Systems Thinking in Biology: Validating an Assessment Method

Daphne van Geelen
62808070
Supervisor: Melde Gilissen
Second corrector: Wouter van Joolingen
University Utrecht

Abstract

Systems thinking (ST) is the ability to understand and interpret complex systems which is essential in biology education because it indicates biological reasoning. Until now, no general method of assessment was developed to measure the extent of students' ST based on this definition of ST. The aim of this research is to determine to what extent the designed assessment method visualizes the extent of students' ST in biology education, and to determine to what extent there is progress in students' ST after being taught about ST. In this research three assessment methods in the form of drawing tasks were developed, tested and analysed throughout one year on two different schools, to study if there was an increase in the students' ST. In one school, the students got lessons about ST (n= 47) and the other served as a control (n=41). ST was measured by the use of the seven characteristics of systems: boundary, hierarchy, components, interactions, input-output, feedback, dynamics. The results suggest that the test could measure the extent of students' ST, apart from the characteristic of boundary, and feedback in the second test. The results show that there was an increase in the explicit use of systems language in the experiment group, although no solid conclusions could be made from the results of the drawings due to differences in teaching methods. These outcomes show, that although improvements are necessary, the test showed potential in measuring students' ST.

Key concepts

Systems thinking, biology education, assessment methods, crosscutting concepts, systemic reasoning, drawing tasks

Introduction

Biology is a science that works on several different levels. It works from a microscopical level to an organism level, a population level, and to an ecosystems level and back (Verhoeff, 2003; Knippels & Waarlo, 2018; Verhoeff, Knippels, Gilissen & Boersma, 2018). These different levels are all connected, from one system to another, and are hierarchically ordered. This is the so-called yo-yo strategy (Knippels & Waarlo, 2018) and is an approach within ST. To connect these different levels and to understand the different interactions and the complex relationships they have, ST is required. ST is defined as the ability to understand and interpret complex systems (Evagorou, Korfiatis, Nicolaou, & Constantinou, 2009). Although this definition is broad, ST can be seen as an important cognitive skill, to figure out the relations between different components and the bigger picture of the system (Cox et al., 2018). Gilissen et al. (2019) specify that ST is essential in biology education because it indicates biological reasoning.

In current education, biology is divided in separate themes and students find it a challenge to identify the connection between these different themes and the different levels at which they work (Tripto, Assaraf, & Amit, 2013; Verhoeff, Boerwinkel, & Waarlo, 2009). The National Research Council (2011), the Next Generation Science Standard (NGSS, 2013) and even the Dutch commission for renewal of biology education (Boersma, Waarlo, & Klaassen, 2011) underline the importance of ST in biology education. ST is also included in the examination programs for Biology in the Netherlands as a domain-specific skill (Boersma, 2010).

The study of Gilissen et al. (2019) found that teachers and teacher educators spend little time on ST in their education. The knowledge gap that Gilissen et al. (2019) define is that ST in current biology education is not fully applied by teachers and teacher educators, therefore a heuristic needs to be developed. They state that teachers and teacher educator should learn more about implementing ST in their education, to foster students' ST. Teachers and teacher educators can use the universal characteristics of systems to enhance students' ST. According to Gilissen et al. (2019), these characteristics of systems are derived from important aspects of ST:

- Identification of a system
- Universal characteristics of a system

- Visualisation of a system, or modelling

With the *identification of system of interest*, the **boundaries** of a system need to be identified and thereby the **components** that include the system. Gilissen, Knippels, & van Joolingen (submitted) state that, although a boundary is important for the selection of the system, zooming in and out is important as well in understanding a system on multiple levels.

In biology systems, certain perspectives are considered when it comes to describing a system. These perspectives are part of the *identification of the universal characteristics of biological systems* (Gilissen et al., 2019):

- **Input – output** of a system
- Self-regulation or **feedback** mechanisms in a system
- Levels of biological organisation or **hierarchy** in a system
- Development of a system, or **dynamics**
- **Emergence** of a system were components of a system work together to achieve a goal on a different organisational level.

Gilissen et al. (submitted) argue that *modelling* helps the students to visualize a system and how the characteristics relate to each other. Whether the students create their own model or whether they learn to interpret an existing model. Models can be distinguished in two categories: qualitative models and quantitative models. Qualitative models represent a system in an abstract approach. It is used for thinking forward and backward between systems (Verhoeff et al., 2018), therefore overseeing the **interactions** between and in a system. Quantitative models focus more on the quantitative changes in a system (Hmelo-Silver et al., 2017; Verhoeff et al., 2018). Gilissen et al. (submitted) argue that for understanding ST in biology, qualitative models are easier to understand than quantitative ones, and they suggest starting with qualitative models in the beginning of teaching ST. For deeper understanding of the system, quantitative models could be introduced to students (Gilissen et al., submitted).

Based on these perspectives on ST, seven characteristics are defined to describe a system. These seven characteristics are summarized in table 1. They can be used to measure the level of students' ST. By using the characteristics, the students show that they understand what a system consist of and therefore use it in their defining systems. Gilissen et al (submitted) is conducting a larger research where they want to develop to a heuristic for teachers to implement ST in biology education. Through lesson studies they hope to find a method to imply ST in biology education, and to measure the extent of students' ST an assessment method is developed.

Table 1. Summary of the system characteristics (Gilissen et al., submitted)

System characteristic	Description
Boundary	A biological system can be defined by its boundary. This boundary can be concrete (cell membrane) or abstract (ecosystem)
Hierarchy	A biological system consists on multiple levels. These different levels the systems work as a subsystem of its own and are categorised on its organisational level.
Components	A biological system consists of different parts, which all have its own function
Interactions	These parts within a subsystem have all kinds of interaction with each other. Without those different parts and interactions, a function would not work.
Input output	A biological system is an open system. This means that an exchange happens between the environment and the system.
Feedback	A feedback mechanism has influence on the system at the time. A positive feedback enhances the changes in the system, where negative feedback reverses the changes in the system.
Dynamics	Each biological system changes over time. By these changes a system occurs in a certain state. These states are not always predictable.
Emergence	An overall characteristic of ST were different components of a system have a function on another organisational level. to understand the overall mechanism, one must descend to a lower organisational level and back to the highest level to understand the impact. Also referred as the yo-yo strategy (Knippels & Waarlo, 2018).

Different studies have developed methods to measure the extents of students' ST in biology, nevertheless these are about specific subjects and define ST in a different way. Therefore, no *general* method of assessment is developed on ST in biology. Gilissen et al. (submitted) has developed a pilot assessment method and in this thesis the designed assessment method is tested, analysed, and possible recommendations for improvements are given.

The aim of this research is (1) to determine to what extent the designed assessment method visualizes the extent of students' ST in biology education, and (2) to determine to what extent there is progress in students' ST after being taught about ST. Therefore, the research questions are:

1) *To what extent does the designed assessment method measures students' ST?*
2) *To what extent do students show progress in ST after being taught about ST?*

Theoretical Background

ST in Biology Education

To measure students' ST, different studies have been done (Brandstädter, et al., 2012; Dauer et al., 2013; Tripto et al., 2016; Tripto, Assaraf & Amit, 2018; Tripto, et al., 2013; Riess & Mischo, 2010; Danish et al., 2017; Hmelo-silver et al., 2017; Yoon & Hmelo-Silver, 2017). Most of them used a tool to measure the extent of students' ST, nevertheless due to different definitions of ST and due to the specific subjects, these assessment methods were not completely suitable for this study.

Tools to assess students' ST

The strengths and weaknesses of several tools that are used to measure the extent of students' ST were analysed in order to determine whether the developed drawing test in the current research was effective, reliable and valid in relation to the other tests.

Concept mapping. A lot of previous studies used concept mapping to visualize the extent of students' ST (Brandstädter et al., 2012; Dauer et al., 2013; Tripto et al., 2013; Tripto et al., 2016; Tripto et al., 2018). Concept maps give students the opportunity to organise their own knowledge (Novak, 1990; Tripto et al., 2013; Tripto et al., 2018). Concept maps can be a representation of the students' mind and the organization of the concepts within it, it can be a mental model (Tripto et al., 2013; Tripto et al., 2018). According to Tripto et al. (2013), ST, and concept mapping share the properties of structure, dynamism and hierarchy. Therefore, concept mapping can be a suitable tool to measure the extent of students' ST. One downside might be that concept mapping in its own is a skill to be mastered hence it should be taught to the students beforehand (Tripto et al., 2018).

Computer simulations. Computer simulations have the advantage to make connections between different phenomena and their mechanisms are more dynamic (Hmelo-Silver et al., 2017). They state that simulation models provide a rich context for productive reasoning, that could guide students' inquiry and their ST (Danish et al., 2017; Hmelo-Silver et al., 2017). Simulations also tend to provide more insight in the underlying mechanisms of a system and on a different scale (Hmelo-Silver et al., 2017). Therefore, the micro and macro levels could be easily explored and connections between the different levels are easier to understand. The downside of the use of computer simulations, is the need for a computer, the installation of the program and the instruction time to learn to use the simulation. For a large-scale research, it might be too time-consuming.

