International Electronic Journal of Elementary Education
March 2022, Volume 14, Issue 4, 539-555

The Effect of Teaching with Anatomical Models in Science Education on Primary School Children's Understanding of Human Organs

Karsten Damerau[a,*], Melanie Marita Beudels[b], Christine Börtitz[c], Nadja El Balti[d], Laura Funk[e], Debora Westerholt[f], Julia Cohnen[g], Angelika Preisfeld[h]

Received : 24 January 2022
Revised : 29 March 2022
Accepted : 30 March 2022
DOI : 10.26822/iejee.2022.262

[a,*] **Corresponding Author:** Karsten Damerau, Europa University of Flensburg, Germany.
E-mail: karsten.damerau@uni-flensburg.de
ORCID: https://orcid.org/0000-0001-6061-8679

[b] Melanie Marita Beudels, University of Wuppertal, Germany.
E-mail: melanie.beudels@uni-wuppertal.de
ORCID: https://orcid.org/0000-0001-9018-1623

[c] Christine Börtitz, Europa University of Flensburg, Germany.
E-mail: christine.boertitz@uni-flensburg.de
ORCID: https://orcid.org/0000-0001-7383-5648

[d] Nadja El Balti, Europa University of Flensburg, Germany.
E-mail: nadja.el-balti@uni-flensburg.de
ORCID: https://orcid.org/0000-0003-0323-7700

[e] Laura Funk, University of Wuppertal, Germany.
E-mail: laura.funk98@gmx.de
ORCID: https://orcid.org/0000-0003-3758-7901

[f] Debora Westerholt, University of Wuppertal, Germany.
E-mail: westerholt@uni-wuppertal.de
ORCID: https://orcid.org/0000-0002-7598-9955

[g] Julia Cohnen, University of Wuppertal, Germany.
E-mail: juliacoh95@gmail.com
ORCID: https://orcid.org/0000-0003-3422-6545

[h] Angelika Preisfeld, University of Wuppertal, Germany.
E-mail: apreis@uni-wuppertal.de
ORCID: https://orcid.org/0000-0001-8579-0731

KURA group
Copyright ©
www.iejee.com
ISSN: 1307-9298

Abstract

Primary school students often do not have differentiated conceptions of human organs and organ systems. As understanding the inner structure of the human body is an elementary prerequisite for the development of health awareness, appropriate forms of teaching must be developed to modify students' fragmented preconceptions about the inside of their bodies into scientifically accurate concepts. Anatomical models are considered a medium to raise awareness of organs and their systemic integration; however, only a few studies have investigated their effectiveness in the context of primary school. This intervention study, therefore, examines the effect of anatomical models on the conceptions of inner organs in German primary school students (N = 45) in a pretest, posttest, and follow-up test design with anatomy teaching between pre- and posttests. Concepts were measured using students' drawings in two treatment groups (anatomical models versus anatomical illustrations). While in both treatment groups students' conceptions changed toward more scientific concepts, there was little difference in the changes between the two groups. There were even indications that the students of the control group showed more pronounced increases, for example, in long-term systemic integration of the organs. We discuss the reasons for this and furnish recommendations for effective teaching practices.

Keywords:

Conceptual Change, Models, Primary School, Science Education, Students' Drawings

Introduction

The ongoing COVID-19 pandemic has engendered anxieties and concerns among many children, decreasing their health-related quality of life and restricting their psychological well-being (Ravens-Sieberer et al., 2020). Children are now more vulnerable to adverse aspects of health, disease, respiratory failure, and ventilation compared to the time before the pandemic. Awareness about one's own body as well as health-promoting measures can counteract health-related anxieties (Gaudion, 2000). Knowledge about their organs enables children to take responsibility for their own health and lifestyle. Accordingly, curricula of primary school general studies (Sachunterricht) in Germany require addressing one's own body, its nutrition and care, to develop health-promoting attitudes and behaviors (MSB NRW, 2021). Primary school children have only fragmentary concepts about internal processes and human organs (Óskarsdóttir, 2006; Piko & Bak, 2006). Considering their classroom and everyday experiences, they develop – within the scope of their possibilities – ideas about what happens inside them in the event of an illness (Gesellschaft für Didaktik des Sachunterrichts, 2013).

Student Conceptions

"Student conceptions" refer to the concepts students develop before, during, and after lessons (Möller, 2018). This term has not yet been fully clarified in the literature (Krumphals et al., 2020). Expressions such as everyday conceptions, prior experiences, prior knowledge, preconceptions, or misconceptions are often used in similar contexts, the latter indicating a contradiction between the students' conceptions and scientifically viable concepts (Möller, 2018). In the following, the terms preconceptions and student conceptions are used to refer to students' mental representations of human internal organs and organ systems. Whereas initially research centered on preconceptions, nowadays, research purview extends to conceptions generated during teaching processes in addition to postconceptions (Möller, 2018). Preconceptions arise from primary everyday experiences and expressions, common thought patterns, and information and opinions influencing the individual (Möller, 1999). These are significant for scientific learning because, on the one hand, they can interfere with learning pathways and scientific findings, while on the other hand, they are the starting point of learning (Duit, 1997).

The distinction between "deep structure" concepts and "current constructions" is relevant for learning contexts. While the former concepts are anchored strongly in form of convictions and are thus highly resistant to change, the latter concepts are developed spontaneously in a given situation (Hartinger & Murmann, 2018). As preconcepts are based upon everyday conceptions that are often acquired over many years, they tend to be more deeply rooted than novel concepts developed after a few lessons (Barke, 2006). Therefore, preconcepts have a high resistance to change and can rarely be altered radically through teaching (Adamina et al., 2018).

One strategy to transform the often scientifically questionable preconceptions of primary school children into scientific conceptions is to provoke a cognitive conflict within the student (Grimm et al., 2020). Students have to realize that they cannot explain observable phenomena (e.g., respiratory rate at different loads) with their previous concepts of human organs and organ systems. Such a conflict motivates them to follow the lessons and develop new thought patterns (Barke, 2006). For a conceptual change to take place, four conditions must be met (Posner et al., 1982): First, there must be dissatisfaction with existing conceptions (Posner et al., 1982), i.e., the student has realized, for example, that his or her perceptions about the respiratory system are not sufficient to explain the lifting of the chest during inhalation. Second, the new concept must be comprehensible (Posner et al., 1982), i.e., teaching methods and media used must be suitable for adequately grasping the content, e.g., the structure and function of the respiratory tract. Third, the new concept should appear plausible (Posner et al., 1982). In other words, realizing the large volume of the lung as along with its ability to change in volume, for example, should help understand the lifting of the chest during inhalation. Fourth, compatibility is pertinent to conceptual change (Posner et al., 1982). For instance, understanding the respiratory system is essential for comprehending the importance of the cardiovascular system. During the process of conceptual change, new information is integrated into existing knowledge structures by expanding, differentiating, or even abandoning existing concepts to add new content to the preconcepts and replace the latter with adequate concepts. The extent of the transformation is expressed by terms such as conceptual growth, addition, revision, emplacement, or enrichment (Möller, 2015).

There are still relatively few studies on anatomical preconceptions about the human body of children under the age of seven years (Bietenhard et al., 2018). However, already seven-year-olds have knowledge of several organs. When asked to draw their organs, primary school children frequently mention heart, brain, and stomach (Stiftung Haus der kleinen Forscher, 2016) or skeleton, components of the cardiovascular system, and the gastrointestinal tract. Other organs such as liver or lungs are hardly ever mentioned (Manokore & Reiss, 2005; Spägele & Flintjer, 2011). Third-grade pupils reported in an interview that they have already seen illustrations of the brain, but their interpretations were noted to vary to a great extent. For example, when seeing an illustration of the cerebral cortex, they recognized tubers, pipes, or bubbles. Moreover, students were found to be unaware of the fact that the color coding of blood vessels depending on the oxygen content of the blood (red/blue) or the colored representation of brain areas does not correspond to the actual color of the vessels or the brain (Vocilka, 2005). About half of the children starting school were observed to be familiar with the function of the heart, the muscles, and the stomach, whereby the function of the heart is linked to the externally perceptible effect of beating (Spägele & Flintjer, 2011). The representation of the heart in children's and adolescents' drawings usually resembles a symbolic representation of the heart. Yet, it is not always clear whether children actually assume that this corresponds to the shape of the real organ or whether shapes of media representations are deliberately adopted. As 13- to 14-year-olds still prefer this form of representation, it can be assumed to be a deliberate symbolic representation (Gellert, 1962; Reiss & Tunnicliffe, 2001). On the whole, children are more familiar with organs that can be perceived directly or indirectly from the outside (e.g., heart, lungs) than those that cannot be felt (e.g., liver, kidneys). While primary school children can localize the position of the brain and the heart in the body relatively well, the position of the lung is often unknown (Óskarsdóttir, 2006). Spägele and Flintjer (2011) concluded that scientific perceptions are all the more developed the more direct and tangible they are in a child's environment. Inner processes or organs are hardly known and explained by deducing from outer to inner processes (Spägele & Flintjer, 2011). In children's minds, air, for example, is invisible, has no physical weight, and often only exists in a moving form as perceptible wind. From these perceptions, it can be concluded that there are no consistent ideas about the mechanism of respiration in the human body as well as the properties of air. This lack of conceptual knowledge has major consequences for other areas, such as the function of the lung (Spägele, 2008). An understanding of the systemic interrelationships between the organs is not fully developed neither in primary school (Óskarsdóttir, 2006) nor at the age of 15 (Stiftung Haus der kleinen Forscher, 2016). In general, only a few students are able to draw organ systems in a way that allows their classification as a system. Among them, the most frequently depicted are the digestive system and the respiratory tract (Reiss & Tunnicliffe, 2001). Likewise, understanding the blood circulation, including the path of the blood through the pulmonary circulation, is still difficult for students at lower secondary level (Riemeier et al., 2010). Even in other biological contexts, it can be observed that students face difficulties in linking concepts of different organization levels vertically (here: isolated organs and organ systems; Hammann, 2019).

