the 11th grade (16-18 years old) from 16 schools in different areas in The Netherlands participated in the study. 43 students (10%) did not complete the entire assignment, therefore only 387 students were included in the study. All students majored in biology. Participating schools were recruited via an announcement. For the areas in The Netherlands that were not yet represented, the researchers contacted schools personally. Participation was on a voluntary basis.

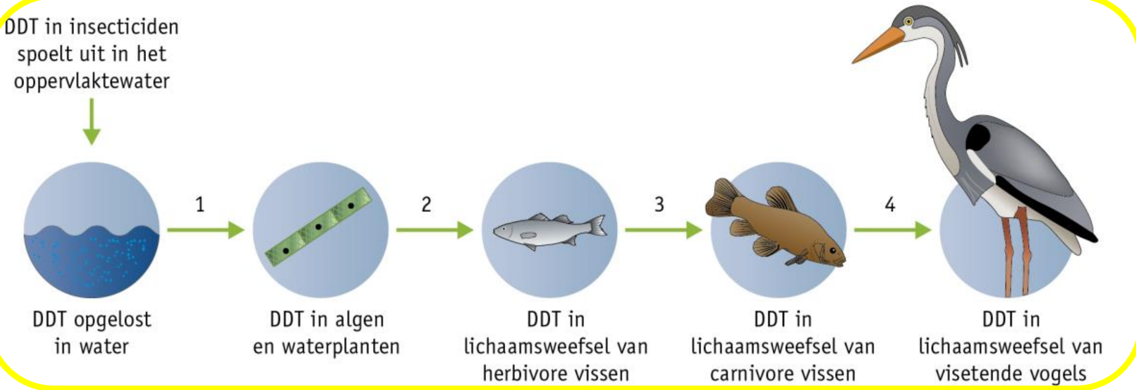
3.2 Instruments

To measure the meta-modeling knowledge, we developed an online assignment consisting of closed-question tasks. These tasks are different from the forced-choice tasks that were used in previous research (Gogolin and Krüger, 2018). The disadvantage of a forced-choice task is that students can only display one level of understanding. Thus, the result of a forced-choice task only shows the preferred level of understanding. We deviate from the forced-choice task since we are interested in all levels of understanding that the student is capable of, rather than only their preferred level of understanding. Therefore, we developed closed-question tasks.

In the closed question tasks, the participants were shown a model in context with corresponding statements (see figure 3). Each task included three statements, and for each statement the participant had to agree or disagree. The statements in the closed question task were related to a specific level of understanding (level I, II or III). Each of these statements was created using the descriptive framework by Jansen, Knippels and Van Joolingen (2019). In the context, the modeling-purpose and the modeling-aspect were identified (see table 2 and 3 for examples).

Example closed question task

Frank does research on water pollution. He discovered that in areas where DDT (a chemical insecticide) pollutes the water, there is higher mortality among fish-eating birds. He creates the model below.



Using this model, Frank wants to investigate whether water pollution by DDT is the cause of mortality among fish-eating birds.

Maud also researches water pollution by DDT. She makes a different model from the same process. When is Maud's model truly different from Frank's?

If Maud creates a model for an area with DDT pollution where other fish species appear.

[ ] Yes

[ ] No

If Maud elaborates on step 2 of the process.

[ ] Yes

[ ] No

If Maud has a different assumption about the route via which DDT ends up in the fish-eating birds.

[ ] Yes

[ ] No

Figure 3 – An example of a closed question task – The task starts with an introduction (red). Then the model itself is shown (yellow), followed by the modeling-purpose (green) and the modeling-aspect (purple). The statement items (orange) each correspond to a level of meta-modeling knowledge. The levels of the statements are in the order: level I, level II and level III. – This specific task has a *macroscopic model* with a *research-tool modeling purpose* and assesses the modeling-aspect *multiple models*.

Table 2 – Example sentences that identify the modeling-purpose.

Modeling-purpose	Example sentence
no explicit purpose	-
aesthetic purpose (level I)	Frank wants to hang this model on his office wall to remind him how much he liked his research.
explanatory purpose (level II)	Using this model, Frank wants to explain to his colleagues how water pollution can lead to mortality among fish-eating birds.
research tool purpose (level III)	Using this model, Frank wants to investigate whether water pollution by DDT is the cause of mortality among fish-eating birds.

Table 3 – Example sentences that identify the modeling-aspect.

Modeling-aspect	Example sentence
nature of models	Frank discusses the model with his colleagues. One of them asks what the models actually shows. How can Frank tell his colleague what the model shows?
multiple models	Maud also researches water pollution by DDT. She makes a different model from the same process. When is Maud's model truly different from Frank's?

Six different models were used to create the assignment (divided over two model types: 3 macroscopic models and 3 (sub)microscopic models, see appendix), four different modeling-purposes and two aspects of meta-modeling knowledge. This leads to 6 x 4 x 2 = 48 tasks in total.