Questionnaires. Riess and Mischo (2010) used a questionnaire as assessment method to measure the extent of students' ST. In the questionnaire, they used multiple choice questions and open questions to elaborate on the models the students had to draw before. They also used a form of concept mapping in their questionnaire, but they already provided the students with the concepts. In the open questions the students could explain their previous answers, but only in words. With multiple choice questions it might be easy to process all the answers, but the mental model behind it is not visible. One also needs to take the gambling chance in mind when a student does not know the answer and just guesses it.

Drawing tasks. Scientific models of biological processes are often objectified in a pictorial representation, or a drawing, to make the process easier to understand. The models show a specific spatial arrangement of the process in a specific point in time, like the concept of meiosis (Kindfield, 1994). Kindfield stated that in biology instructions, diagrams, models and drawings often *'(…) are treated as little more than illustrative devices'* (1994). While the research of Maienschein (1991) claims that models could represent theoretical ideas and have a central role in understanding the underlying mechanisms. Kindfield (1994) found in her research that with the development of understanding the content (of meiosis) the pictorial skills of the student developed. Therefore, a more complete drawing of a model could help students understand the theory and the mechanisms behind it.

Method

To answer the research questions, an assessment method was developed and made by students of secondary schools. This assessment method should measure the extent of ST of these students. Two schools were involved in the research. A secondary school in Enschede and in

Nieuwegein. The students of the school in Enschede participate in a larger study on ST in biology. These students followed lessons about ST in biology. The school in Nieuwegein served as a control group to verify the validation of the assessment method and to measure if there were any differences between the extent of ST in students who had the lessons about ST and students who did not receive lessons about ST.

A school year is divided in four periods per year. Each period, students made a pre-test at the beginning, and the post-test at the end of a period. At least four weeks are in between the pre- and the post-test of each subject. The tests were therefore distributed throughout the year. The control group made each test only one time.

Designed assessment method

Due to different definitions of ST, these above-mentioned tests could not be directly used. Through drawings, students can expose their true understanding and conceptualisation of ideas and concepts (Dikmenli, 2010). That is why for this research an assessment method with a drawing task was developed. It has an easy set-up, no need for computers and can be taken anywhere with a pencil and a paper. With only simple drawings, teachers or researchers can have access to large amount of data of the mental models of ST (Dikmenli, 2010). Extra space was made to give some form of description of their systems drawing. Thus giving the students the opportunity to explain themselves and make their mental models more visible like Ries and Mischo (2010) did with their open tasks. The construction of the assessment method is explained in the section below.

Design of the test

Pre- and post-tests were conducted to measure the baseline and the development of ST in the students. Four tests were designed, each consist of four items and starts with six general questions like name, age, gender, class, teacher and school. Each test had its own subject, but the tests had the same structure. The subjects of the tests were:

- Respiration (in the lungs)
- Photosynthesis
- Digestion
- Ecosystem (not addressed in this study)

In the first item, the student was asked to draw everything they know about the given subject/process. For the example of respiration: *"Make a drawing as extensive as possible of everything you know about respiration and name everything you have drawn. You are not assessed on your drawing skills. The drawing must be understandable for a fellow student, so keep it readable."* The student was asked to name everything he/she had drawn. The drawing supports the completeness of the extent of students' ST (Dikmenli, 2010). In item two, the student had extra space to write down and elaborate on what he/she could not say in the drawings; hence, the students' mental model is becoming more visible (Ries and Mischo, 2010). Within these two items the characteristics of boundary, components, interactions and input-output were scored.

In item three, the student got a specific case (different in each subject) and he/she had to explain what the feedback mechanisms and the dynamics in this case were. The cases per subject were:

- Respiration: Difference in human respiration in sprint and rest
- Photosynthesis: Difference in growth/photosynthesis when one of two trees stands in the shadow of a house
- Digestion: The feeling of hunger and saturation and the fluctuation between these two.

In the fourth item, the student needed to make a connection between different systems. In the first test about the respiration system these different systems were given:

- Cell of a plant;
- A muscle;
- The digestion system;
- A cat;
- A forest.

In the second and third test, the students were asked to make a connection between the earlier subjects. Thus, in the second test about photosynthesis, the students were asked to make a connection between the human respiratory system and photosynthesis. In the third test the students were asked to make a connection between human respiration, photosynthesis and the human digestion. The tests can be found in Appendix A-C.

Selection of Participants

All participants were 4th grade senior general secondary education (4 havo) students. The students were approximately between 15-17 years old. In the Netherlands, the havo students have their exam after five years of secondary school. When testing the growth in ST on 4th graders, they already have most of the knowledge to connect multiple systems, but there is still improvement possible before the exams. Havo students have difficulty seeing coherence and abstraction and it should be good for them to learn more about ST.

The designed assessment method was a pilot, so no large groups of participants were needed. In each of the two schools, two classes of 4 havo students made the tests (resp. n=24, n=23, n=19, n=22). These participants were coded as experiment group and control group and divided in experiment group 1 and 2 and control group 1 and 2 when necessary.

Ethical Guidelines

In the research project, the guidelines of the Ethical Committee of the Faculty of Science (UU) were followed. Participants stayed anonymous in the study and received a code for their test, based on gender and location of school. The parents of the students filled in a consent form developed by the head researcher. Only the data from the students who gave consent have been used.

Data Collection

The test. A paper test had been developed and made by the students. The students made the test individually and without their books. When the students finished their test, they handed the test in at the teacher, without discussing it with other students.

Coding. Afterwards the test was coded by two coders with the help of a codebook (Dikmenli, 2010). The codebook was developed by two coders in different rounds.

In the code book, the coders found the descriptions of the characteristics that the test tests. These characteristics are determined in the research of Gilissen et al. (submitted) and are *boundary, components, interaction, input and output, feedback, dynamics and hierarchy*. The codebook was used to score the presence of the characteristics in the drawings of the students in Microsoft Excel with

- 0 (no presence of characteristic);
- 1 (complete presence of characteristic).

Furthermore, the coder recorded when students explicitly used the system characteristics in their drawings. When the students named one or more characteristics by name in their explanation, this was marked with a yellow cell at the characteristics mentioned and the specific line was written down.

The coders started by each taking the same five tests and analysed if they found the same characteristics in the tests. After the coding of the first five tests, the similarities and disagreements were discussed until they came to an agreement. These agreements were written down to form a code book. The coders took the next five tests to test the first code framework until they coded the tests the same without any disagreements according to the code framework. This was after 4 sessions of each five students.

Each test had its own subject and therefore its own codebook. There were some general rules in the codebook that applied to all the test, and these were written down at the beginning of the framework. Each characteristic had its own description. The codebooks can be found in appendix D-F

Validation codebook. To test the validation of the codebook, a third coder was asked to code a number of tests. To decide the number of test the third coder needs to code, the formula of Cicchetti (1976) was used. The formula states that to calculate the number of tests one must code is $2n^2$, where n is the number of variables tested. In our case the number of variables tested were the characteristics of ST, which were a total number of seven. This made the number of variables tested ninety-eight. To calculate the amount of test coded, we divided this number by the number of characteristics (because in every test all the characteristics were measured) and we got a number of fourteen tests to be coded. Because we had three subjects, we made a total of fifteen tests to be coded, five tests per subject. These tests were selected randomly over the pre- and post-tests, as it was only important to code the tests and the growth of the students' ST is not of interest. When the third coder coded the tests, he or she needed to be instructed. To do that, each coder got a "test run" with an extra test per subject. This made the total number of tests needed for a third coder eighteen, six per subject. An example of how to score the students' drawing, so item one and two, is given in figure 1.