Previous studies have confirmed that teaching human anatomy can develop primary school children's ideas about internal human organs toward more scientific ideas (e.g., Óskarsdóttir, 2006). However, it remains largely unclear which teaching methods or media give rise to this influence. One possible way to make new concepts easier to understand and more plausible for students is to use models (Posner et al., 1982).

Anatomical Models in Science Class

Models are central working and teaching aids both in natural sciences (Krüger et al., 2018) and in natural science teaching (Przywarra & Risch, 2019). On the one hand, they serve as media for conveying familiar content and, on the other hand, as tools for generating new knowledge (Krüger et al., 2018). The development of initial model conceptions of natural phenomena as well as becoming familiar with the interpretative character of models are therefore integral components of the basic science education (cf. Gesellschaft für Didaktik des Sachunterrichts, 2013). In contrast to the complexity of reality, a good model is characterized by the fact that it only depicts those essential characteristics for which it was created to gain knowledge of the scientific phenomenon or illustrate complex inter-relationships. Model selection should be both goal-oriented as well as situation- and addressee-oriented (Heitzmann, 2019). Through their characteristics, models enable children to recognize structural relationships and make it easier for them to learn challenging content in general studies (Hardy et al., 2004). In natural science lessons, models are typically employed as a medium for imparting knowledge. For example, a plastic structural model of a human body can be used as a teaching and learning tool to illustrate the difficult-to-access and hidden internal organs of the human body (Marika Haider, 2019; Stiftung Haus der kleinen Forscher, 2016). Furthermore, models have the potential to promote the development of transferable knowledge and long-term implementation of learning content as well as motivation (Marika Haider, 2019). In addition, students use models to help explain and articulate their understanding of scientific phenomena (Harrison & Treagust, 1993).

Even students' conceptual comprehension can be supported by a systematic use of models in the classroom (Dilber & Duzgun, 2008). Models achieve this by first creating dissatisfaction in children with their preconceptions and then providing an understandable and plausible explanation (Dilber & Duzgun, 2008). To achieve explanatory power, models used in the classroom should therefore be adapted to the background and living conditions of the students (Glynn et al., 1994). If this succeeds, non-viable preconceptions can be transformed into scientific concepts by using models in the classroom, as demonstrated by a control group study in primary schools general studies on electricity using models. Even lower-performing students benefit from the use of models (Michael Haider, 2010). Similarly, the studies by Möller et al. (2002) and Hardy et al. (2004) on the topic "swimming and sinking" in general studies at primary school show that the use of predetermined and student self-constructed forms of representation can stimulate a change of scientific concepts, because they help students abandon non-viable preconceptions.

However, models always represent a learning diversion, because they require the student to make a conclusion by finding resemblance of the model's features with the original. If this is not successful, the model cannot contribute to students' understanding (Marika Haider, 2019). A plastic torso used in class is clearly different from the real human body – and not only in terms of material. Nevertheless, it offers many visual links to the real human body such that the position, size, and appearance of the organs are easier for children to understand (Stiftung Haus der kleinen Forscher, 2016). However, in addition to a possible lack of analogy, there is a risk that students understand models as a replica of the original, although models are basically a simplified representation of reality (Heitzmann, 2019). On the one hand, this bears the risk of simplifying scientific relationships (e.g., the systemic relationships of human internal organs), and on the other hand, there exists the problem of a poor expandability of knowledge acquired from the model (Marika Haider, 2019). To counteract misconceptions and develop an adequate understanding of models and the original system they represent, it is, therefore, necessary to address the differences between a model and the original in class (Meisert, 2014; Vocilka, 2005). Therefore, when using anatomical models in general studies, model criticism should be practiced with students to counteract an unreflective use of models in the sense of model competence.

That the use of anatomical models in teaching can be effective is known from medical education, in which the use of anatomical models has a long-standing tradition (Narang et al., 2021; Talairach-Vielmas, 2014). From studies in this context, it is known that three-dimensional anatomical models – whether plastic models (Smith et al., 2018) or digital three-dimensional models (Haque et al., 2021; Nicholson et al., 2006; Zilverschoon et al., 2021) – can have a greater learning effect than two-dimensional representations. However, there is hardly any control group study on the effectiveness of the use of anatomical models in the context of human biology topics in general studies at primary school. This study was conducted with the aim to investigate whether the use of anatomical models can stimulate a conceptual change related to the knowledge of organs and their correct position, systemic integration, and appearance.

Method

Sample Size and Study Design

Forty-five third-grade students (N_{female} = 26, N_{male} = 19, average age = 8.7 years) from an elementary school in North Rhine-Westphalia, Germany, participated in this study, which was conducted in a pretest, posttest, and follow-up test design. The student participants were divided into two treatment groups (hereinafter referred to as model group, MG, N_{MG} = 24 and *control group*, CG, N_{CG} = 21). Both groups participated in five teaching units (seven hours in total) on internal organs of the human body between the pre- and posttests. Neither of them had been taught about the subject before. While the MG was taught using anatomical models, the CG was taught without haptic models but with two-dimensional illustrations of the internal organs. In the first teaching unit (one lesson), both groups dealt with the location of the internal organs using an anatomical plastic torso (MG) and a poster showing the position of the internal organs (CG). In the second teaching unit (two lessons), the human heart and lungs were discussed, with MG building a functional model of the lungs while CG worked out the function of the lungs with the help of illustrations. The third teaching unit (one lesson) linked the respiratory system with the cardiovascular system. In this lesson, the MG performed a role-play on blood circulation by assigning different roles to the students (e.g., heart and oxygen), whereas the students in the CG watched a two-dimensional explanatory film. In the penultimate teaching unit (two lessons), both treatment groups dealt with the human digestive tract exploiting the method of learning stations. While the MG executed some model experiments (for example, simulating the esophagus when pushing a marble through a garden hose), the CG worked with illustrated worksheets. In the final teaching unit (one lesson), both groups solved a quiz on the internal organs of the human body for revision. Pre- and posttests were directly implemented at the beginning of the first and at the end of the last lesson, respectively, with the follow-up test after four weeks.

Measuring Instrument

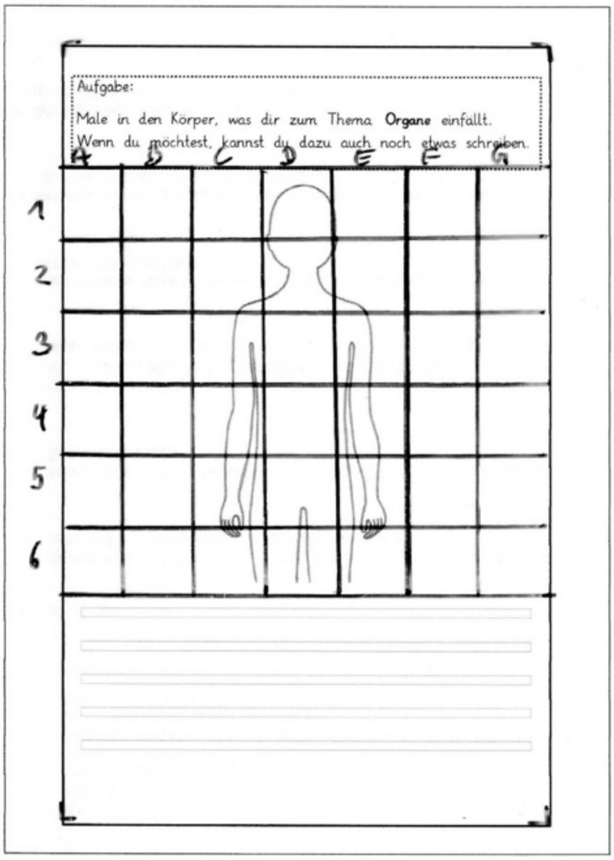
As it is not possible to directly measure students' conceptions (Adamina, 2008), students' drawings of human organs were used to capture their conceptions in a pretest, posttest, and follow-up-test. The use of children's drawings has already proven successful in various studies to measure thoughts and concepts about human anatomy (Bietenhard et al., 2018; Piko & Bak, 2006; Ranaweera & Montplaisir, 2010). In this study, the students were asked to draw the internal organs on a blank paper showing only the outline of the human torso. While this outline was primarily intended to help students with less artistic talent, in particular, to concentrate on the internal organs (Óskarsdóttir et al., 2011), it also gave us an objective frame of reference to evaluate the correct location of the individual organs. Drawings are not just a time-efficient way to measure concepts, they are interesting, motivate students, and enable international comparability as they do not depend on written or spoken language (Reiss & Tunnicliffe, 2001). Nevertheless, it makes sense to give students an additional opportunity to express themselves in writing. As this can facilitate the analysis of children's drawings (Adamina, 2008), we included some blank lines in the measuring instrument (Figure 1).