Each participant was presented with 12 closed question tasks (see table 4 for an example of a complete assignment). Half of these tasks included a macroscopic model, and the other half of these tasks included an (sub)microscopic model. Furthermore, six tasks were presented with no explicit modeling-purpose, two tasks with a level I modeling-purpose, two tasks with a level II modeling-purpose and two tasks with a level III modeling-purpose. Since we used six different models, participants encountered each model twice: once with a non-explicit modeling-purpose (set 1) and once with an explicit modeling purpose (set 2). In order to prevent an influence of explicit modeling-purposes on the non-explicit modeling purpose tasks, all tasks in set 1 were presented before all tasks in set 2. Within each set, the order of the tasks was randomized to prevent order effects.

Table 4 – An example of a complete assignment for one participant

Task	Model	Modeling-purpose	Modeling-aspect
1	1: macroscopic	none	nature of models
2	2: macroscopic	none	multiple models
3	3: macroscopic	none	nature of models
4	4: (sub)microscopic	none	multiple models
5	5: (sub)microscopic	none	nature of models
6	6: (sub)microscopic	none	multiple models
7	1: macroscopic	aesthetic	multiple models
8	2: macroscopic	explanatory	nature of models
9	3: macroscopic	research tool	multiple models
10	4: (sub)microscopic	aesthetic	nature of models
11	5: (sub)microscopic	explanatory	multiple models
12	6: (sub)microscopic	research tool	nature of models

3.3 Pilot

Before conducting the full experiment, a pilot with nine pre-university education students (6 female, 3 male) has been executed. In this pilot the participants were asked to complete the task individually. After they completed the task, a semi-structured interview was conducted. The goal of the interview was four-fold. The first goal was to find out whether the tasks were clear and understood by the participants. The second goal was to determine how the participants reasoned to select the answers. The third goal of the interview was to observe whether the previous tasks influenced the participant's answers on the following tasks. The fourth and last goal was to find out whether the length and the number of questions was adequate for the participants.

The outcome of the pilot resulted in minor linguistic phrases, the use of personal names instead of function titles (e.g. *Lisa* instead of *the researcher*) and the location of the modeling purpose (below the picture instead of above it). Even though the participants mentioned that they were confronted with each model twice and noticed that the accompanying text was different, we decided to make this more explicit in the full experiment. Therefore, a screen was added before the models appeared for the second time that makes students aware that the models would be repeated and that they should read the text carefully.

3.4 Procedure

All participants received a short introduction about the procedure of the experiment by the researcher (either in real-life or with a short video-clip), after which the participants were given the opportunity to ask questions. Then, the assignment was administered to the participants digitally via laptops or tablets. Because we wanted the students to be able to see every detail of the models that were presented, we only used large screens, meaning that no mobile phones were used.

3.5 Analysis

Each answered statement was given a score (1 for agree / yes and 0 for disagree / no). The mean score of agreement with each statement level is calculated and ranges between 0 (total disagreement) and 1 (total agreement).

To answer RQ1, only the tasks without an explicit modeling-purpose have been considered. In order to find out whether there are differences in the students' agreement between level I, level II or level III statements a one-way ANOVA was used to compare the mean agreement scores between the three statement levels.

To answer RQ2, all tasks will be considered. For each of the modeling-aspects, a two-way ANOVA was used to investigate

1) the effect of the *statement level* (level I, level II or level III),
2) the effect of the *modeling-purpose* (none, aesthetic, explanatory or research tool), and
3) the interaction effect

on the participants' agreement scores. If there would be an effect of statement level, it indicates that the statement level influences how many students agree with the statement. If there would be an effect of the modeling-purpose, it indicates that the sentence indicating the modeling-purpose influences how many students agree with the statement, irrespective of the statement level. If there would be an interaction effect, it indicates that the agreement scores depend on which statement level is used in which modeling-purpose. In other words, the interaction effect represents whether students are more likely to agree with a level II statement when it was provided in a level II modeling-purpose than when it was provided in a level III modeling-purpose.

For RQ3, a two-way ANOVA was used to investigate

1) the effect of *statement level* (level I, level II or level III),
2) the effect of *model type* (macroscopic or (sub)microscopic), and
3) the interaction effect

on the participants' agreement scores. If there would be an effect of statement level, it indicates that the statement level influences how many students agree with the statement, irrespective of the model type. If there would be an effect of model type, it indicates that the model type influences how many students agree with the statement, irrespective of statement level. If there would be an interaction effect, it indicates that the agreement scores depend on which statement level is used in which model type.

For all three research questions it was suitable to use an ANOVA, since the dependent variable (mean agreement scores) was continuous and all independent variables (statement level; modeling purpose; model type) were categorical. If an effect of an independent variable with more than two categories (statement level; modeling-purpose) was significant, a Bonferroni post-hoc test was used to determine which categories differed from each other.