Data Analysis

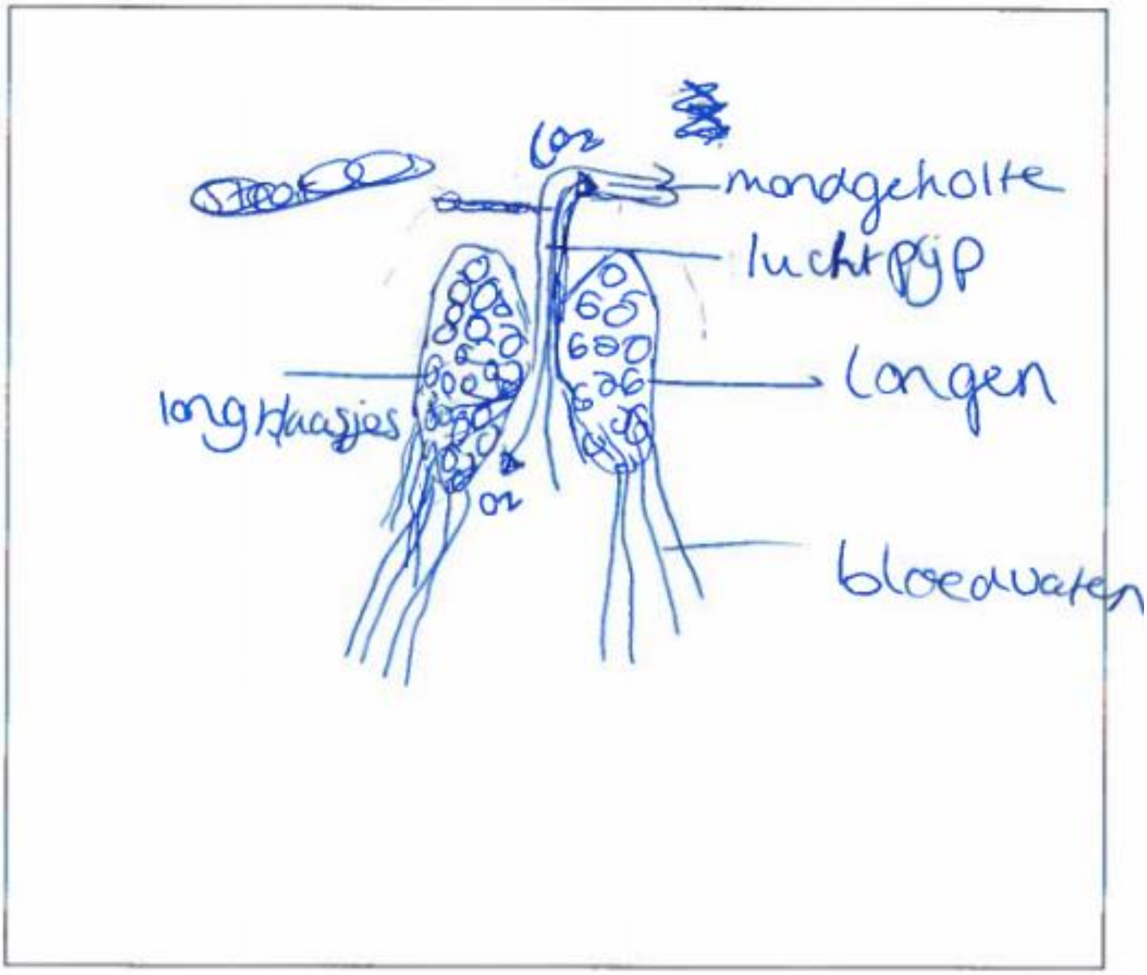
To test the interrater reliability of the third coder, a *Cohen's Kappa* test was calculated in IBM SPSS Statistics 25. The interrater reliability is the level of accuracy when two or more coders are coding data of a study (Lodico et al., 2010; Cox, et al., 2018; Gilissen et al., 2019). The codebook was proven to be moderate reliable when the Cohen's Kappa is between, 0.40 and 0.60, substantial reliable when the kappa is between 0.60 and 0.80 and excellent when exceeds 0.80 (Landis & Koch, 1977; Van der Ark & Ten Hove, 2018).

Considering the study was a pilot test, no statistical analysis was performed on the results of the test.

Results

The first thing that was noticed during the development of the codebook is that the characteristic 'boundary' could not be scored. The definition of the boundary of a system depends on one's own vision and is therefore subjective. Also, when students exceeded the boundary they automatically thought on multiple levels, and that is a quality of ST that you want to achieve. Due to the open question on item one and two no directions to the boundary were given and therefore it could not be measured and scored. This characteristic is not shown in the results. furthermore, in the second test about photosynthesis, the characteristic 'feedback' could not be scored due to the form of questioning in item three. The question was about two trees who stood next to each other. A house was built between the trees and now one tree stood into the shadow. The students had to explain what happened to the tree and show the dynamics and the feedback in the system. Only the sun influences the tree, whereas the tree does not influence the sun and therefore there is no feedback mechanism in this question.

Figure 1. Example of scoring item one and two.



Tekening ademhaling

1. Maak een zo uitgebreid mogelijke tekening van alles wat jij weet over ademhaling én benoem alles wat je hebt getekend. Je wordt niet beoordeeld op je tekenvaardigheden. De tekening moet begrijpbaar zijn voor een mede-leerling, dus houd het leesbaar.

2. Leg in het vak hieronder uit wat je hebt getekend.

ik heb getekend hoe ~~lucht~~ zuurstof ~~door~~ via de mondgeholte aan via de lucht pijp naar de longen gaan en ook weer terug met CO2 .

Components: at least lungs and windpipe are drawn, student also draws alveoli and mouth +1. Interactions: the oxygen and carbon dioxide flow through the windpipe to the lungs and back. The interaction of the input-output with different components is made visible +1. Input-output: oxygen and carbon dioxide flow through lungs and windpipe, visualised with arrows +1. Hierarchy: The student drew the alveoli, therefore zooming in, and the arteries are from a different system, therefore connects multiple systems together +1.

First the results of the experiment group were compared to the control group, after that the experiment group is split in the two different classes. They were compared if there are any difference between the classes. The explicit use of systems language in the tests was analysed in the experiment group to see whether students' make more use of systems language at the end of the schoolyear.

Interrater reliability

The interrater reliability calculated between the third coder and the first two coders was 0.821, which is excellent when following the guidelines from Landis & Koch (1977). Therefore, the procedure by coding the test with the codebook was proven to be reliable.

Test results experiment group versus control group

Due to different numbers of participants in each test, the results are shown in relative results. That means that the experiment group – with a score of 0.97 for components in the pre-test of respiration – 97 percent of the students were scored for that characteristic.

Table 2. Relative results experiment group per test per characteristics.

	Respiration (n=35)		Photosynthesis (n=40)		Digestion (n=40)	
Test	Pre	Post	Pre	Post	Pre	Post
Boundary	x	x	x	x	x	x
Components	0.97	0.97	0.53	0.73	0.93	0.98
Interactions	0.51	0.40	0.60	0.63	0.65	0.45
Input-output	0.49	0.34	0.73	0.75	0.48	0.25
Feedback	0.60	0.51	x	x	0.70	0.70
Dynamics	0.60	0.49	0.83	0.83	0.85	0.83
Hierarchy	0.83	0.77	0.48	0.48	0.55	0.58

The numbers in the table represent the relative numbers that are scored for the given characteristic for the experiment group. This means that for the experiment group in the pre-test of respiration, 0.97 of 97 percent of 35 students scored for the characteristic of components. A x indicates that the characteristics could not be scored.

In the results of the experiment group (table 2) you see that a high percentage of the group scored high in the first test on components (0.97) and on hierarchy (0.83). for the other characteristics in the first test, 49-60 percent got scored on the characteristics. Over the period of the tests the results fluctuated. Where the characteristics of components and hierarchy showed a decline in the second test about photosynthesis and an increase in the third test about digestion. While in the characteristic of input and output, there was an increase in the second test and a decline in the third test. The last test (post-test of digestion system) showed that 25 percent scored on that characteristic.

Table 3. Relative results control group per test per characteristics.

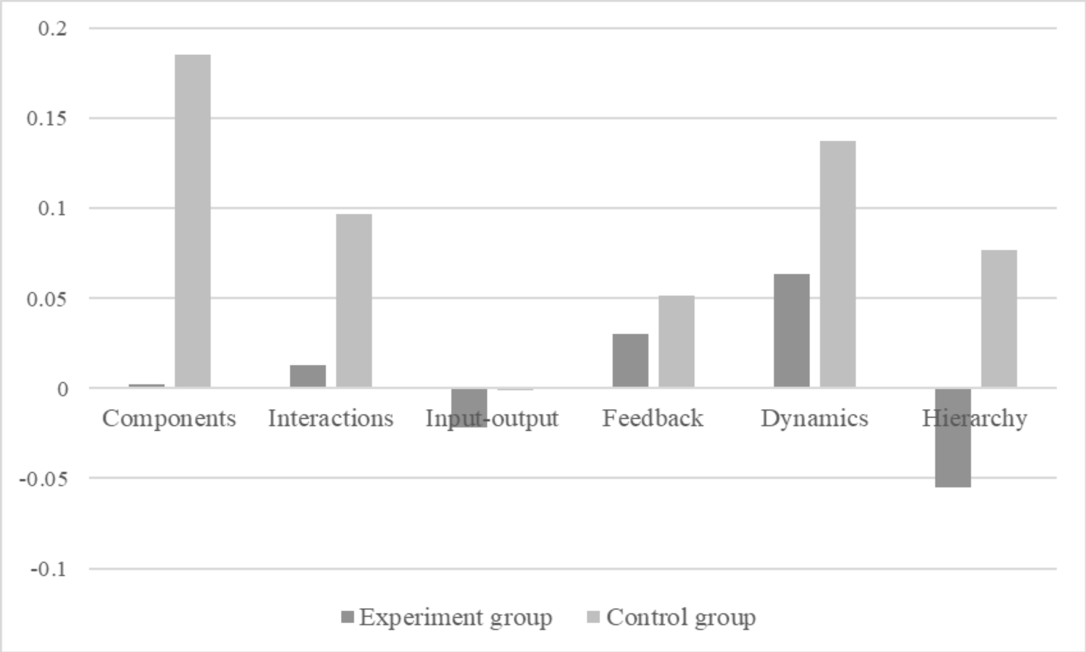
	Respiration (n=27)	Photosynthesis (n=27)	Digestion (n=37)
Boundary	x	x	x
Components	0.63	0.67	1.00
Interactions	0.48	0.74	0.68
Input-output	0.41	0.70	0.41
Feedback	0.52	x	0.62
Dynamics	0.48	0.93	0.76
Hierarchy	0.67	0.67	0.51

The numbers in the table represent the relative numbers that are scored for the given characteristic for the control group. This means that for the control group in the pre-test of respiration, 0.63 or 63 percent of 27 students scored for the characteristic of components. A x indicates that the characteristics could not be scored.