Data Analysis

All statistical analyses were performed using Statistical Package for the Social Sciences (SPSS) 28. Because students were free to draw any organs they liked, three experts in biology didactics developed an evaluation key to ensure the objectivity of the evaluation and avoid the risk of assessing artistic quality (Bietenhard et al., 2018; Reiss & Tunnicliffe, 2001). The key was initially divided into four main categories, namely, organ presence, organ position, systemic integration, and realism. Owing to spontaneously observed similarities between the drawings and the representations used in the treatments, the additional main category spontaneous observations was introduced later. Each main category was divided into subcategories, in the case of organ presence, organ position, and realism represented by several organs (heart, lung, stomach, etc.). The main category systemic integration was subdivided into the subcategories cardiovascular system, respiratory system, and digestive system. The main category spontaneous observations comprised the subcategories covered half similarity, dark spot similarity, vascular connection similarity, box-shaped similarity, and striking lines similarity. These are the conspicuous features on the representations used in the treatment: The human model torso used in the MG showed a half-covered intestine and a dark spot where the large intestine was incised, whereas the poster that was used in the CG showed the small intestine as wavy lines and the large intestine as box shaped. The poster also showed the heart with attachments of blood vessels. Each organ (or organ system in case of the main category systemic integration) in a main category was coded with two (organ presence: organ present/absent) or three variable expressions (organ position: correct position/wrong position/absent; realism: more realistic/less realistic (symbolic)/absent; spontaneous observations: similarity/dissimilarity/absent). To ensure objectivity in determining the organ positions, a grid was placed on the outline of the human torso during evaluation (Figure 1). The main category systemic integration was coded as correct systemic integration/incorrect systemic integration/missing systemic integration. A correct systemic integration was coded in case one or more correct connections between organs of an organ system were present in the drawing (e.g., if the stomach was connected to the intestine in the correct order).

Figure 1 depicts the measuring instrument, and the grid placed over it for ascertaining the position of the organs. For example, if the heart was drawn in D3, this was considered the correct position. The top of the worksheet displays task for the students (translated from German into English: Draw into the body whatever comes to your mind on the subject of organs. If you like, you can also write something about it.)

Figure 1
Measuring Instrument and the Grid Placed Over it for Determining the Position of the Organs



As the students were completely free in drawing, it was necessary to empirically verify the objectivity of the evaluation by calculating an inter-rater agreement (Moosbrugger & Kelava, 2020). For this purpose, a second rater evaluated 44.4% of the sample. Because our dataset consisted of categorical data, Cohen's kappa (κ) was used to calculate interrater reliability (Wirtz & Kutschmann, 2007). A value of κ < .00 is regarded as poor, .00–.20 as slight, .21–.40 as fair, .41–.60 as moderate, .61–.80 as substantial, and .81–1.00 as almost perfect (Landis & Koch, 1977). Because of inaccuracies in the first version of the evaluation key, the main categories systemic integration of the organs and representation of the whole intestine were again completely evaluated by the second rater and an additional 44.4% by a third rater. We report the final k for each measured characteristic in the results. Our data were categorical because of the fact that there was no theoretical reason to combine the subcategories into one scale, and the frequency of undrawn (missing) organs was found highly relevant for assessing changes in students' conceptions. Thus, the frequencies and changes in the frequency of each individual organ drawing between the measurement times were evaluated. Thereafter, we calculated the changes in frequency between (1) pre- and posttests and (2) pretest and follow-up test for each subcategory (e.g., for the frequency of correctly or incorrectly positioned hearts) as follows: For the main category organ presence, a change from absent to present was coded as appearance, a change from present to absent as disappearance and in all other cases as unchanged. In the case of the main category organ position, changes from "missing" or a wrong position in the pretest to a correct position in the posttest or follow-up test were coded as correction. Changes to a wrong position in the second test were coded as falsification, unchanged variable characteristics as unchanged, and the remainder as unidentifiable. The same procedure was used to evaluate the changes in the main categories systemic integration, realism, and spontaneous observations. We subsequently used these changes in a 3 x 2 and 4 x 2 contingency table comparison where "2" represents the two treatment groups and "3" and "4" are the number of variable characteristics. As there were one or more expected values of five or less in all contingency tables (Comiskey et al., 2014), Fisher's exact test was used to detect relationships of the conceptual changes between the treatment groups. Cramer's V (*V*) is given as a measure of effect size (small effect *V* = .1, medium effect *V* = .3, large effect *V* = .5; Cohen, 1988).

Results

Inter-rater reliability: We used kappa statistics to calculate the inter-rater reliability for each organ within each construct and time of measurement (see Table 1). Only eight of 105 calculations showed a fair agreement with Cohen´s κ between 0.27 and 0.40. In all the other cases, Cohen´s k ranged between 0.41 and 1.00, indicating a moderate-to-perfect agreement (Landis & Koch, 1977). In cases where one of the two variables used to calculate Cohen´s κ was a constant, the percentage of agreement was calculated with high values between 90 and 100.

Table 1

Cohen´s k Coefficient (N = 20) or, in Cases Where One of the Two Variables Used to Calculate Cohen´s k was a Constant, the Percentage of Agreement

Construct	Organ	Pretest	Posttest	Follow-up test
Organ presence	Heart	1.000	constant (100)	constant (100)
	Lung	.857	1.000	.773
	Stomach	.886	.886	.798
	Liver	.444	.528	.700
	Esophagus	.694	.694	.400
	Trachea	.444	.875	.612
	Intestine (whole)	.765	.857	.857
	Small intestine	constant (95)	.432	.612
	Large intestine	.643	.432	.600
	Brain	.900	1.000	1.000
	Blood vessels	.783	.900	.792
Organ position	Heart	.831	.692	.348
	Lung	.733	.394	.324
	Stomach	.431	.569	.649
	Liver	.459	.396	.442
	Esophagus	.450	.694	.447
	Trachea	.351	.610	.409
	Intestine (whole)	.556	.814	.583
	Small intestine	constant (90)	.423	.396
	Large intestine	.643	.423	.435
	Brain	.809	1.000	1.000
Systemic integration	Cardiovascular system	.902	.912	.906
	Respiratory system	1.000	.840	1.000
	Digestive system	.593	.551	.535
Realism	Heart	.524	.524	.700
	Lung	.857	.573	.462
	Lung details	.864	1.000	.835
	Esophagus	.724	.614	.400
	Brain	.900	.922	1.000
	Intestine (whole)	.672	.764	.665
Spontaneous observations	"Covered half" similarity	.765	.822	.725
	"Dark spot" similarity	.765	.780	.857
	"Vascular connection" similarity	.437	.266	.588
	"Box-shaped" similarity	.765	.837	.769
	"Striking lines" similarity	.765	.435	.612

Organ presence: The relative frequency of the organs drawn by both treatment groups is shown in Table 2. The heart, the brain, and the esophagus, in particular, were frequently drawn by both groups at the time of the pretest. The lungs and the stomach were more common in the CG, while the intestine was drawn more often in the MG. None of the groups differentiated between small and large intestine in their pretest drawings. In the posttest and follow-up test, not only did the students draw most organs more frequently, but they also drew a more detailed intestine by differentiating between the small and large intestine.

Table 3 shows the changes in the relative frequency of the organs from (1) pre- to posttest and (2) pretest to follow-up test taking into account the presence and absence of the organs. Fisher's exact test (p) and Cramér's V (V) are reported to determine relationships between group affiliation and changes in the organs' presence This table should be interpreted as follows: 8.3% of the MG drew a heart in the posttest without a heart being present in their pretest drawing. None of the students reduced the heart from their drawing from pre- to posttest. Consequently, 91.7% showed no change in the presence or absence of a heart. Taking into account the changes between the pre- and posttests, Fisher's exact test only showed one relationship to the treatment. Results showed a significant relationship between the treatment and the change in the presence of the liver with a more frequent occurrence of the organ in the MG ($p \leq .01$, $V = .466$). In the long run, however, CG seemed to have increased the incidence of trachea ($p \leq .001$, $V = .536$), small intestine ($p \leq .05$, $V = .339$), large intestine ($p \leq .01$, $V = .429$), and blood vessels ($p \leq .05$, $V = .419$) more often in comparison to the MG. Nevertheless, it must be taken into account that some students who drew a certain organ in the pretest no longer did so at the later measurement points, as can be clearly seen when looking at the blood vessels in the pretest/follow-up test, for example.