4. Results

4.1 Meta-modeling knowledge with no explicit modeling-purpose (RQ1)

The mean agreement scores of the participants for the tasks in which no explicit modeling-purpose is present are shown in table 5. For the modeling-aspect *nature of models*, there is a significant effect of *statement level* on the mean agreement scores, $F(2, 1158) = 54.399$, $p < .001$. Post-hoc tests (Bonferroni) show that there are significant differences between level I (M = 0.76) and level III (0.60) and between level II (0.80) and level III. There is no significant difference between level I and level II. Thus, for the modeling-aspect *nature of models* students agree most with level I and II statements and least with level III statements.

For the modeling-aspect *multiple models*, there is a significant effect of *statement level* on the mean agreement scores, $F(2,1158) = 506.161$, $p < .001$. Post-hoc tests show that there are significant differences between level I (M = 0.19) and level II (M = 0.53), between level I and level III (M = 0.84) and between level II and level III. Thus, for the modeling-aspect *multiple models* students agree most with level III statements, then level II statements and least with level I statements.

Table 5 – Mean agreement scores for the tasks with no explicit modeling purpose.

	Nature of models		**Multiple models**	
Statement level	**Mean**	**SD**	**Mean**	**SD**
Level I	0.76	0.25	0.19	0.22
Level II	0.80	0.24	0.53	0.37
Level III	0.60	0.33	0.84	0.23

4.2 Meta-modeling knowledge with explicitly given modeling-purposes (RQ2)

The mean agreement scores of the participants for the tasks with either an aesthetic modeling purpose (level I purpose), an explanatory purpose (level II purpose) or a research tool purpose (level III purpose) are shown in table 6 and 7.

For the modeling-aspect *nature of models*, there is a significant effect of *statement level* on the mean agreement scores, $F(2, 4632) = 67.994$, $p < .001$. There is no significant effect of *modeling-purpose* on the mean agreement scores, $F(2, 4632) = 0.465$, $p = .707$. Post-hoc tests (Bonferroni) show that there are significant differences between level I and level II, between level I and level III and between level II and level III. Thus, for the modeling-aspect *nature of models* the agreement scores are determined by the statement level, but not by the modeling-purposes.

For the modeling-aspect *multiple models*, there is a significant effect of *statement level* on the agreement scores, $F(2, 4632) = 957.340$, $p < .001$ and a significant effect of *modeling-purpose* on the agreement scores, $F(3, 4632) = 2.880$, $p = .035$. The interaction effect is not significant, $F(6, 4632) = 0.765$, $p = .597$. Post-hoc tests (Bonferroni) show that for *statement level* there are significant differences between level I and level II, between level I and level III and between level II and level III.

Table 6 – Mean agreement scores for modeling-aspect nature of models in different modeling-purposes.

Statement level	Aesthetic		Explanatory		Research tool	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Level I	0.80	0.40	0.72	0.45	0.75	0.43
Level II	0.78	0.41	0.81	0.39	0.79	0.41
Level III	0.63	0.48	0.65	0.48	0.66	0.48

Table 7 – Mean agreement scores for modeling-aspect multiple models in different modeling-purposes.

Statement level	Aesthetic		Explanatory		Research tool	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Level I	0.22	0.42	0.26	0.44	0.26	0.44
Level II	0.53	0.50	0.56	0.50	0.57	0.50
Level III	0.87	0.34	0.84	0.36	0.88	0.33

4.3 Meta-modeling knowledge for macroscopic and (sub)microscopic models (RQ3)

Since there is no interaction effect between the statement level and the modeling-purpose (RQ2), no distinction is made between tasks with different modeling-purposes in this section. The mean agreement scores of the participants for the tasks with either a macroscopic or (sub)microscopic model are shown in table 8 and 9.

For the modeling-aspect *nature of models*, there is a significant effect of *statement level* on the agreement scores, $F(2, 2316) = 54.006$, $p < .001$, and a significant effect of *model-type* on the agreement scores, $F(1, 2316) = 10.664$, $p = .001$. The interaction effect is also significant, $F(2, 2316) = 42.365$, $p < .001$. Post-hoc tests show that there are significant differences between level I and level II, between level I and level III, and between level II and level III. Thus, for the modeling-aspect *nature of models* the agreement scores are determined by the statement level, the model-type and the interaction between the two. The mean agreement scores for level I statements are higher for macroscopic models than for (sub)microscopic models, while this difference is not observed for level II and level III statements (figure 4).

For the modeling-aspect *multiple models*, there is a significant effect of *statement level* on the agreement scores, $F(2, 2316) = 745.813$, $p < .001$. There is no significant effect of *model-type* on the agreement scores, $F(2, 2316) = 0.026$, $p = .872$. Post-hoc tests show that there are significant differences between level I and level II, between level I and level III, and between level II and level III.

Table 8 – Mean agreement scores for modeling-aspect nature of models for different model-types.