The relative results of the control group (table 3) shows that there was an increase in the number of students that got scored on the characteristics of components, interactions and dynamics. Even though they had no lessons about ST, the number of students that scored on the characteristics grew over time. For input and output there was an increase in number of students that got scored on the characteristic in the second test, but a decline in the third test, the same as in the experiment group.

To show if there is any progress in students' ST, by scoring them on the characteristics, the growth factor of the trendline of the characteristics was taken and visualised in graph 1. The graph shows that the experiment group has an increase in scoring for the characteristics of interaction, feedback and dynamics. The experiment group showed no increase or decline in the scoring of the characteristics of components and a decline in the characteristics input-output and hierarchy. The control group showed an increase in scoring on components, interactions, feedback and dynamics. They showed no increase or decline in scoring on input-output; however, they showed a decline in scoring on hierarchy.

Graph 1. Growth factor per characteristics experiment compared to control group



A positive number in the graph shows an increase in scoring of that given characteristics over all the tests. A negative number in the graph shows a decline in scoring of that given characteristics over all the tests.

Test result experiment group

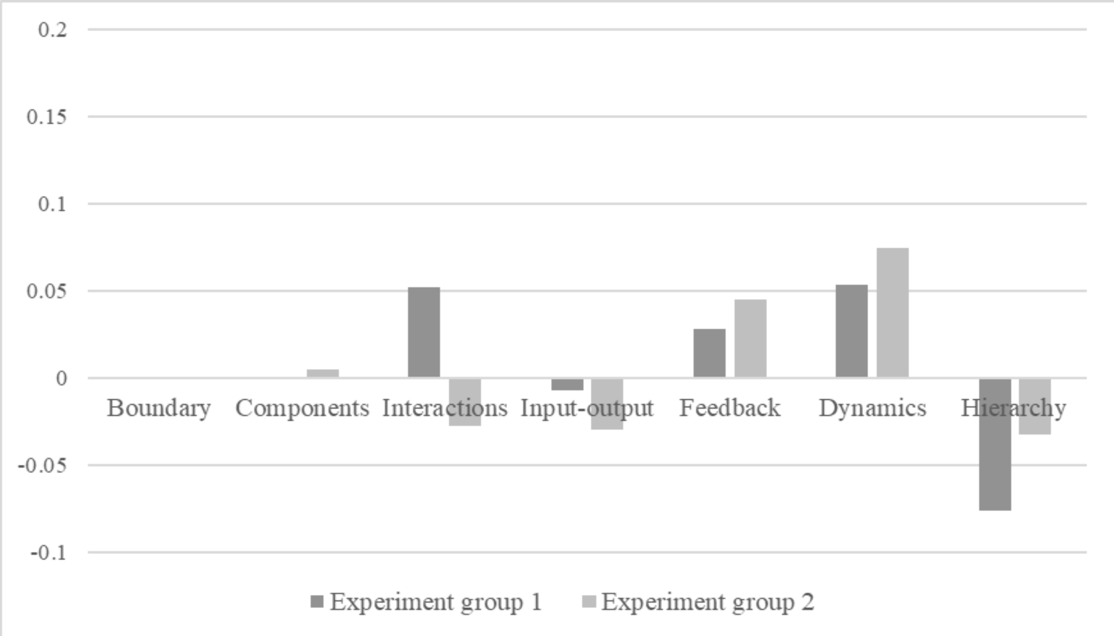
The results of the two experiment groups were compared to each other to see whether there was a difference between these two groups. These groups had both a different teacher, therefore a difference in results can be explained. Table 4 shows the results of the two groups per test. Graph 2 shows the increase or decline in scoring for the characteristics per class.

The results show that both classes scored high on the characteristic of components in the first test and that that less students scored in the second test, but for the last test they suddenly score high again for the characteristics of components. In the first test there was also no difference between the pre and the post tests, while in the other two tests the post-tests scored higher than the pre-tests. Overall, the students showed only a little increase in the use of the characteristics of components.

Experiment group 2 scored in the pre-test of respiration a lot higher than the post-test on the characteristics of interactions, but also a lot higher than experiment group 1. Overall, in experiment group 2 the interactions in the pre- and post-test showed a decrease in scoring. As if the students drew less interactions in the post test. Also, experiment group 1 showed a decline in scoring on interactions in the third test about digestion. Still, experiment group 1 showed an increase in scoring on the characteristics of interactions in total, while experiment group 2 showed a decline in scoring for the characteristics of interactions.

Input-output was scored lower in the first and the last tests for the post-test compared to the pre-test. While in the second test little more students were scored higher in the post-test in experiment group 1, but no differences in experiment group 2. Overall, a decline in scoring for input-output was present in both classes.

Graph 2. Growth factor per characteristics experiment group 1 compared to experiment group 2



A positive number in the graph shows an increase in scoring of that given characteristics. A negative number in the graph shows a decline in scoring of that given characteristics over all the tests.

The characteristic of feedback had only been scored in two tests. It showed an increase in number of students that got scored on the test. That means that more students used a feedback mechanism in their explanation in the third test about digestion.

For dynamics the students showed an overall increase in their scoring. Only experiment group 2 scored lower in the second test about photosynthesis. Therefore, it looked like experiment group 2 had a higher increase in scoring (more growth) than experiment group 1, but experiment group 1 scored higher on the second test than the third. Both classes used more dynamics over time and showed progress in using the characteristics dynamics in their explanation.

Last, the students of experiment group 1 scored higher on the characteristics of hierarchy in the first test than experiment group 2, however over time, the score of hierarchy declined when the students made the other tests. This suggests that less students thought or drew on different levels.

Table 4. Relative results of the two experiment groups, experiment group 1 and experiment group 2 per test per characteristics

	Respiration				Photosynthesis				Digestion			
Class	Experiment group 1 (n=18)		Experiment group 2 (n=17)		Experiment group 1 (n=23)		Experiment group 2 (n=17)		Experiment group 1 (n=20)		Experiment group 2 (n=20)	
Test	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Boundary	x	x	x	x	x	x	x	x	x	x	x	x
Components	1.00	1.00	0.94	0.94	0.57	0.74	0.47	0.71	0.95	1.00	0.90	0.95
Interactions	0.39	0.39	0.65	0.41	0.65	0.74	0.53	0.47	0.70	0.55	0.60	0.35
Input-output	0.44	0.28	0.53	0.41	0.74	0.78	0.71	0.71	0.50	0.25	0.45	0.30
Feedback	0.67	0.61	0.53	0.41	x	x	x	x	0.80	0.75	0.60	0.70
Dynamics	0.61	0.61	0.59	0.35	0.96	0.91	0.65	0.71	0.85	0.85	0.85	0.80
Hierarchy	0.94	1.00	0.71	0.53	0.57	0.57	0.35	0.35	0.60	0.65	0.50	0.50

The numbers in the table represent the relative numbers that are scored for the given characteristic for the experiment groups. This means that for the experiment group 1 in the pre-test of respiration, 1.00 or 100 percent of 18 students scored for the characteristic of components. A x indicates that the characteristics could not be scored

Explicit use of systemic language by students

Systems language was defined as the use of the word system, the specific characteristics, a synonym or a description of the characteristics in the right context. In this section the amount of systems language used by the students over time, therefore per test, is analysed. The number of used characteristics is shown in table 5. When there is no number in the table, no explicit use of that characteristics was scored.

Test 1: Respiration. In the first test almost only use of system language was found in the word 'system', or within the characteristic hierarchy. Only one student names boundary or interactions, and then in the form of a relation between the components:

"They are all systems within biology.") "They all form a system and they all have a system." "It belongs to an organisation level." "You see that everything that does not belong to the respiration system belongs is outside of it, so that is the boundary. They have all the same relationship with each other, namely breathing, they are parts of the whole."

Table 5. explicit use of characteristics in absolute numbers per test per class of the experiment groups.

	Respiration				Photosynthesis				Digestion			
Class	Experiment group 1 (n=18)		Experiment group 2 (n=17)		Experiment group 1 (n=23)		Experiment group 2 (n=17)		Experiment group 1 (n=20)		Experiment group 2 (n=20)	
Test	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Boundary		1										
Components									1	3	2	2
Interactions		1			2	2			1	3	2	
Input-output					4	7	4	4	3	3	6	5
Feedback								1		1	2	2
Dynamics							1		1	1		1
Hierarchy	2	1	1	3				1		2		
System	1	6		1		1		4	6	10	3	4

Each number in the table represents the absolute number of students that used the specific characteristic in the selected test. For example, in the pre-test of respiration in experiment group 1, two students explicitly used the word hierarchy and one student used the word system in their explanation.