Organ position: The percentage of correctly and incorrectly positioned organs in the pretest, posttest, and follow-up test is shown in Table 4. This table should be interpreted as follows: 66.7% of the MG drew a heart at the correct position, whereas 25.0%

Table 2

Relative Frequency of the Organs Drawn of the MG and the CG for Each Time of Measurement

Organ	Pretest		Posttest		Follow-up test	
	MG	CG	MG	CG	MG	CG
Heart	91.7	90.5	100.0	100.0	91.7	95.2
Lung	29.2	42.9	91.7	90.5	79.2	85.7
Stomach	37.5	42.9	83.3	47.6	58.3	33.3
Liver	0.0	14.3	58.3	23.8	45.8	28.6
Esophagus	50.0	57.1	50.0	71.4	33.3	61.9
Trachea	25.0	28.6	50.0	71.4	4.2	61.9
Intestine (whole)	54.2	38.1	87.5	66.7	79.2	76.2
Small intestine	4.2	0.0	62.5	66.7	41.7	71.4
Large intestine	8.3	0.0	58.3	66.7	41.7	76.2
Brain	45.8	76.2	83.3	71.4	87.5	81.0
Blood vessels	33.3	23.8	45.8	42.9	41.7	42.9

Table 3

Percentage of Change in the Presence of the Organs (Presence/Absence) Between (1) Pre- and Posttests and (2) Pretest and Follow-up Test for the MG and CG

Organ	Pre-/Posttest			Pretest/Follow-up test		
	MG	CG	*p (V)*	MG	CG	*p (V)*
Heart	8.3/0.0	9.5/0.0	1.000	4.2/4.2	9.5/4.8	.791
Lung	62.5/0.0	52.4/4.8	.643	62.6/12.5	47.6/4.8	.289
Stomach	54.2/8.3	19.0/14.3	.061	37.5/16.7	19.0/28.6	.368
Liver	58.3/0.0	14.3/4.8	≤.01 (.466)	45.8/0.0	19.0/4.8	.084
Esophagus	33.3/33.3	23.8/9.5	.062	8.3/25.0	19.0/14.3	.543
Trachea	29.2/4.2	47.6/4.8	.476	0.0/20.8	42.9/9.5	≤.001 (.536)
Intestine (whole)	41.7/8.3	33.3/4.8	.727	33.3/8.3	38.1/0.0	.596
Small intestine	58.3/0.0	66.7/0.0	.759	37.5/0.0	71.4/0.0	≤.05 (.339)
Large intestine	50.0/0.0	66.7/0.0	.366	33.3/0.0	76.2/0.0	≤.01 (.429)
Brain	45.8/8.3	14.3/19.0	.074	50.0/8.3	19.0/14.3	.117
Blood vessels	29.2/16.7	38.1/19.0	.728	16.7/8.3	42.9/23.8	≤.05 (.419)

positioned it incorrectly. As a result, 8.3% of the MG did not draw a heart in the pretest. Around 50% to three quarters of both groups chose a correct position for the heart, brain, and esophagus. As a consequence of the low frequency in the pretest (see Table 2), only a few students positioned the small and large intestine correctly. For almost all organs, it can be observed that their positions became more correct from the pre- to the posttest.

Table 5 shows the changes in the correctness of the organ positions between (1) pre- and posttests and (2) pretest and follow-up test. Fisher's exact test (p) and Cramér's V (*V*) are reported to detect relationships between group affiliation and changes in the correctness. The table should be interpreted as follows: 58.3 % of the MG changed the position of the stomach from "missing" or an incorrect position in the pretest to a correct position in the posttest. None of the students changed it to a wrong position and 33.3% did not change the position of the stomach at all. The remainder, 8.4%, changed from a correct or incorrect position to "missing." Table 5 shows that a relationship with the treatment was found only in case of the liver ($p \leq .01$, $V = .466$). The students of the MG improved their positioning of the organ more noticeably than the students of the CG. The latter showed a stronger increase in the positioning of the trachea ($p \leq .01$, $V = .536$) as well as the small ($p \leq .05$, $V = .339$) and the large intestine ($p \leq .01$, $V = .429$) in the follow-up test.

Systemic integration of the organs: As shown in Table 6, there is only a relatively small number of students who connected the heart with blood vessels or the

Table 4

Relative Frequency of the Organs in the Correct and the Incorrect (Correct/Wrong) Positions in the Drawings of the MG and the CG at all Measurement Times

Organ	Pretest		Posttest		Follow-up test	
	MG	CG	MG	CG	MG	CG
Heart	66.7/25.0	57.1/33.3	87.5/12.5	71.4/28.6	87.5/4.2	81.0/14.3
Lung	16.7/12.5	33.3/9.5	79.2/12.5	71.4/19.0	70.8/8.3	76.2/9.5
Stomach	33.3/4.2	38.1/4.8	83.3/0.0	47.6/0.0	54.2/4.2	28.6/4.8
Liver	0.0/0.0	4.8/9.5	41.7/16.7	9.5/14.3	20.8/25.0	23.8/4.8
Esophagus	41.7/8.3	42.9/14.3	45.8/4.2	71.4/0.0	33.3/0.0	57.1/4.8
Trachea	25.0/0.0	23.8/4.8	50.0/0.0	71.4/0.0	4.2/0.0	52.4/9.5
Intestine (whole)	45.8/8.3	23.8/14.3	79.2/8.3	47.6/19.0	58.3/20.8	57.1/19.0
Small intestine	4.2/0.0	0.0/0.0	62.5/0.0	66.7/0.0	41.7/0.0	71.4/0.0
Large intestine	8.3/0.0	0.0/0.0	58.3/0.0	66.7/0.0	41.7/0.0	76.2/0.0
Brain	45.8/0.0	76.2/0.0	83.3/0.0	71.4/0.0	87.5/0.0	81.0/0.0

Table 5

Percentage of Changes (Correction/Falsification/Unchanged) in the Correctness of Organ Positions Between (1) Pre- and Posttests and (2) Pretest and Follow-up Test Shown for the MG and the CG

Changes	Organ	MG	CG	p (V)
Pre-/Posttest	Heart	29.2/12.5/58.3	28.6/19.0/52.4	.917
	Lung	66.7/12.5/20.8	47.6/14.3/33.3	.471
	Stomach	58.3/0.0/33.3	23.8/0.0/61.9	.070.
	Liver	41.7/16.7/41.7	9.5/4.8/81.0	≤.01 (.466)
	Esophagus	29.2/4.2/33.3	33.3/0.0/57.1	.130
	Trachea	29.2/0.0/66.7	47.6/0.0/47.6	.476
	Intestine (whole)	41.7/8.3/41.7	33.3/19.0/42.9	.725
	Small intestine	58.3/0.0/41.7	66.7/0.0/33.3	.759
	Large intestine	50.0/0.0/50.0	66.7/0.0/33.3	.366
	Brain	45.8/0.0/45.8	14.3/0.0/66.7	.074
Pre-/ Follow-up test	Heart	29.2/4.2/62.5	33.3/4.8/57.1	.936
	Lung	62.5/8.3/16.7	47.6/4.8/42.9	.287
	Stomach	33.3/4.2/45.8	19.0/4.8/47.6	.699
	Liver	20.8/25.0/54.2	23.8/0.0/71.4	.051
	Esophagus	8.3/0.0/66.7	33.3/4.8/47.6	.106
	Trachea	0.0/0.0/79.2	33.3/9.5/47.6	≤.01 (.536)
	Intestine (whole)	25.0/16.7/50.0	47.6/14.3/38.1	.336
	Small intestine	37.5/0.0/62.5	71.4/0.0/28.6	≤.05 (.339)
	Large intestine	33.3/0.0/66.7	76.2/0.0/23.8	≤.01 (.429)
	Brain	50.0/0.0/41.7	19.0/0.0/66.7	.117

lungs with the trachea in the pretest. The table should be interpreted as follows: 29.2% of the students in MG drew a heart connected to blood vessels (= correct systemic integration), 4.2% drew a heart separated from blood vessels (= wrong systemic integration). The remaining 66.6% did not draw a heart or blood vessels. Neither the rudiments of a cardiovascular system nor a respiratory system is, therefore, clearly recognizable in most drawings. However, about half of the sample connected components of the digestive tract with each other in their pretest drawings. There was a short-term increase in systemic drawing from pre- to posttest for the cardiovascular and the respiratory systems. The correct systemic integration seemed to be more stable in the CG than in the MG in the long term (see Table 6).

Table 7 provides detailed information about the improvements and deteriorations in the systemic integration. Fisher's exact test (p) and Cramér's V (V) are reported to detect relationships between the group affiliation and the change in the systemic integration. The table should be interpreted as follows: 25.0% of the MG changed their drawing of a heart to a more systemic integration, 12.5% to a less systemic integration, and 50.0 % did not change the systemic integration at all (and the absence of the organ) from pre- to posttest. The remaining 12.5% could not be classified as increase or decrease because of a missing heart in the posttest. No significant relationship was found between the treatments and the change in the systemic drawing of the digestive tract from pre- to posttest or from pretest to follow-up test. However, there were significant relationships in the pretest–follow-up changes in the systemic integration of the heart ($p \leq .01$, $V = .536$) and in the pretest–posttest ($p \leq .001$, $V = .571$) as well as pretest–follow-up changes in the systemic integration of the lungs ($p \leq .001$, $V = .700$). All cases had a large effect size with the CG showing a more pronounced increase in the systemic integration than the MG.

Realism of the drawing: The realism of the students' drawings is reported in Table 8. This table should be interpreted as follows: 4.2% of the MG drew a heart in the approximately realistic form, whereas 87.5% drew a less realistic (often symbolic) heart. The remaining 8.3% did not draw a heart in the pretest. In the MG, there was an increase in the frequency of realistic drawings from pre- to posttest and pretest to follow-up test for all organs except the esophagus. The frequency of realistic drawings in the MG was already high in the pretest, remained constant in the posttest, and decreased in the follow-up test.