Statement level	Macroscopic model		(Sub)microscopic model	
	Mean	**SD**	**Mean**	**SD**
Level I	0.80	0.28	0.60	0.32
Level II	0.73	0.30	0.75	0.39
Level III	0.55	0.32	0.62	0.37

Table 9 – Mean agreement scores for modeling-aspect multiple models for different model-types.

Statement level	Macroscopic model		(Sub)microscopic model	
	Mean	**SD**	**Mean**	**SD**
Level I	0.15	0.21	0.27	0.24
Level II	0.53	0.38	0.48	0.37
Level III	0.84	0.27	0.77	0.28

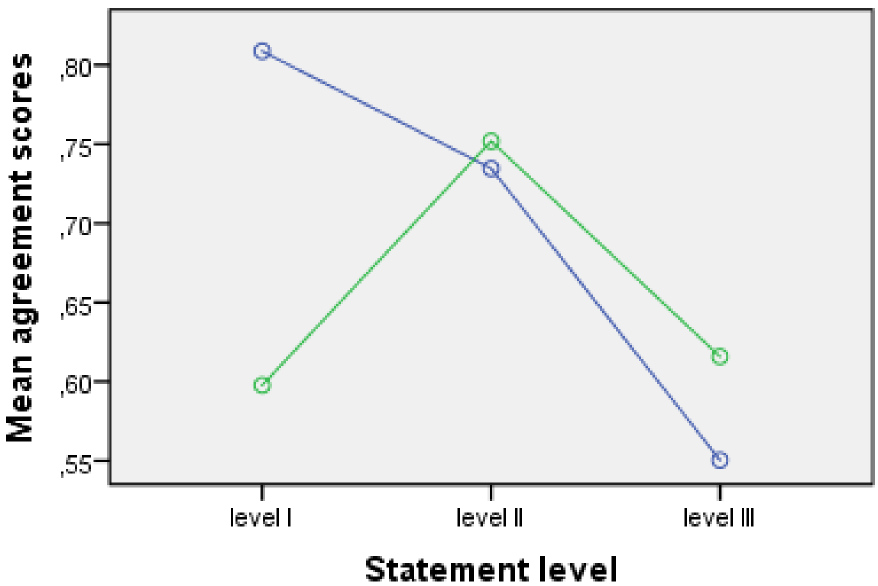


Figure 4 – Interaction effect between *statement level* and *modeling-type* on the mean agreement scores. Blue line = macroscopic models. Green line = (sub)microscopic models.

4.4 Exploratory analyses

In order to obtain more insight into individual answer behavior, the answer patterns of the participants are analyzed. For each modeling-aspect and each modeling-purpose, table 10 and 11 show the answer patterns of the participants. An answer pattern is the agreement pattern for the three statements of one closed question task (see figure 3). If a student agrees with statement levels I and II, but not with statement level III, then the answer pattern is agree-agree-disagree, which corresponds with pattern F in table 10 and 11. Eight different answer patterns are possible (level I: disagree/agree, level II: disagree/agree, level III: disagree/agree; thus 2 x 2 x 2 = 8 patterns).

For the modeling-aspect *nature of models*, there are two most common patterns. The first most common pattern (34 – 39%) is pattern H, which means agreeing with all statements. The second most common pattern (21 – 27% of all answers) is pattern F, which means agreeing with level I and level II statements, while disagreeing with level III statements. The least common pattern is pattern A, meaning disagreement with all statements (1 – 2 % of all answers).

For the modeling-aspect *multiple models*, there are also two most common patterns. The first most common pattern here (34 – 38%) is pattern C, which means disagreeing with level I statements and agreeing with level II and level III statements. The second most common pattern (28 – 33%) is pattern D, which means disagreeing with level I and level II statements and agreeing with level III statements. The least common pattern (1 – 2%) is pattern E, meaning agreement with level I statements and disagreement with level II and level III statements.

Table 10 – Answer patterns for the modeling-aspect *nature of models* – For each modeling-purpose the percentage of participants that answered according to each pattern is shown, - indicates disagreement, + indicates agreement.

Answer pattern \ Modeling-purpose				No explicit purpose	Aesthetic purpose	Explanatory purpose	Research tool purpose
	Level I	Level II	Level III	Participants (%)	Participants (%)	Participants (%)	Participants (%)
A	-	-	-	1	1	2	1
B	-	+	-	8	4	7	8
C	-	+	+	12	11	15	11
D	-	-	+	3	5	4	5
E	+	-	-	6	6	5	3
F	+	+	-	26	27	21	23
G	+	-	+	10	11	8	13
H	+	+	+	34	36	39	36

Table 11 – Answer patterns for the modeling-aspect *multiple models* – For each modeling-purpose the percentage of participants that answered according to each pattern is shown, - indicates disagreement, + indicates agreement.