Experiment group 2 only showed explicit use of systems language in the characteristics of hierarchy and one student that used the explicit word system in his or her explanation. In the post test, an increase in the use of the word systems was noticed, especially in experiment group 1. Students in experiment group 2 used more hierarchy in their post-test.

Test 2: photosynthesis. The second test showed a shift to other characteristics that are explicitly used by the students. More explicit use of system language was shown on the characteristics of interactions and a high increase in the use of input-output. Also, dynamics was used by one student, in the pre-test, and hierarchy is only mentioned twice in the post-test:

> *"There is an input and an output." "Both take something up from the environment." "With both, a substance goes in and out and it is a system within an organism." "process, function, feedback loop, level (organism)." "They are both systems that take something from outside the body and converts it to something an organism need." "Both processes absorb and process a certain substance so that the organism can use the substances."*

This suggests that the students found more comparisons in input-output in the second test than in the first, whether that was due to more knowledge or due to a different question in item four. In item four of the second test, students had to explain the similarities between the respiration system (first test) and photosynthesis (second test). A lot of students only used the direct word and did not make a connection between the two systems (*"a dynamic" "input-output"*). Still, it was interesting to see that they realised that those characteristics were found in multiple systems.

Test 3: digestion. The students still used a lot of input and output in their system language in the third test. There were also more characteristics used, components, interactions, input-output were used the most. A couple of students used feedback and dynamics and the connection with the word systems was more used than in the previous tests:

> *"All three are systems with all little building blocks to become that way. So, they work all together in their system." "Components that work together, certain function, input/output, process." "multiple organs are used. They have to work together to get to a good result." "Input and output. They are all systems with systems characteristics." "It has an input and an output with dynamics." "They are all three systems with a feedback." "it all has a boundary, components, input and output." "All input-output. Lots of systems characteristics are the same." "Working together of multiple organs. There are forms of feedback in all the systems. In all the systems different interactions take place."*

It was noticeable that not only the fourth item stimulated the system language in the students, however they used some system language in their explanation of their drawings. So more explicit use of the characteristics by the students in the third test. Instead of only using the characteristic to use it, they explained what they mean, and they made connections between the characteristics and the different systems.

Discussion

The results visualised in graph 1 suggest that while the experiment group followed lessons about ST, they did not show an increase in all the characteristics of systems. While the control group, whom we expected to show no increase or decline in scoring, scored higher on the increase over the tests. This suggests that the experiment group did not learn to use the characteristics of systems in their explanation or drawing. The control group showed an increase in the scoring for the characteristics this could mean that the tests over time were somehow easier to interpret or that the control group had more knowledge about the subjects. When a student has more knowledge of a subject, he or she is more likely to draw more characteristics in their drawings. In the second test, the students of the control group just finished the chapter about photosynthesis. Therefore, they had more knowledge about the topic. This explains the higher percentage scored in the second test in the control group (see table 2), compared to the pre-test of the experiment group. It could also depend how they experience the subject of the test. Not all students have the same level of understanding and therefore one might find the subject of respiration harder than digestion or maybe photosynthesis is something he or she excels in. This might explain the lower score on hierarchy in the third test compared to the first test.

Another reason why there was a difference in growth between the control group and the experiment group, could be the relative high percentage of the experiment group that scored in the first test on components and hierarchy. In these two characteristics approximately 97 percent (components) and 80 percent (hierarchy) students scored for those characteristics, and in the third test, they scored just as high for components, however lower for hierarchy. Therefore, real growth is not visible, although the knowledge was already present.

The size of the participating groups were not equal and also differ per test. Therefore, the results are not entirely reliable. To solve this problem, we opted to use relative results.

Some students claimed to be test tired. With six tests over the year on top of the regular tests they got, students wondered why they were doing this. They did not receive a grade for it and some of the experiment group students did not see the purpose of these tests. The control group was left blank as well, so they were not influenced on the subject of the research. They only knew that they had to make the test in order to help the researcher out on the research. Only after they finished all the tests, a little presentation was held to tell them what they were doing. By doing this the researcher hoped that the students were a “true” control group, however, it also influenced their motivation. Because they did not know the purpose behind the test, they were less motivated on making the test.

We found that not all characteristics could be scored in the drawing task. As mentioned in the previous chapter, the characteristic of boundary could not be scored in the drawings of the students. It is discussed whether it is more important to think (and draw) on multiple levels than on just one level. Still, it is important for students to learn that a certain system also has a boundary and that certain processes take place within that boundary. To implement the characteristics of boundary, this kind of test is not suitable. The drawing task is to open and therefore students do not know if they can cross the boundary or not.

The characteristics of feedback that could not be scored in the second test of photosynthesis also has an influence on the results. Therefore, this item should be changed.

The drawing test could measure the characteristics of systems, and therefore be used to define the extent of students' ST, although it needs some improvements to make sure all the characteristics can be measured. By measuring all the characteristics students' ST is becoming more visual. However, does a drawing reflect the true thinking process of a student? A drawing only reflects a students' visual representation of a certain system or a process. When a student is more visually oriented, he or she can express him or herself more easily by drawings. Nevertheless, not all students are visually oriented and find it hard to draw their ideas, therefore, their drawings might be incomplete.

Improvements

To reflect the true thinking process of a student, a group of students can be asked to make the test and thereby thinking out loud (Cox, et al., 2018). A drawing test on its own could limit the students who do not like to draw, or who do not know how to draw certain systems and processes. Even with an extra space to write down an explanation, students find it sometimes a challenge to write down exactly what they mean. With thinking out loud, a student tells the researcher what he or she is thinking while taking the test. Hence, more characteristics of systems could be found in the thinking process, instead of solely the drawing.

To prevent the students getting test tired, the tests could be integrated within the regular tests or used as formative tests. Thereby, the students use ST in the regular subjects and see the purpose of ST other than participating in a research project.

In order to be able to score all system characteristics, while maintaining the same structure of the tests, systems language can be explicitly implemented in the assignment. For example, ask an extra question to let the students explain the boundary or boundaries within the drawn system(s). This allows them to still think hierarchical, although check whether the students actually understand how systems can be defined within their boundaries.

The same can also be done with the third assignment of the photosynthesis test. Where there is actually no feedback present in the system at that time, but let the students explain this. Ask the question whether they can see which characteristic of systems cannot be properly depicted or explained and let them explain why they have chosen this characteristic.

Another option to test systems characteristics, is to give the students an already existing system. Using a photo, image, model or a story, the students then explain the existing system characteristics, what they consist of and how they interact. With this, the knowledge is already sufficiently present in the image, model or the story and therefore students only have to fill in the systems characteristics or explain how they interact. This kind of assignment operate on lower levels of Bloom's taxonomy (Krathwohl, 2002) because the students only have to understand and apply the system characteristics. With the current assessment setup, the students have to draw the system, create it, and therefore operate on the highest level of cognitive thinking (Krathwohl, 2002). However, this assignment with a given system can perhaps be used as a starting assignment to introduce ST, or just to test the students on a lower level of cognitive thinking.

The assignment can also be applied to a system that is independent of a biological subject. The students receive a topic and they have to design a model that meets the system characteristics that are not linked biologically. For an example, a supermarket. Multiple components like the shelfs and the products, but also the staff and the buyers have interactions with each other. Products come in and out and the sale of the products influence the dynamics of the products. All this interacts with the factories where the products are packed, and they again interact with the farmers that cultivate the products (multiple organisational levels). With this kind of assignment, regular ST skills are practiced instead of only biological ST. Therefore, this kind of assignment can also be used cross-subject but is therefore not specifically tested in biology. An advantage of this can be that students come to see that systems are all around us and that it is therefore interesting to see how those characteristics keep coming back. If students can see this, they may be able to look at (biological) phenomena from a systems' perspective and they may apply the characteristics themselves in learning to understand and analyse systems and create a more coherent understanding of biology.

Conclusion

The aim of this research was (1) to determine to what extent the designed assessment method visualizes the extent of students' ST in biology education, and (2) to determine to what extent there was progress in students' ST after being taught about ST.

The results suggest that the test could measure the extent of ST apart from one characteristic (boundary) and question three in the second test could not measure any feedback. A few improvements for the assessment method are given. For instance, implementing the systems language in the questions, to use a system that is not bound to biology, a question where students fill in and explain the characteristics instead of creating the system themselves. Also, the timing of the test and the subject taught is important, and to get the best results, these two need to be combined.

The results of the tests furthermore suggest that there is an increase in the use of systems language in the students of the experiment group, while the control group did not use (more) systems language explicitly. The two experiment groups had different results, which suggest that not only the lessons about ST had an influence on students' ST, but perhaps also the setting in the group and or the teacher. Due to the different results of the experiment groups the overall results of the experiment group as a whole is not reliable to compare to the control group.