Table 6

Relative Frequency of the Systemic Integration (Correct/Wrong Systemic Integration) of the Cardiovascular, Respiratory, and Digestive System for the MG and CG at All Times of Measurement

System	Pretest		Posttest		Follow-up test	
	MG	CG	MG	CG	MG	CG
Cardiovascular system	29.2/4.2	23.8/19.0	45.8/12.5	42.9/4.8	20.8/16.7	42.9/0.0
Respiratory system	16.7/8.3	0.0/0.0	29.2/12.5	61.9/14.3	4.2/16.7	61.9/4.8
Digestive system	54.2/8.3	33.3/9.5	50.0/29.2	33.3/23.8	37.5/20.8	14.3/38.1

Table 7

Percentage of Changes (More Systemic/Less Systemic/No Changes) in the Relative Frequency of the Systemic Integration of the Cardiovascular, Respiratory, and Digestive System Between (1) Pre- and Posttests and (2) Pretest and Follow-up Test for the MG and CG

Changes	Organ	MG	CG	*p (V)*
Pre-/Posttest	Cardiovascular system	25.0/12.5/50.0	33.3/4.8/38.1	.572
	Respiratory system	12.5/8.3/75.0	61.9/14.3/23.8	≤ 0.001 (.571)
	Digestive system	20.8/25.0/41.7	19.0/23.8/47.6	1.000
Pretest/ Follow-up test	Cardiovascular system	8.3/16.7/62.5	38.1/0.0/28.6	≤ 0.01 (.536)
	Respiratory system	0.0/12.5/70.8	61.9/4.8/33.3	≤ 0.001 (.700)
	Digestive system	12.5/20.8/50.0	9.5/33.3/47.6	.794

Table 8

Relative Frequency of More Realistic and Less Realistic (More Realistic/Less Realistic) Drawings for the MG and CG at All Times of Measurement

Organ	Pretest		Posttest		Follow-up test	
	MG	CG	MG	CG	MG	CG
Heart	4.2/87.5	0.0/90.5	50.0/50.0	61.9/38.1	29.2/62.5	52.4/42.9
Lung	8.3/20.8	0.0/42.9	33.3/58.3	47.6/42.9	12.5/66.7	28.6/57.1
Lung details	16.7/16.7	28.6/14.3	91.7/0.0	90.5/0.0	70.8/4.2	85.7/0.0
Esophagus	50.0/0.0	47.6/9.5	50.0/0.0	66.7/4.8	29.2/8.3	61.9/0.0
Brain	12.5/33.3	4.8/71.4	16.7/66.7	23.8/47.6	29.2/58.3	19.0/61.9
Intestine (whole)	29.2/25.0	4.8/33.3	41.7/45.8	9.5/57.1	37.5/41.7	14.3/61.9

Table 9 shows the development of the reality content from pre- to posttest and pretest to follow-up test. Fisher's exact test (*p*) and Cramér's V (*V*) are reported to detect relationships between the group affiliation and the change in realism. It should be interpreted as follows: 47.6% of the CG changed the realism of their lung drawing from "symbolic" or "missing" in the pretest to "realistic" in the posttest. Furthermore, 23.8% changed it from "realistic" or "missing" to a symbolic drawing, and 23.8% did not change realism (or absence) of their lung drawing. The increase or decrease of realism was not assessable for the remaining 4.8% because of the absence of lungs in their posttest drawings. A high percentage of students in both treatment groups increased the realism of their drawings. For some organs (e.g., the lungs and the brain in the MG), there was also a considerable increase in unrealistic, symbolic drawings. This does not mean that the students were getting worse overall, but it was because of the overall increase in the number of organs drawn after the treatment (see Table 2). A significant relation of group affiliation and changes in realism was shown only in the case of the brain with an increase of more symbolic drawings in the MG and of more realistic drawings in the CG ($p \leq .05$, $V = .424$) from pre- to posttest.

Spontaneous observations: A systematic examination of the initially randomly observed similarities between the drawings and the model torso or poster led to the results summarized in Table 10. This table should be interpreted as follows: None of the pretest drawings of the MG showed similarities to the model torso in the sense of a half-covered intestine, while 54.2% of the drawings showed an intestine without this similarity. The remainder 45.8% did not draw an intestine in the pretest. While no particular similarities between the drawings and the torso or poster were noticeable in the pretest, this changed in the posttest and remained until the follow-up test. About half of the drawings of the MG showed similarities to the intestine of the model torso used in their treatment. This means they drew an intestine that was half covered, as shown by the torso. There was no resemblance of this kind in the CG. In another striking feature (the incised large intestine, which looked like a dark spot in the model torso), there were only a few drawings of the MG that resembled the model torso in the posttest. However, it is still noticeable that in the posttest and follow-up test of the CG, conspicuous features of the poster used in this treatment frequently appeared in the drawings, that were not present in their pretest and that were not (in case of the box-shaped larger intestine and in case of striking lines in the smaller intestine) or considerably less present (in case of striking vascular connections at the heart) in the MG's drawings.

Table 9

Percentage Change in Realism (More Realistic/Less Realistic/Unchanged) of Organ Positions Between (1) Pre- and Posttests and (2) Pretest and Follow-up Test in the MG and CG

Changes	Organ	MG	CG	*p* (*V*)
Pre-/Posttest	Heart	50.0/8.3/41.7	61.9/4.8/33.3	.727
	Lung	33.3/41.7/25.0	47.6/23.8/23.8	.453
	Lunge details	79.2/0.0/16.7	61.9/0.0/33.3	.498
	Esophagus	33.3/0.0/33.3	28.6/4.8/57.1	.120
	Brain	8.3/45.8/37.5	23.8/9.5/47.6	≤.05 (.424)
	Intestine (whole)	29.2/33.3/29.2	9.5/28.6/57.1	.209
Pretest/ Follow-up test	Heart	25.0/4.2/66.7	52.4/4.8/38.1	.182
	Lung	8.3/54.2/25.0	28.6/33.3/33.3	.221
	Lunge details	66.7/0.0/20.8	57.1/0.0/38.1	.348
	Esophagus	12.5/8.3/54.2	28.6/0.0/57.1	.317
	Brain	20.8/41.7/29.2	14.3/14.3/57.1	.129
	Intestine (whole)	25.0/37.5/29.2	9.5/33.3/57.1	.149

Table 10

Relative Frequencies of Similarities (Similarity/Dissimilarity) to the Model Torso and to the Poster Used in the Treatment

Resemblance to	Pretest		Posttest		Follow-up test	
	MG	CG	MG	CG	MG	CG
the intestine on the torso (covered half)	0.0/54.2	0.0/38.1	37.5/50.0	0.0/66.7	29.2/50.0	0.0/76.2
the large intestine on the torso (dark spot)	0.0/54.2	0.0/38.1	8.3/79.2	0.0/66.7	0.0/79.2	0.0/76.2
the large intestine on the poster (box-shaped)	0.0/54.2	0.0/38.1	0.0/91.7	47.6/19.0	0.0/91.7	61.9/14.3
the small intestine on the poster (striking lines)	0.0/54.2	0.0/38.1	16.7/70.8	42.9/23.8	20.8/58.3	28.6/47.6
the heart on the poster (vascular connections)	0.0/91.7	0.0/90.5	0.0/91.7	57.1/42.9	0.0/91.7	61.9/33.3

The results of Fisher's exact test comparing the changes of similarities from (1) pre- to posttest and (2) pretest to follow-up test in both treatment groups confirm dependencies with the treatment (see Table 11). In the MG, a more substantial increase in similarities to the model corpus can be observed compared to the CG, both in the posttest and the follow-up test. On the other hand, in the drawings of the CG, features of the poster used in the treatment accumulate.

Discussion

The results of this study showed that already in the pretest about one-third to one-half of the students were able to draw some internal organs and position them correctly in the body. In line with previous studies, many students drew the heart, the brain, and individual organs of the digestive system. Internal organs without a directly noticeable external effect, such as the liver, were rarely present (Manokore & Reiss, 2005; Reiss & Tunnicliffe, 2001; Spägele & Flintjer, 2011; Stiftung Haus der kleinen Forscher, 2016). The fact that no organs of the reproductive system were drawn may be related to missing concepts or a taboo, as mentioned by Prokop and Fancovicová (2006). As the primary school students in this study have not yet had lessons on the inner organs, the results showed the relevance of their everyday experiences for their preconceptions. Accordingly, it is not surprising that students make relatively few systemic connections of the heart in the pretest as there are no externally visible connections of the cardiovascular system. At this point in time, the elements of the respiratory system are also barely connected. The less scientific ideas of many students were also indicated by the often symbolic representation of the internal organs such as the heart, in line with previous studies (Reiss & Tunnicliffe, 2001).

This can be explained by the fact that students obviously learn about organ systems at different ages (Bartoszeck et al., 2011; Reiss & Tunnicliffe, 2001). It seems as if they first learn that the human body consists of individual organs, then recognize their position, and only later understand that the organs are interconnected in order to form functional organ systems (Reiss & Tunnicliffe, 2001). Surprisingly, about one-third (CG) to one-half (MG) of the students connect at least individual organs of the digestive system (often esophagus and stomach). This finding is in line with the data of Garcia-Barros et al. (2011), who reported more adequate conceptions of four- to seven-year-old children about their digestive system than about their respiratory system. Possibly, kids ask themselves at a young age what happens to their food after the chewing and swallowing process. Also, it is more likely that they have experienced "tummy ache" more often than breathing problems at this young age level, which brings the stomach to their awareness. Another reason why parts of the digestive system might be more dominant in their conceptual thinking than the respiratory system could be that air is mostly perceived as invisible and therefore non-existent (Spägele, 2008). However, the drawings of the digestive system in this study are not very detailed in the pretest, as shown by the low frequency of a differentiated small and large intestine.