Answer pattern \ Modeling-purpose				No explicit purpose	Aesthetic purpose	Explanatory purpose	Research tool purpose
	Level I	Level II	Level III	Participants (%)	Participants (%)	Participants (%)	Participants (%)
A	-	-	-	2	3	4	2
B	-	+	-	9	6	7	6
C	-	+	+	35	36	34	38
D	-	-	+	33	32	28	28
E	+	-	-	2	1	1	1
F	+	+	-	2	2	4	3
G	+	-	+	10	10	12	11
H	+	+	+	6	9	9	11

5. Conclusion and discussion

This study investigated students' meta-modeling knowledge of biological concept-process models, using the modeling-aspects *nature of models* and *multiple models* of the theoretical framework by Jansen et al. (2019). This study addressed the following three research questions.

RQ1: What is students' meta-modeling knowledge of concept-process models for the aspects *nature of models* and *multiple models* if no explicit modeling-purpose is present?

RQ2: How does students' meta-modeling knowledge of concept-process models for the aspects *nature of models* and *multiple models* differ in the presence of an explicit aesthetic, explanatory or research tool modeling-purpose?

RQ3: How does students' meta-modeling knowledge of concept-process models for the aspects *nature of models* and *multiple models* differ for models on the macroscopic level or on the (sub)microscopic level?

The results for RQ1 show that when students are presented with biological concept-process models without an explicit modeling-purpose, they mostly think on level I and level II for the modeling-aspect *nature of models* and they mostly think on level III for the modeling-aspect *multiple models*. For RQ2 it was found that the students' meta-modeling knowledge of concept-process models for the modeling-aspects *nature of models* and *multiple models* did not differ in the presence of an explicit aesthetic, explanatory or research tool modeling-purpose. Lastly, the results for RQ3 indicated that students' meta-modeling knowledge differs between models on the macroscopic level and models on the (sub)microscopic level. For the modeling-aspect *nature of models*, students mostly think on level I for the macroscopic models, but for the (sub)microscopic models they mostly think on level II. For macroscopic models, the students also reach level III thinking less often than for (sub)microscopic models. This difference between macroscopic models and (sub)microscopic models is not observed for the modeling-aspect *multiple models*.

It is noteworthy that students' level of understanding in this study reached level III more often than in previous studies. In this study 60 percent of the students reached level III for *nature of models* and 84 percent of the students reached level III for *multiple models*, when no explicit modeling-purpose was present. In contrast, Grünkorn et al. (2014) found that 4 and 9 percent of the students reached level III for *nature of models* and *multiple models*, respectively. Krell et al. (2015) found that 15 and 34 percent of the students reached level III for *nature of models* and *multiple models*, respectively. There are two possible explanations for this discrepancy. The first explanation is that the concept-process models in this study trigger a more advanced level of understanding than the scale models used by Grünkorn et al. (2014). Concept-process models are more complex and abstract than scale models, because they show processes rather than fixed states and they often show multiple processes in one model (Harrisson and Treagust, 2000). Therefore, interpreting concept-process models requires more abstract thinking. This can possibly trigger the more advanced level of meta-modeling understanding. The second explanation is that the difference is caused by the used method. In this study, students could display more than one level of understanding (i.e. a student could agree with more than one statement). In the study by Krell et al. (2015), students could only display one level of understanding (i.e. they were obliged to choose one statement). The exploratory analyses showed that many students made use of the opportunity to agree with more than one statement. Therefore, the majority of students might not prefer thinking on level III as shown by Krell et al. (2015), but that does not indicate that they can never display level III thinking. To determine which of the two explanations for the discrepancy is correct, further research is needed.

When students are presented with the same biological concept-process models in the presence of an explicit modeling-purpose, their level of understanding does not change depending on the level of the provided explicit modeling-purpose. This contradicts the research by Krell (2019) where they find that students' level of understanding changes depending on the provided modeling-purpose for the modeling-aspects *testing models* and *changing models*. A possible explanation for this contradiction is that the modeling-aspects *nature of models* and *multiple models* are inherently different from the modeling-aspects *testing models* and *changing models*. According to Grünkorn et al. (2014) the modeling-aspects *nature of models* and *multiple models* deal with the ontological and epistemological understanding of models, while the modeling-aspects *purpose of models*, *testing models* and *changing models* are about how models are used in science. The explicit modeling-purpose that is provided in the current study and the study by Krell (2019) reflects the modeling-aspect *purpose of models*. This modeling-aspect is more closely related to the other modeling-aspects about the use in science (*testing models* and *changing models*) than to the modeling-aspects about the ontological and epistemological understanding (*nature of models* and *multiple models*). This inherent difference between the modeling-aspects provides a possible explanation for why an explicit modeling-purpose influences the students' meta-modeling knowledge of *testing models* and *changing models*, but not the meta-modeling knowledge of *nature of models* and *multiple models.*