The results of this research give several leads for improvement of the test and the setting of the test in which students' ST can be measured.

Reference list

Boersma, K.T., Kamp, M. J. A., Van den Oever, L. & Schaik, H. H. (2010). *Naar actueel relevant en samenhangend biologieonderwijs. Eindrapportage van de commissie Vernieuwing Biologie Onderwijs, met nieuwe examenprogramma's voor havo en vwo.* [Towards actual, relevant and coherent biology education. Final report of the board for the innovation of biology education, with new examinations programs for general upper secondary and pre-university education]. Utrecht: CVBO.

Boersma, K., Waarlo, A. J., & Klaassen, K. (2011). The feasibility of systems thinking in biology education. *Journal of Biological Education*, *45*(4), 190–197.

Brandstädter, K., Harms, U., & Grossschedl, J. (2012). Assessing system thinking through different concept-mapping practices. *International Journal of Science Education*, *34*(14), 2147-2170.

Cicchetti, D. V. (1976). Assessing inter-rater reliability for rating scales: resolving some basic issues. British Journal of Psychiatry, 129, 452e456.

Cox, M., Elen, J., & Steegen, A. (2018). A test to measure students' systems thinking abilities in geography. *European Journal of Geography*, *9*(1), 105-120

Danish, J., Saleh, A., Andrade, A., & Bryan, B. (2017). Observing complex systems thinking in the zone of proximal development. *Instructional Science*. doi:10.1007/s11251-016-9391-z.

Dauer, J. T., Momsen, J. L., Speth, E. B., Makohon-Moore, S. C., & Long, T. M. (2013). Analyzing change in students' gene-to-evolution models in college-level introductory biology. *Journal of Research in Science Teaching*, *50*(6), 639-659.

Dikmenli, M. (2010). Misconceptions of cell division held by student teachers in biology: A drawing analysis. *Scientific Research and Essays*, *5*(2), 235-247.

Evagorou, M., Korfiatis, K., Nicolaou, C., & Constantinou, C. (2009). An investigation of the potential of interactive simulations for developing system thinking skills in elementary school: A case study with fifth-graders and sixth-graders. *International Journal of Science Education*, *31*(5), 655-674.

Gilissen, M. G.R, Knippels, M. C. P.J., Verhoeff, R. P., & van Joolingen, W. R. (2019). Teachers' and educators' perspectives on systems thinking and its implementation in Dutch biology education. *Journal of Biological Education*, 1-12.

Gilissen, M. G. R, Knippels, M. C. P. J., & van Joolingen, W. R. (submitted). Heuristics for introducing systems thinking into the biology classroom.

Hmelo-Silver, C. E., Jordan, R., Eberbach, C., & Sinha, S. (2017). Systems learning with a conceptual representation: A quasi-experimental study. *Instructional Science*. doi:10.1007/s11251-016-9392-y.

Hoogland, K., Pepin, B., Bakker, A., de Koning, J., & Gravemeijer, K. (2016). Representing contextual mathematical problems in descriptive or depictive form: Design of an instrument and validation of its uses. *Studies in Educational Evaluation*, *50*, 22-32.

Kindfield, A. C. (1994). Biology diagrams: Tools to think with. *The Journal of the Learning Sciences*, *3*(1), 1-36.

Knippels, M. C., & Waarlo, A. (2018). Development, Uptake, and Wider Applicability of the Yo-yo Strategy in Biology Education Research: A Reappraisal. *Education Sciences*, *8*(3), 129

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice*, *41*(4), 212-218.

Landis, J. R., & Koch, G. G. (1977). The measu rement of observer agreement for cate gorical data. *Biometrics*, *33*, 159174. doi:10.2307/2529310

Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2010). *Methods in educational research: From theory to practice* (Vol. 28). John Wiley & Sons.

Maienschein, J. (1991). From presentation to representation in EB Wilson'sThe Cell. *Biology and Philosophy*, *6*(2), 227-254.

National Research Council. (2011). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. *Social Sciences* (Vol. 1).

NGSS Lead States. (2013). Next Generation Science Standards: For States, by States. *Achieve, Inc. on Behalf of the Twenty-Six States and Partners That Collaborated on the NGSS*, (November), 1–103.

Novak, J. D. (1990). Concept maps and Vee diagrams: Two metacognitive tools to facilitate meaningful learning. *Instructional Science*, *19*(1), 29-52.

Riess, W., & Mischo, C. (2010). Promoting systems thinking through biology lessons. *International Journal of Science Education*, *32*(6), 705-725.

Tripto, J., Assaraf, O. B. Z., & Amit, M. (2013). Mapping what they know: Concept maps as an effective tool for assessing students' systems thinking. *American Journal of Operations Research*, *3*(01), 245.

Tripto, J., Assaraf, O. B. Z., Snapir, Z., & Amit, M. (2016). How is the body's systemic nature manifested amongst high school biology students? *Instructional Science*. doi:10.1007/s11251-016-9390-0.

Tripto, J., Assaraf, O. B. Z., & Amit, M. (2018) Recurring patterns in the development of high school biology students' system thinking over time. *Instructional Science*, 1-42.

Van der Ark, A., & ten Hove, D. (2018). Do we agree? Interrater reliability in education. *PEDAGOGISCHE STUDIEN*, *95*(5-6), 361-371.

Verhoeff, R. P. (2003). *Towards systems thinking in cell biology education* (Doctoral dissertation).

Verhoeff, R., Jan Boerwinkel, D., & Jan Waarlo, A. (2009). Genomics in school. Science and Society Series on Convergence Research. *EMBO Reports*, *10*(2), 120–124.

Verhoeff, R. P., Knippels, M. C., Gilissen, M. & Boersma, K. T. (2018). The theoretical nature of systems thinking. Perspectives on systems thinking in biology education. In *Frontiers in Education* (Vol. 3, p. 40). Frontiers.

Yoon, S. A. & Hmelo-Silver, C. (2017) Introduction to special issue: models and tools for systems learning and instruction. *Instructional Science*, *45*(1), 1-4.

APENDIX A: Test 1: Respiration (in the lungs) (Dutch)

Opdracht biologie

Leeftijd: …………………………………………………

Geslacht: ………………………………………………

School: ……………………………………………………

Klas: ………………………………………………………

Docent: …………………………………………………….

Methode biologie: ……………………………………..

Datum: ……………………………………………………..

Tekening ademhaling

1. Maak een zo uitgebreid mogelijke tekening van alles wat jij weet over ademhaling én benoem alles wat je hebt getekend. Je wordt niet beoordeeld op je tekenvaardigheden. De tekening moet begrijpbaar zijn voor een mede-leerling, dus houd het leesbaar.

2. Leg in het vak hieronder uit wat je hebt getekend.

3. Wanneer je van rust overgaat tot een sprint zal je ademhaling versnellen. Licht aan de hand van je tekening toe wat er precies verandert aan de ademhaling als je gaat sprinten.

4. Wat zijn overeenkomsten tussen de volgende biologische objecten?

- een plantencel
- een spier
- het verteringsstelsel
- een kat
- een bos

Einde van de opdracht. Controleer je antwoorden nogmaals en lever vervolgens de opdracht in bij je docent.

APENDIX B: Test 2: Photosynthesis (Dutch)

Naam: ………………………….. Docent: Datum:……………..

Tekening fotosynthese

1. Maak een zo uitgebreid mogelijke tekening van alles wat jij weet over de fotosynthese én benoem alles wat je hebt getekend. Je wordt niet beoordeeld op je tekenvaardigheden. De tekening moet begrijpbaar zijn voor een mede-leerling, dus houd het leesbaar.

2. Leg in het vak hieronder uit wat je hebt getekend.

Naam: ………………………….. Docent: Datum:……………..

3. Precies tussen twee jonge bomen in wordt een heel groot huis gebouwd waardoor één boom altijd in de schaduw van het huis komt te staan. Na een paar jaar is er een verschil te zien tussen de twee bomen. Geef aan wat dit verschil is en leg aan de hand van je tekening (die je bij opdr. 1 hebt gemaakt) uit hoe dit verschil is ontstaan.

4. Je hebt nu een tekening van de fotosynthese gemaakt en de vorige keer van de ademhaling. Wat zijn overeenkomsten tussen deze twee processen en hoe kun je deze overeenkomsten terugvinden in je tekening?

Einde van de opdracht. Controleer je antwoorden nogmaals en lever vervolgens de opdracht in bij je docent.

APENDIX C: Test 3: Digestion system (Dutch)

Naam: ………………………….. Docent: Datum:……………..

Tekening vertering

1. Maak een zo uitgebreid mogelijke tekening van alles wat jij weet over de vertering én benoem alles wat je hebt getekend. Je wordt niet beoordeeld op je tekenvaardigheden. De tekening moet begrijpbaar zijn voor een mede-leerling, dus houd het leesbaar.