For most organs, as well as the digestive system, there is an increase in the frequency of drawn organs as well as correctly positioned organs in both treatment groups showing a positive effect of both treatments. However, this effect is more noticeable in the CG than in the MG, especially for the organs of the digestive system and trachea and in the long term. Furthermore, while some students improved the systemic integration of organs in the posttest, others deteriorated. This was the case for the intestinal system in both treatment groups. In the CG, however, a clear short-term (pre–post) as well as long-term (pre–follow-up) improvement in systemic integration of the respiratory system was observed. The observations of Carvalho et al. (2004) could provide an explanation for the better performance of the CG. They assumed that the absence of teaching nutrient absorption and the inadequately illustrated pathway of food from the stomach via the intestine to the anus in teaching material can give rise to confusion at the intestinal

Table 11

Percentage Changes in Similarities with the Model Torso and the Poster (similar/less similar/unchanged) from pretest to posttest or from pretest to follow-up test in the model group (MG) and the control group CG

Changes	Resemblance to	MG	CG	*p (V)*
Pre-/Posttest	the intestine on the torso (covered half)	37.5/37.5/16.7	0.0/33.3/61.9	≤.001 (.562)
	the large intestine on the torso (dark spot)	8.3/41.7/41.7	0.0/33.3/61.9	.505
	the large intestine on the poster (box-shaped)	8.3/37.5/45.8	47.6/4.8/42.9	≤.01 (.519)
	the small intestine on the poster (striking lines)	16.7/37.5/37.5	42.9/9.5/42.9	.071
	the heart on the poster (vascular connections)	8.3/8.3/83.3	57.1/4.8/38.1	≤.001 (.527)
Pre-/Follow-up test	the intestine on the torso (covered half)	29.2/25.0/37.5	0.0/38.1/61.9	≤.05 (.468)
	the large intestine on the torso (dark spot)	0.0/33.3/58.3	0.0/38.1/61.9	.596
	the large intestine on the poster (box-shaped)	4.2/33.3/54.2	61.9/4.8/33.3	≤.001 (.657)
	the small intestine on the poster (striking lines)	20.8/25.0/45.8	28.6/23.8/47.6	.757
	the heart on the poster (vascular connections)	0.00/4.2/91.7	61.9/4.8/28.6	≤.001 (.700)

March 2022, Volume 14, Issue 4, 539-555

level. Both the human torso and the poster used in our treatments include the digestive and the respiratory systems. In both media, the continuous path of the digestive system may remain unclear because of the way the extended intestine is represented. Especially in the three-dimensional model torso, the path of the food through the digestive system and the path of air to the respiratory system can hardly be traced. In particular, the path of the air from the mouth to the lungs is probably more apparent in the two-dimensional poster. However, this does not automatically mean a higher increase in realistic representations of the organs in the CG. Apart from the exclusively realistic depiction of the two lungs in both treatments in the short term (pre–post) and the long term (pre–follow-up), some students changed their drawings of the other organs to a more realistic depiction in both the short and long terms, while unrealistic drawings occurred in others. A significant relationship between change in realism and treatment group exists only in case of the brain, with a stronger increase in realism in the CG and a stronger increase in unrealistic, symbolic drawings in the MG. This is probably just a consequence of the plastic torso used in the MG which, in contrast to the poster of the control group, did not include a brain. This is a limitation of the present study that needs to be optimized in future studies. It would be interesting to see if there would still be more realistic drawings in the CG. This would be an indication that the two-dimensional poster probably has a more focusing, attention-grabbing effect, as this type of representation may link better to representations the students are familiar with. This kind of connection with prior knowledge is highly relevant for gaining the attention of primary school children (Shin & Shin, 2016). However, both, the models in the MG and the two-dimensional representations used in the CG, seem to have caught the students' attention. This was evidenced in the increase in similarities of the students' drawings with the model torso in the MG as well as with the poster in the CG. Such similarities of student drawings to models used in treatment were also observed by Copolo and Hounshell (1995) for high school chemistry classes. The unreflective adoption of models' unwanted accessories into students' conceptual thinking, e.g., drawing the semi-concealed intestine of the model torso or the rectangular intestine of the poster are from a scientific point of view undesirable effects. This finding also indicates that the ability to draw organs does not necessarily require knowledge about their functions—nor a correct inner positioning of the organs (Prokop & Fancovicová, 2006). This result is of relevance for practical anatomy teaching in primary schools as it emphasizes the need to support students in understanding external representations—such as physical models (Ingham & Gilbert, 1991) or graphical illustrations (Scaife & Rogers, 1996)—as analogies and not as mere replicas. The level of understanding an analogy and the effective integration of the information conveyed by an analogy correlates highly with the level of conceptual understanding of students (Mason, 1994). It should be supported by the teacher using supportive cues, prompting questions, or elaborations of the anatomical analogy (Richland et al., 2007) to make the analogy explicit to the students and prevent misleading students' learning process (Duit et al., 2001).

Potential differences in external representations (children's drawings) and internal mental representations of human organs are not just a limitation of this study but also a limitation in terms of measuring students´ conceptions of internal organs in general, as mental models only exist in the mind of the subject (Buckley & Boulter, 2000), and it is therefore questionable whether it is possible to externally measure them. In line with other studies (Óskarsdóttir, 2006; Reiss & Tunnicliffe, 2001; Riemeier et al., 2010), we used (two-dimensional) student's drawings to capture their concepts of internal anatomy. As the students were not to be influenced in their drawings, no support could be given to them, which may have made it difficult for them to translate their mental image of the inner organs into a two-dimensional drawing. Especially the finding that students need assistance in mentally transferring from three dimensions to two dimensions (Copolo & Hounshell, 1995) could be another reason for the results described above. It cannot be ruled out that some students of the MG, for example, obtained a differentiated concept of the inner organs because of the three-dimensional model used in the treatment but were unable to translate their mental concept to a two-dimensional drawing (Adamina, 2008; Prokop & Fancovicová, 2006; Reiss & Tunnicliffe, 2001). In this context, the spatial imagination of primary school students should also be taken into consideration, as it is probably not as high as that of adolescents (Yılmaz, 2009). Taking this into account, it would be interesting to modify the capturing of student concepts in future studies. For example, students could be asked to mold organs from plasticine and insert them into a physical model torso, or to draw three-dimensional organs in a virtual torso using augmented reality.

Furthermore, there is a risk of over-interpretation of the drawings, so additional written information can facilitate their interpretation (Piko & Bak, 2006). For this reason, we provided five lines in addition to the outline of the human body in the measuring instrument to give the students the opportunity to write down their ideas about the human organs. Although this option was rarely used, some students wrote down the names of the organs drawn, which was helpful for the interpretation. Nevertheless, some interpretations remained difficult, which is why we involved a second and, where necessary, a third rater to ensure inter-rater reliability. In future studies, the procedure could be further optimized by asking the students to label the organs they drew using horizontal lines. To optimize insights into treatment-dependent changes in student conceptions in future studies, it would be helpful to analyze the cognitive processes while drawing in real time using the "thinking aloud" method (Olson et al., 1984) or to conduct interviews as suggested by Prokop and Fancovicová (2006).

Other factors affecting the results could be the cognitive developmental level (Copolo & Hounshell, 1995), drawing ability (Prokop & Fancovicová, 2006), and the students' desire to draw. However, as the students in both treatment groups belonged to the same grade and no student submitted a blank sheet of paper, the relevance of these additional factors for the observed inter-treatment differences seem to be negligible. Nevertheless, these potential factors should be controlled in future studies.

Conclusion

With its experimental approach, this study complements previous research on primary school students' ideas about the human body. This is based upon the assumption that the use of physical anatomical models can contribute to a change in students' conceptions about inner organs compared to a CG. As the results show, the intervention not only led to a change in the student's conceptions of the presence, position, systemic integration, and realism of inner organs in the MG treatment but also in the CG. Significant differences in conceptual change between both treatments were only present in relatively few aspects but with more pronounced increases in the CG in terms of the long-term change of systemic integration of the cardiovascular and respiratory systems. There are theoretically derived indications that conceptual change depends on the type of representation offered in both treatments, which presumably supports the analogy conclusion to varying degrees. Future experimental studies will have to show which aspects of these anatomical representations account for the observed differences. This insight into students' conceptions is highly relevant to primary education as it can enable teachers to develop science lessons suitable for conceptual change in relation to human anatomy.

Acknowledgments

We acknowledge financial support by Land Schleswig-Holstein within the funding programme Open Access-Publikationsfonds.

References

Adamina, M. (2008). *Vorstellungen von Schülerinnen und Schülern zu raum-, zeit- und geschichtsbezogenen Themen: Eine explorative Studie in Klassen des 1., 3., 5. und 7. Schuljahres im Kanton Bern* [Dissertation]. Westfälischen Wilhelms-Universität, Münster.