The results for the third research question showed a captivating difference between the students' meta-modeling knowledge for models on the macroscopic level and models on the (sub)microscopic level. Students showed lower level meta-modeling knowledge for macroscopic models than for (sub)microscopic models. This observation is in line with the qualitative research results by Jansen et al. (2019) that suggested more advanced meta-modeling knowledge for (sub)microscopic models than for macroscopic models. The discrepancy between meta-modeling knowledge for macroscopic and (sub)microscopic levels can be explained by a difference in abstractness. Macroscopic models are visible to the naked eye, which makes them less abstract than (sub)microscopic models that are not visible to the naked eye (Tsui and Treagust, 2013). The abstract thinking skills that are needed to interpret (sub)microscopic models can possibly trigger the abstract thinking skills that are needed to reach the more advanced levels of meta-modeling knowledge. This explanation resembles the explanation for the difference in meta-modeling knowledge between scale models and concept-process models. Thus, a single explanation of model abstractness and more advanced meta-modeling knowledge can provide a possible explanation for two separate observations in this study. However, it is important to realize that the difference between macroscopic and (sub)microscopic models is only found for the modeling-aspect *nature of models*, but not for the modeling-aspect *multiple models*. Further research should investigate whether this is a persistent difference and whether similar differences between macroscopic and (sub)microscopic models can be observed for other aspects of meta-modeling knowledge.

References

Burns, T. W., O'Connor, D. J., & Stocklmayer, S. M. (2003). Science communication: a contemporary definition. *Public understanding of science, 12*(2), 183-202.

College voor Toetsen en Examens. (2016). *Biologie VWO, Syllabus centraal examen 2018 (bij het nieuwe examenprogramma).*

Crawford, B., & Cullin, M. (2005). Dynamic assessments of preservice teachers' knowledge of models and modelling. In *Research and the quality of science education.* Dordrecht: Springer

Gogolin, S., & Krüger, D. (2018). Students' understanding of the nature and purpose of models. *Journal of Research in Science Teaching*, *24*, 357–368.

Grosslight, L., Unger, C., Jay, E., & Smith, C. L. (1991). Understanding models and their use in science: Conceptions of middle and high school students and experts. *Journal of Research in Science Teaching, 28*(9), 799 –822.

Grünkorn, J., Upmeier zu Belzen, A. U., & Krüger, D. (2014). Assessing students' understandings of biological models and their use in science to evaluate a theoretical framework. *International Journal of Science Education*, *36*(10), 1651-1684.

Harrison, A. G., & Treagust, D. F. (2000). A typology of school science models. *International Journal of Science Education*, *22*(9), 1011-1026.

Hodson, D. (2003). Time for action: Science education for an alternative future. *International Journal of Science Education, 25*(6), 645-670.

Jansen, S., Knippels, M. C. P., & van Joolingen, W. R. (2019). Assessing students' understanding of models of biological processes: a revised framework. *International Journal of Science Education*, *41*(8), 981-994.

Justi, R. S., & Gilbert, J. K. (2003). Teachers' views on the nature of models. *International Journal of Science Education, 25*, 1369–1386

Krell, M. *(2019).* Assessment of meta-modeling knowledge: Learning from triadic concepts of models in the philosophy of science. *Science Education Research Letters, 2019,* 1-7.

Krell, M., Reinisch, B., & Krüger, D. (2015). Analyzing students' understanding of models and modeling referring to the disciplines biology, chemistry, and physics. *Research in Science Education, 45*(3), 367-393.

Mahr, B. (2009). Information science and the logic of models. *Software & Systems Modeling*, *8*(3), 365-383.

Odenbaugh, J. (2005). Idealized, inaccurate but successful: A pragmatic approach to evaluating models in theoretical ecology. *Biology & Philosophy, 20*(2–3), 231–255.

Reece, J.B., Urry, L.A., Cain, M.L., Wasserman, S.A., Minorsky, P.V., & Jackson, R.B. (2011). *Campbell Biology.* San Fransisco: Pearson. (9th ed.).

Svoboda, J., & Passmore, C. (2013). The strategies of modeling in biology education. *Science & Education*, *22*(1), 119-142.

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., ... & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching, 46*(6), 632-654.

Tsui, C. Y., & Treagust, D. F. (2013). Introduction to multiple representations: Their importance in biology and biological education. In *Multiple representations in biological education* (3-18). Springer, Dordrecht.