2. Leg in het vak hieronder uit wat je hebt getekend.

Naam: ………………………….. Docent: Datum:……………..

3. Gedurende de dag heb je soms het gevoel dat je honger hebt en soms juist een gevoel van verzadiging.

a. Waardoor ontstaat deze schommeling?

b. Leg uit hoe het honger- en verzadigingsgevoel ontstaat.

4. Je hebt nu een tekening van de vertering gemaakt en de vorige keren van de ademhaling en fotosynthese. Wat zijn overeenkomsten tussen deze drie processen?

Einde van de opdracht. Controleer je antwoorden nogmaals en lever vervolgens de opdracht in bij je docent.

APENDIX D: Code book test 1: Respiration (of the lungs) (Dutch)

Codeboek ademhaling

Algemeen:

- Bij het scoren van de kenmerken wordt niet zozeer gekeken naar de volledigheid of de correctheid. Bijvoorbeeld:
  - Zuurstof en koolstofdioxide omdraaien
  - Longblaasjes vergeten
- De kenmerken kunnen gescoord worden in de tekst of in de tekening.
- De kenmerken worden gescoord op aanwezigheid (1) of afwezigheid (0). Kenmerken kunnen meerdere keren voorkomen
- Wanneer een kenmerk expliciet is benoemd wordt het vakje geel gemarkeerd.
- Per kenmerk scoren

Opdracht 1+2+4	? scoor: componenten, interacties, input & output en genestheid.
Opdracht 3+4	? scoor: feedback en dynamiek.
Opdracht 1 t/m4	? expliciet benoemen van systeemkenmerken.

Componenten:

- De tekening of tekst bevat verschillende componenten van het ademhalingssysteem.
  - Als minimale is vereist: longen en luchtpijp.
  - Meer componenten is goed, maar geen hogere score.
- De componenten hoeven niet expliciet benoemd te worden om gescoord (1) te worden.
- Om geel gemarkeerd te worden moet de term ‘componenten’ of ‘onderdelen’ expliciet worden genoemd.

Interacties:

De relatie tussen de verschillende componenten en de input en output moet aangegeven zijn. Dit kan aangegeven zijn door middel van pijlen of in de tekst met woorden.

- Minimale interactie dat wordt gescoord is het transport van lucht (koolstofdioxide of zuurstof) van buiten het systeem door de luchtpijp naar de longen.
- Het noemen/tekenen van longen en luchtpijp in interactie met lucht (koolstofdioxide of zuurstof) is van belang. Zonder het noemen van die elementen geen punt
- Het gaat hier om interactie tussen componenten, dus geen fragmentarische weergave. De pijl moet dan door de luchtpijp naar de longen (en terug) getekend zijn.
- Verschil met input en output is dat hier de interacties met verschillende componenten WEL aanwezig moet zijn.

Input output:

- Er is aangegeven dat zuurstof en koolstofdioxide in en uit de longen gaat. Dit kan in de tekening aangegeven zijn of in woorden.
- Goed zou bijvoorbeeld zijn: er wordt zuurstof ingeademd en koolstofdioxide uitgeademd (hier mogen koolstofdioxide en zuurstof omgedraaid zijn). Hier ontbreekt dus de interactie met de componenten.
- Verschil met interacties is dat de interacties met de verschillende componenten NIET aanwezig hoeft te zijn.

Feedback:
Een snellere ademhaling heeft tot gevolg dat de input output groter wordt, ofwel toename uitwisseling lucht (koolstofdioxide zuurstof), maar ook toename volume longen.

- Minimaal benoemd dat er een toename in de uitwisseling van lucht is of toename volume longen bij het overgaan tot sprint.
- Alleen benoemen van sneller ademhalen is geen goed antwoord, aangezien dit in de vraag terugkomt.
- Leerlingen hebben nog geen informatie gehad over het regelcentrum en hoe de feedbackloop daar loopt, dus dit kan niet genoemd worden.
- Toename in vraag naar zuurstof

Dynamiek:

- Er moet een concreet voorbeeld gegeven zijn van de rust situatie en de sprint situatie
  - Voor de sprint …, tijdens de sprint … OF
  - Toename input output OF volume longen veranderd.
  - Geen punt: sneller ademen/versnellen ademhaling. Is al genoemd in de opdracht.
- Verschil met feedback is het benoemen van begin en eindsituatie (zie eerste inspringende punt). Daarbij geven de woorden *meer* en *minder* dit verschil ook al aan.

Hiërarchie/genestheid:

- Het getekende systeem bestaat uit een groter geheel met ademhalingsstelsel (omhulsel van een persoon met ademhalingsstelsel), OF
- Er wordt ingezoomd binnen het systeem (ademhalingsstelsel met longblaasjes), OF
- Er zijn meerdere systemen met elkaar verbonden (ademhalingsstelsel en bloedvatenstelsel), OF
- Er wordt genoemd dat er verschillende organisatieniveaus zijn.

Opdracht 4:

- Er moet een relatie gelegd worden met *alle* objecten. Hierbij wordt er wel gekeken naar de correctheid van de uitspraak:
  - Wanneer er wordt gezegd dat het allemaal organismen zijn is dit niet correct.
  - Correct: het zijn/bestaat allemaal (uit) cellen.
- Wanneer wordt toegelicht dat er verschillende organisatieniveaus aanwezig zijn is dit ook correct. Bij het noemen van ‘organisatieniveaus’ kan ‘hiërarchie’ geel worden gemarkeerd.

APENDIX E: Code book test 2: Photosynthesis (Dutch)

Codeboek fotosynthese

Algemeen:

- Bij het scoren van de kenmerken wordt niet zozeer gekeken naar de volledigheid of de correctheid. Bijvoorbeeld:
  - Zuurstof en koolstofdioxide omdraaien
  - Water komt via het blad binnen
  - Zonlicht vergeten
- De kenmerken kunnen gescoord worden in tekst of in de tekening.
- De kenmerken worden gescoord op aanwezigheid (1) of afwezigheid (0). Kenmerken kunnen meerdere keren voorkomen
- Wanneer een kenmerk expliciet is benoemd wordt het vakje geel gemarkeerd.
- Het coderen vindt per kenmerk plaats. Dus eerst wordt het kenmerk 'componenten' gescoord voor alle leerlingen, vervolgens 'interacties', enzovoorts.

Opdracht 1 + 2 + 4 → scoren componenten, interacties, input en output en genestheid.
Opdracht 3 + 4 → scoren feedback en dynamiek.
Opdracht 1 t/m 4 → expliciet benoemen systeemkenmerken

Componenten:
Het systeem kan op verschillende organisatieniveaus zijn weergegeven. Afhankelijk van het gekozen systeem moeten bepaalde componenten zijn genoemd. Alleen het hoofdsysteem wordt gescoord. Soms tekenen leerlingen bijvoorbeeld nog gras om de plant heen, maar dan wordt alleen de tekening van de plant gescoord.
Input en output zijn hier niet als component gescoord.

- Plant/boom/bloem: minstens stengel en blad getekend.
  - Het tekenen van alleen een stengel en een bloem is niet voldoende. Is er een blad bij getekend, dan is dit wel voldoende.
  - Bij het tekenen van een boom met bladerdek is dit niet voldoende. Wanneer bij een boom een apart blad getekend wordt is dit wel goed.
- Blad: minstens bladsteel en nerven.
- Cel: minstens bladgroenkorrel en celwand (met huidmondjes) getekend.
- Bladgroenkorrel: lln hebben nog geen kennis over de componenten van een bladgroenkorrel, dus wanneer een lln een bladgroenkorrel heeft getekend geen punt toekennen.

Interacties:
De relatie tussen de verschillende componenten en de input output moet aangegeven zijn. Koolstofdioxide, water, zonlicht en mineralen zorgen voor de productie van glucose (voedingsstoffen) en/of zuurstof. Uit de tekening en/of in de tekst moet deze interactie naar voren komen. Als één tot drie elementen ontbreken of zijn omgedraaid wordt dit alsnog goed gerekend.

Energie wordt goedgerekend. Het gaat hier om de relatie tussen de input en de output met de componenten. Dus dat er iets ingaat en er wat veranderd

Voorbeelden:

- Alleen het opschrijven van de formule is niet voldoende, omdat het niet duidelijk is hoe dit zich verhoudt tot het systeem.
- Het noemen van omzetten of omzetting van beginstof tot eindstof(fen) wordt gescoord als een vorm van interactie, bv: water en koolstofdioxide wordt omgezet tot glucose (of energie).
- Met pijlen kunnen in de tekening de interacties zijn aangegeven tussen de beginstoffen en de eindstof(fen).
- Niet goed: "Hier wordt licht omgezet" → glucose productie mist.
- Wel goed: het groeien van een bloem in het zonlicht (component met element)

<u>Input output:</u>
Er moet minstens één input (CO2, H2O, zonlicht) en één output (glucose/energie, O2, H2O) genoemd of getekend zijn. Wanneer er maar 1 wordt gegeven is dit niet voldoende. Alleen het noemen van de fotosynthese reactie is niet voldoende, aangezien het dan niet in relatie staat tot elkaar en het gehele systeem. Uit de tekening moet echt blijken dat een stof het systeem in gaat en uitgaat door middel van pijlen.