Adamina, M., Kübler, M., Kalcsics, K., Bietenhard, S [S.], & Engeli, E. (Eds.). (2018). *„Wie ich mir das denke und vorstelle..." - Vorstellungen von Schülerinnen und Schülern zu Lerngegenständen des Sachunterrichts und des Fachbereichs Natur, Mensch, Gesellschaft.* Klinkhardt.

Barke, H.-D. (2006). *Chemiedidaktik: Diagnose und Korrektur von Schülervorstellungen*. Springer-Verlag.

Bartoszeck, A. B., Machado, D. Z., & Amann-Gainotti, M. (2011). Graphic representation of organs and organ systems: Psychological view and developmental patterns. *EURASIA Journal of Mathematics, Science and Technology Education*, *7*(1), 41–51. https://doi.org/10.12973/ejmste/75177

Bietenhard, S [Sophia], Gafner Knopf, A.-M., & Jaun-Holderegger, B. (2018). Befinden und sich kennen: Glück, Gesundheit, Körper. In M. Adamina, M. Kübler, K. Kalcsics, S. Bietenhard, & E. Engeli (Eds.), *„Wie ich mir das denke und vorstelle..." - Vorstellungen von Schülerinnen und Schülern zu Lerngegen-ständen des Sachunterrichts und des Fachbereichs Natur, Mensch, Gesellschaft.* (pp. 63–84). Klinkhardt.

Buckley, B. C., & Boulter, C. J. (2000). Investigating the role of representations and expressed models in building mental models. In J. K. Gilbert & C. J. Boulter (Eds.), *Developing Models in Science Education* (pp. 119–135). Springer, Dordrecht. https://doi.org/10.1007/978-94-010-0876-1_6

Carvalho, G. S., Silva, R., Lima, N., Coquet, E., & Clément, P. (2004). Portuguese primary school children's conceptions about digestion: Identification of learning obstacles. *International Journal of Science Education*, *26*(9), 1111–1130. https://doi.org/10.1080/0950069042000177235

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Erlbaum.

Comiskey, C., Coyne, I., Lalor, J., & Begley, C. (2014). A national cross-sectional study measuring predictors for improved service user outcomes across clinical nurse or midwife specialist, advanced nurse practitioner and control sites. *Journal of Advanced Nursing*, *70*(5), 1128–1137. https://doi.org/10.1111/jan.12273

Copolo, C. E., & Hounshell, P. B. (1995). Using three-dimensional models to teach molecular structures in high school chemistry. *Journal of Science Education and Technology*, *4*(4), 295–305. https://doi.org/10.1007/BF02211261

Dilber, R., & Duzgun, B. (2008). Effectiveness of analogy on students' success and elimination of misconceptions. *Latin-American Journal of Physics Education*, *2*(3), 174–183. https://dialnet.unirioja.es/descarga/articulo/2734605.pdf

Duit, R. (1997). Alltagsvorstellungen und Konzeptwechsel im naturwissenschaftlichen Unterricht: Forschungsstand und Perspektiven für den Sachunterricht in der Primarstufe. In W. Köhnlein, B. Marquardt-Mau, & H. Schreier (Eds.), *Forschungen zur Didaktik des Sachunterrichts: Bd. 1. Kinder auf dem Wege zum Verstehen der Welt* (pp. 233–246). Klinkhardt. https://scholar.google.com/citations?user=9ojlhuaaaaaj&hl=de&oi=sra

Duit, R., Roth, W.-M., Komorek, M., & Wilbers, J. (2001). Fostering conceptual change by analogies—between Scylla and Charybdis. *Learning and Instruction*, *11*(4–5), 283–303. https://doi.org/10.1016/S0959-4752(00)00034-7

Garcia-Barros, S., Martínez-Losada, C., & Garrido, M. (2011). What do children aged four to seven know about the digestive system and the respiratory system of the human being and of other animals? *International Journal of Science Education*, *33*(15), 2095–2122. https://doi.org/10.1080/09500693.2010.541528

Gaudion, C. (2000). Children's knowledge of their internal anatomy. In E. A. Glasper & L. Ireland (Eds.), *Evidence-based Child Health Care: Challenges for Practice* (pp. 148–162). Macmillan Education UK. https://doi.org/10.1007/978-0-333-98239-6_9

Gellert, E. (1962). Children's conceptions of the content and functions of the human body. *Genetic Psychology Monographs*, *65*, 293–405. https://www.semanticscholar.org/paper/Children's-conceptions-of-the-content-and-functions-Gellert/8e7b5cddd3a7bf06213880658697015b7ee4a618

Gesellschaft für Didaktik des Sachunterrichts. (2013). *Perspektivrahmen Sachunterricht*. Julius Klinkhardt.

Glynn, S. M., Law, M., Gibson, N. M., & Hawkins, C. H. (1994). Teaching science with analogies: A resource for teachers and textbook authors. *National Reading Resource*, *7*, 9–20. https://eric.ed.gov/?id=ED378554

Grimm, H., Todorova, M., & Möller, K. (2020). Schülervorstellungen in einem inquiry-orientierten Sachunterricht verändern – Besteht ein Zusammenhang mit der Förderung adäquaten Schlussfolgerns? *Zeitschrift für Didaktik der Naturwissenschaften*, *26*(1), 37–51. https://doi.org/10.1007/s40573-020-00110-1

Haider, M [Marika]. (2019). *Modellkompetenz Im Naturwissenschaftlichen Sachunterricht: Eine Empirische Studie Zum Lernen Mit Modellen und über Modelle in der Primarstufe*. Logos.

Haider, M [Michael]. (2010). *Der Stellenwert von Analogien für den Erwerb naturwissenschaftlicher Erkenntnisse: Eine Untersuchung im Sachunterricht der Grundschule am Beispiel "Elektrischer Stromkreis"* (1. Aufl.). *Klinkhardt-Forschung*. Klinkhardt. https://elibrary.utb.de/doi/book/10.35468/9783781550117

Hammann, M. (2019). Organisationsebenen biologischer Systeme unterscheiden und vernetzen: Empirische Befunde und Empfehlungen für die Praxis. In J. Groß, M. Hammann, P. Schmiemann, & J. Zabel (Eds.), *Biologiedidaktische Forschung: Erträge für die Praxis* (pp. 77–91). Springer Spektrum. https://doi.org/10.1007/978-3-662-58443-9_5

Haque, A., Minhas, R. S., Faisal, T., Naseem, S., Ambreen, S., & Rafi, A. (2021). Effect on academic performance by learning through 3D anatomy atlas versus 2D presentations. *Pakistan Armed Forces Medical Journal*, *71*(5), 1700–1704. https://doi.org/10.51253/pafmj.v71i5.6963

Hardy, I., Jonen, A., Möller, K., & Stern, E. (2004). Die Integration von Repräsentationsformen in den Sachunterricht der Grundschule. In J. Doll & M. Prenzel (Eds.), *Bildungsqualität von Schule, Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung* (pp. 267–283). Waxmann. https://www.uni-muenster.de/imperia/md/content/didaktik_des_sachunterrichts/dokumente/literaturmoeller/integration.pdf

Harrison, A. G., & Treagust, D. F. (1993). Teaching with analogies: A case study in grade-10 optics. *Journal of Research in Science Teaching*, *30*(10), 1291–1307. https://doi.org/10.1002/tea.3660301010

Hartinger, A [A.], & Murmann, L. (2018). Schülervorstellungen erschließen – Methoden, Analyse, Diagnose. In M. Adamina, M. Kübler, K. Kalcsics, S. Bietenhard, & E. Engeli (Eds.), *„Wie ich mir das denke und vorstelle..." - Vorstellungen von Schülerinnen und Schülern zu Lerngegen-ständen des Sachunterrichts und des Fachbereichs Natur, Mensch, Gesellschaft.* (pp. 51–62). Klinkhardt.

Heitzmann, A. (2019). Modelle verwenden. In P. Labudde & S. Metzger (Eds.), *Fachdidaktik Naturwissenschaft: 1.- 9. Schuljahr* (pp. 89–104). UTB.

Ingham, A. M., & Gilbert, J. K. (1991). The use of analogue models by students of chemistry at higher education level. *International Journal of Science Education*, *13*(2), 193–202. https://doi.org/10.1080/0950069910130206

Krüger, D., Kauertz, A., & Upmeier zu Belzen, A. (2018). Modelle und das Modellieren in den Naturwissenschaften. In D. Krüger, I. Parchmann, & H. Schecker (Eds.), *Theorien in der naturwissenschaftsdidaktischen Forschung* (1st ed., pp. 141–157). Springer. https://doi.org/10.1007/978-3-662-56320-5_9

Krumphals, I., Plotz, T., & Haagen-Schützenhöfer, C. (2020). Schülervorstellungen in den Naturwissenschaftsdidaktiken. In S. Habig (Ed.), *Naturwissenschaftliche Kompetenzen in der Gesellschaft von morgen: Naturwissenschaftliche Kompetenzen in der Gesellschaft von morgen. Gesellschaft für Didaktik der Chemie und Physik Jahrestagung in Wien 2019.* Universität Duisburg-Essen. https://www.pedocs.de/volltexte/2020/20445/pdf/habig_2020_naturwissenschaftliche_kompetenzen_in_der_gesellschaft.pdf#page=95

Landis, J. R., & Koch, G. G. (1977). The Measurement of observer agreement for categorical data. *biometrics, 33*(1), 159–174.