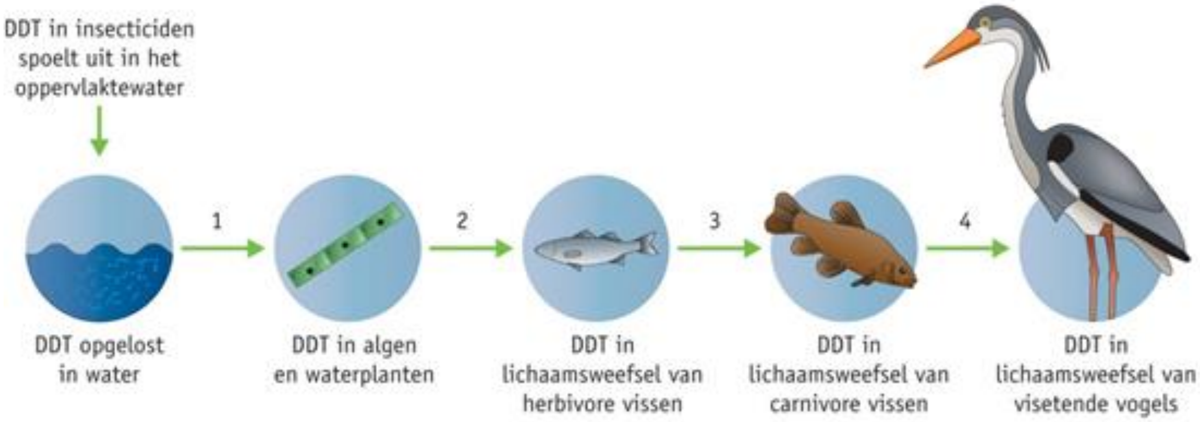
Upmeier zu Belzen, A., & Krüger, D. (2010). Modellkompetenz im Biologieunterricht [Model competence in biology education]. *Zeitschrift für Didaktik der Naturwissenschaften [Journal of science education], 16,* 41–57.

Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science Education, 92*, 941–967.

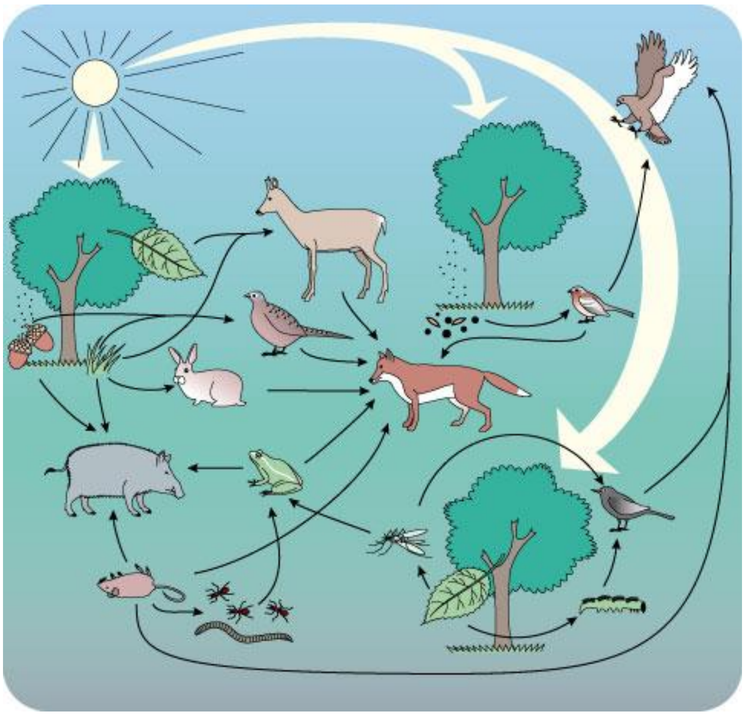
Appendix – Models used in the closed question tasks