- Wanneer leerling zegt water en zonlicht omgezet in glucose, maar niet duidelijk dat het IN de plant(onderdeel)/systeem omgezet wordt, dan is dit niet correct.
- Groei van een bloem/plant is ook een vorm van uitput
- Expliciet benoemen is ook het geval wanneer lln het volgende beschrijft: "*Er gaat iets in/uit.*"

<u>Feedback</u>: NIET SCOREN
Er is geen feedbackloop te onderscheiden hier. De zon beïnvloedt de mate van fotosynthese en daarmee de groei, maar de groei/fotosynthese beïnvloedt niet de zon. Met de aanwezige componenten kan geen feedbackloop worden gemaakt. Het gaat hier wel om een oorzaak-gevolg relatie/interactie, maar niet om feedback. Als lln de relatie leggen tussen afname van zonlicht → minder fotosynthese → minder groei oid kunnen we dit wel scoren als interactie (als ze deze fout hadden gedaan in opdracht 1).

<u>Dynamiek</u>:
De relatie moet gelegd worden tussen afname van zonlicht (door schaduw) en verlaagde fotosynthese of vermindering gezondheid of verkleuring of groei. Fotosynthese hoeft niet perse genoemd te worden.
Wanneer een lln een punt heeft voor feedback, heeft hij/zij automatisch ook een punt voor dynamiek.

<u>Hierarchie</u>:

Minimaal twee verschillende organisatieniveaus benoemd en/of getekend waarbij duidelijk is dat de lln ziet dat het ene onderdeel een deelsysteem is van het andere grotere systeem.
Voorbeeld:

- Goed: Lln die een plant/boom tekent met een pijltje vanuit blad met daarnaast een plantencel, aangezien dit erop wijst dat hij/zij inzoomt op een deelsysteem
- Goed: Lln die een blaadje los tekent en daarnaast een tekening van een boom met bladerdek. Dit illustreert namelijk 2 systemen.
- Niet goed: lln die een plant/boom heeft getekend met een aantal blaadjes.

Soms noemen de lln in de tekst nog een ander organisatieniveau “dit gebeurt in de bladgroenkorrels” lastig om dat te scoren. Ik heb ze als ze in de tekst staan wel gescoord.

Emergentie:
Mogelijk bij vraag 4 expliciet genoemd of beschreven. Iets in de trant van: het geheel is meer dan de som der delen.

Expliciet een systeemkenmerk benoemd bij vraag 4 of elders, met geel arceren en scoren ook als de lln dit niet goed had voor de desbetreffende context.

APENDIX F: Code book test 3: Digestion system (Dutch)

Codeboek vertering

Algemeen:

- Bij het scoren van de kenmerken wordt niet zozeer gekeken naar de volledigheid of de correctheid. Bijvoorbeeld:
  - Voedingsstof en voedingsmiddel omdraaien
  - Naam orgaan vergeten
- De kenmerken worden gescoord in tekst of in de tekening.
- De kenmerken worden gescoord op aanwezigheid (1) of afwezigheid (0). Kenmerken kunnen meerdere keren voorkomen, dit wordt nog steeds als 1 gescoord.
- Wanneer een kenmerk expliciet is benoemd wordt het vakje geel gemarkeerd.
- Per kenmerk scoren, niet per leerling

Opdracht 1+2+4 scoor: componenten, interacties, input & output en genestheid.
Opdracht 3+4 scoor: feedback en dynamiek.
Opdracht 1 t/m4 expliciet benoemen van systeemkenmerken.

Componenten:

- De tekening bevat verschillende componenten van het verteringsstelsel
  - Minimale benoeming: minstens 2 componenten van het verteringsstelsel (denk aan slokdarm, maag, darmen etc.)
- De componenten hoeven niet expliciet genoemd te worden om gescoord (1) te worden, maar mogen ook in de tekening aanwezig zijn. Mits het duidelijk is dat het om het verteringsstelsel gaat (aan de hand van vormen).

Interacties:

De relaties tussen de verschillende componenten en de input en output moeten aangegeven zijn. Dit kan aangegeven zijn door middel van pijlen, nummers of in de tekst met woorden.

- Dit kan gaan over de volgorde van vertering. Langs welke organen de voedselbrij gaat. Denk hierbij dan aan nummers of pijlen met de richting. Correctheid van de volgorde mag afwijken. het woord *via* maakt dit duidelijk.
- Er moet een verbinding met het *verteren van eten/omzetten van voedsel tot voedingsstoffen* en de organen zijn. Dat er in de getekende/benoemde organen het eten verteerd wordt
- "Van belang zijn bij vertering" is niet specifiek genoeg en wordt niet gescoord.
- Of er moet een relatie gelegd zijn tussen voeding, de onverteerde delen (ontlasting) en de opgenomen stoffen
- Verschil met input output is dat het duidelijk moet zijn dat er interacties plaatsvinden tussen de verschillende componenten. dus de voedingsstoffen die opgenomen zijn in een darm, of de verteringssappen uitgescheiden door een orgaan.

Input output:

- Er moet aangegeven zijn dat er eten/voedsel via de mond naar binnen gaat, wat er aan verteringssappen aan te pas komt of wat er opgenomen wordt aan verteerde voedingsstoffen. Dit kan eventueel met pijlen, nummers of in tekst aangegeven worden.
- Het beginnen van vertering alleen is niet voldoende (geen input).

- De pijlen moeten duidelijk de mond IN en evt darmen/anus UIT
- Er moet input genoemd worden (denk aan verteringssappen, voedsel) EN output (verteerd voedsel dat kan worden opgenomen, ontlasting)
- Voedsel in je mond en evt anus eruit is al correct
- Het is voor de leerlingen niet allemaal duidelijk dat het verteringsstelsel nog BUITEN het lichaam is, dus dit niet fout rekenen.
- Verschil met interacties is dat het duidelijk moet zijn dat er iets in gaat en ook weer uit, de interacties tussen de componenten is niet van belang hier.

Voorbeeld:

- Wel: dan gaan de overgebleven stoffen via ontlasting naar buiten
- Niet goed: dan wordt het ontlasting

Feedback: (deelvraag b)

Leerling moet relatie leggen tussen minder eten gevoel van honger meer eten meer *vertering* gevoel van verzadiging minder eten ... etc.

- Hierbij is vertering het belangrijkste element omdat de vertering ervoor zorgt dat er meer suikers in het lichaam opgenomen worden en er minder gevoel van honger is.
- Wanneer er gesproken wordt over een signaal/stof om te gaan eten wanneer je te weinig (voedings)stoffen/glucose/energie in je lichaam hebt is dit goed. wel in combinatie met weinig (voedings)stoffen, alleen een signaal om te gaan eten is niet voldoende.
- Dat de schommeling kan ontstaan door hormonen wordt goed gerekend.
- a en b moeten samen genomen worden, wordt in de ene vraag wel gekoppeld aan tekort aan voedsel/glucose/energie en hoe dit op te lossen en bij b gekoppeld aan een seintje vanuit de hersenen, dan is dit goed.
- Koppeling van meer of minder inspanning als voorbeeld van voeding is goed. (Meer inspanning staat gelijk aan minder voeding, minder inspanning aan meer voeding).

Dynamiek: (deelvraag a)

De relatie moet gelegd worden tussen het eten en het gevoel van verzadiging en het niet eten en het gevoel van honger.

- Het specifiek noemen van vertering is hier niet van belang. Het gaat om de dynamiek.
- Het benoemen van een schommeling van de input en output is van belang (wanneer er een input is (eten) zal er een gevoel van verzadiging ontstaan, wanneer geen input is, zal er een gevoel van honger ontstaan).
- Het benoemen van het opnemen van voedingsstoffen en verbruiken van voedingsstoffen geeft ook dynamiek aan.
- Het noemen van te weinig eten of van eten wordt ook goed gerekend.
- Koppeling van meer of minder inspanning als voorbeeld van voeding is goed. (Meer inspanning staat gelijk aan minder voeding, minder inspanning aan meer voeding).
- De schommeling kan ontstaan door hormonen, wordt goed gerekend.

Hiërarchie/genestheid:

- Het getekende systeem bestaat uit een groter geheel met verteringsstelsel (omhulsel van een persoon met verteringsstelsel), OF
- Er wordt ingezoomd binnen het systeem (verteringsstelsel met inzoomen op darmepitheel, of orgaan), OF

- Er zijn meerdere systemen met elkaar verbonden (verteringsstelsel met bloedvatenstelsel of spierstelsel).
- Alleen verteringsstelsel is geen genestheid (dus zonder omhulsel of in-uit zoomen. (Alleen orgaansysteemniveau?)

Emergentie:
Mogelijk bij vraag 4 expliciet genoemd of beschreven. Iets in de trant van: het geheel is meer dan de som der delen.