Macroscopic models

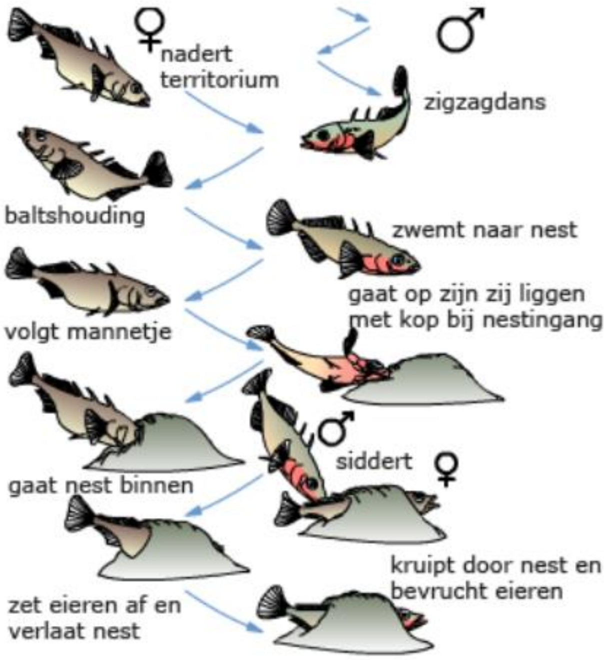
Model 1



Model 2

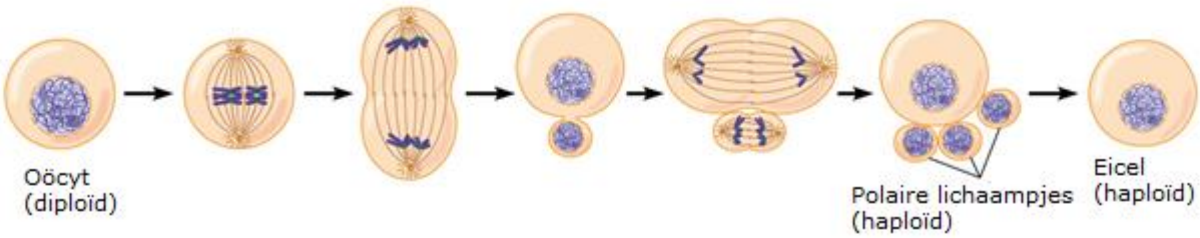


Model 3

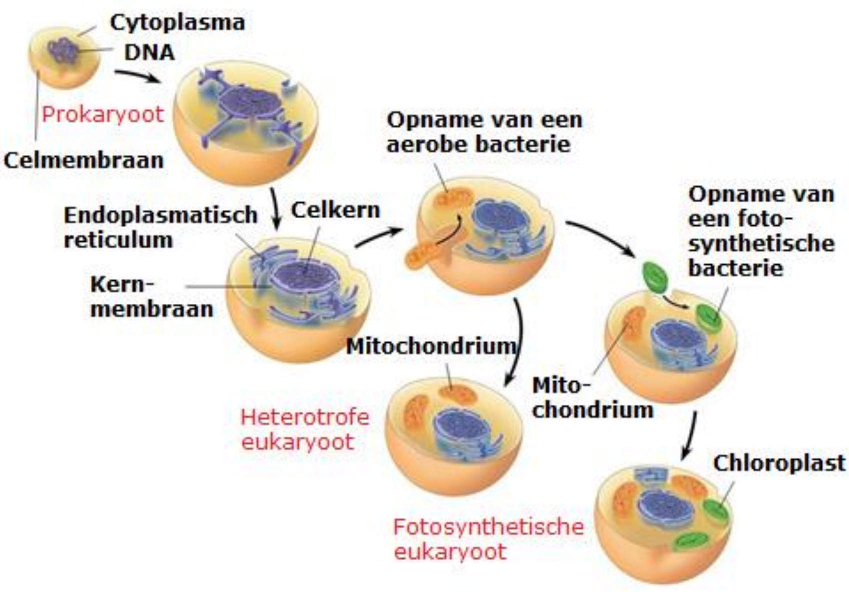


Microscopic models

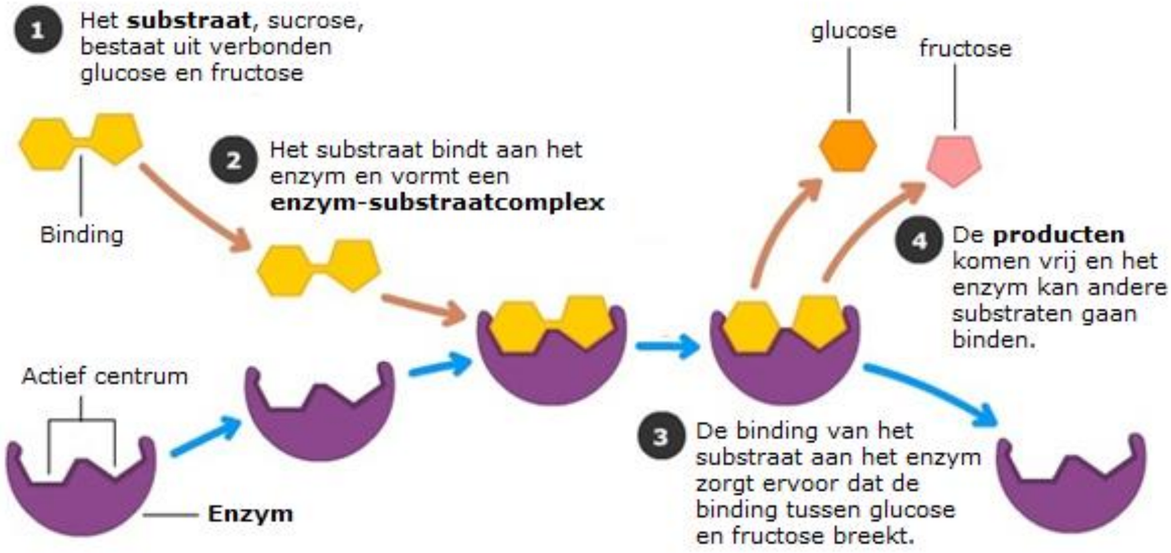
Model 4



Model 5



Model 6



1 Het **substraat**, sucrose, bestaat uit verbonden glucose en fructose
Binding
2 Het substraat bindt aan het enzym en vormt een **enzym-substraatcomplex**
Actief centrum
Enzym
glucose
fructose
3 De binding van het substraat aan het enzym zorgt ervoor dat de binding tussen glucose en fructose breekt.
4 De **producten** komen vrij en het enzym kan andere substraten gaan binden.