Manokore, V., & Reiss, M. J. (2005). Pupils' drawings of what is inside themselves: A case study in Zimbabwe. *Zimbabwe Journal of Educational Research*, *15*(1), 28–43. https://doi.org/10.4314/zjer.v15i1.26020

Mason, L. (1994). Cognitive and metacognitive aspects in conceptual change by analogy. *Instructional Science*, *22*(3), 157–187. https://doi.org/10.1007/BF00892241

Meisert, A. (2014). Mit Modellen lernen. In U. Spörhase-Eichmann & W. Ruppert (Eds.), *Biologie-Methodik: Handbuch für die Sekundarstufe I und II* (3rd ed., pp. 121–127). Cornelsen.

Möller, K. (1999). Konstruktivistisch orientierte Lehr-Lernprozessforschung im naturwissenschaftlich-technischen Bereich des Sachunterrichts. In W. Köhnlein, B. Marquardt-Mau, & H. Schreier (Eds.), *Vielperspektivisches Denken im Sachunterricht* (pp. 125–191). Klinkhardt.

Möller, K. (2015). Genetisches Lernen und Conceptual Change. In J. Kahlert, M. Fölling-Albers, M. Götz, A. Hartinger, S. Miller, & S. Wittkowske (Eds.), *utb-studi-e-book: Vol. 8621. Handbuch Didaktik des Sachunterrichts* (2nd ed., pp. 243–249). Klinkhardt; Verlag Julius Klinkhardt.

Möller, K. (2018). Die Bedeutung von Schülervorstellungen für das Lernen im Sachunterricht. In M. Adamina, M. Kübler, K. Kalcsics, S. Bietenhard, & E. Engeli (Eds.), *„Wie ich mir das denke und vorstelle..." - Vorstellungen von Schülerinnen und Schülern zu Lerngegen-ständen des Sachunterrichts und des Fachbereichs Natur, Mensch, Gesellschaft.* (pp. 35–50). Klinkhardt.

Möller, K., Jonen, A., Hardy, I., & Stern, E. (2002). Die Förderung von naturwissenschaftlichem Verständnis bei Grundschulkindern durch Strukturierung der Lernumgebung. In M. Prenzel & J. Doll (Eds.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen.: Zeitschrift für Pädagogik, Beiheft 45* (pp. 176–191). Beltz. https://www.pedocs.de/frontdoor.php?source_opus=3946

Moosbrugger, H., & Kelava, A. (2020). Qualitätsanforderungen an Tests und Fragebogen („Gütekriterien"). In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion.* Springer. https://doi.org/10.1007/978-3-662-61532-4

MSB NRW. (2021). *Lehrpläne für die Primarstufe in Nordrhein-Westfalen*. MSB NRW. https://www.schulentwicklung.nrw.de/lehrplaene/upload/klp_PS/ps_lp_sammelband_2021_08_02.pdf

Narang, P., Raju, B., Jumah, F., Konar, S. K., Nagaraj, A., Gupta, G., & Nanda, A. (2021). The evolution of 3D anatomical models: A brief historical overview. *World Neurosurgery*, *155*, 135–143. https://doi.org/10.1016/j.wneu.2021.07.133

Nicholson, D. T., Chalk, C., Funnell, W. R. J., & Daniel, S. J. (2006). Can virtual reality improve anatomy education? A randomised controlled study of a computer-generated three-dimensional anatomical ear model. *Medical Education*, *40*(11), 1081–1087. https://doi.org/10.1111/j.1365-2929.2006.02611.x

Olson, G. M., Duffy, S. A., & Mack, R. L. (1984). Thinking-out-loud as a method for studying real-time comprehension processes. In D. E. Kieras & M. A. Just (Eds.), *New Methods in Reading Comprehension Research* (pp. 253–286). Routledge. https://doi.org/10.4324/9780429505379-11

Óskarsdóttir, G. (2006). *The development of children's ideas about the body: How these ideas change in a teaching environment* [PhD]. University of Iceland, Reykjavík.

Óskarsdóttir, G., Stougaard, B., Fleischer, A., Jeronen, E., Lützen, F., & Kråkenes, R. (2011). Children's ideas about the human body – A Nordic case study. *Nordic Studies in Science Education*, *7*(2), 179–189. https://doi.org/10.5617/nordina.240

Piko, B. F., & Bak, J. (2006). Children's perceptions of health and illness: Images and lay concepts in preadolescence. *Health Education Research*, *21*(5), 643–653. https://doi.org/10.1093/her/cyl034

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*(2), 211–227. https://doi.org/10.1002/sce.3730660207

Prokop, P., & Fancovicová, J. (2006). Students´ ideas about the human body: Do they really draw what they know? *Journal of Baltic Science Education*, *2*(10), 86–95.

Przywarra, T., & Risch, B. (2019). Modelleinsatz im Chemieunterricht Illustrativ, haptisch-interaktiv oder digital erweitert. In C. Maurer (Ed.), *Naturwissenschaftliche Bildung als Grundlage für berufliche und gesellschaftliche Teilhabe: Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Kiel 2018* (pp. 616–619). Universität Regensburg.

Ranaweera, S. P. N., & Montplaisir, L. M. (2010). Students' illustrations of the human nervous system as a formative assessment tool. *Anatomical Sciences Education*, *3*(5), 227–233. https://doi.org/10.1002/ase.162

Ravens-Sieberer, U., Kaman, A., Otto, C., Adedeji, A., Devine, J., Erhart, M., Napp, A.-K., Becker, M., Blanck-Stellmacher, U., Löffler, C., Schlack, R., & Hurrelmann, K. (2020). Mental health and quality of life in children and adolescents during the COVID-19 pandemic-results of the copsy study. *Deutsches Arzteblatt International*, *117*(48), 828–829. https://doi.org/10.3238/arztebl.2020.0828

Reiss, M. J., & Tunnicliffe, S. D. (2001). Students' Understandings of human organs and organ systems. *Research in Science Education*, *31*(3), 383–399. https://doi.org/10.1023/A:1013116228261

Richland, L. E., Zur, O., & Holyoak, K. J. (2007). Mathematics. Cognitive supports for analogies in the mathematics classroom. *Science (New York, N.Y.)*, *316*(5828), 1128–1129. https://doi.org/10.1126/science.1142103

Riemeier, T., Jankowski, M., Kersten, B., Pach, S., Rabe, I., Sundermeier, S., & Gropengiesser, H. (2010). Wo das Blut fließt. Schülervorstellungen zu Blut, Herz und Kreislauf beim Menschen. *Zeitschrift Für Didaktik Der Naturwissenschaften*, *16*, 77–93.

Scaife, M., & Rogers, Y. (1996). External cognition: how do graphical representations work? *International Journal of Human-Computer Studies*, *45*(2), 185–213. https://doi.org/10.1006/ijhc.1996.0048

Shin, W.-S., & Shin, D.-H. (2016). An analysis of elementary students' attention characteristics through attention test and the eye tracking on real science classes. *Journal of The Korean Association For Science Education*, *36*(4), 705–715. https://doi.org/10.14697/jkase.2016.36.4.0705

Smith, C. F., Tollemache, N., Covill, D., & Johnston, M. (2018). Take away body parts! An investigation into the use of 3D-printed anatomical models in undergraduate anatomy education. *Anatomical Sciences Education*, *11*(1), 44–53. https://doi.org/10.1002/ase.1718

Spägele, E. (2008). *Naturwissenschaftliches Vorverständnis von Schulanfängern* [Dissertation, Pädagogische Hochschule Weingarten, Weingarten]. RIS. https://hsbwgt.bsz-bw.de/frontdoor/index/index/docid/34

Spägele, E., & Flintjer, B. (2011). Naturwissenschaftliches Vorverständnis von Kindern zum Zeitpunkt der Einschulung: Eine Studie zu außerschulisch erworbenen Kenntnissen und Vorstellungen. *Zeitschrift für Didaktik der Biologie (ZDB) - Biologie Lehren und Lernen*, *18*, 1–22. https://doi.org/10.4119/zdb-1641 (1-22 Seiten / Zeitschrift für Didaktik der Biologie (ZDB) - Biologie Lehren und Lernen, Bd. 18 (2011)).

Stiftung Haus der kleinen Forscher. (2016). *Forschen rund um den Körper*. Stiftung Haus der kleinen Forscher.

Talairach-Vielmas, L. (2014). Anatomical models: A history of disappearance? *Histoire, Médecine Et Santé*, *5*, 9–20.

Vocilka, A. (2005). *Schülervorstellungen zum Zustandekommenkognitiver Leistungen unter besondererBerücksichtigung der Wahrnehmung* [Diplomarbeit]. Pädagogische Hochschule Ludwigsburg, Ludwigsburg.

Wirtz, M., & Kutschmann, M. (2007). Analyse der Beurteilerübereinstimmung für kategoriale Daten mittels Cohens Kappa und alternativer Masse [Analyzing interrater agreement for categorical data using Cohen's kappa and alternative coefficients]. *Die Rehabilitation*, *46*(6), 370–377. https://doi.org/10.1055/s-2007-976535

Yılmaz, H. B. (2009). On the development and measurement of spatial ability. *Nternational Electronic Journal of Elementary Education*, *1*(2), 83–96. https://www.iejee.com/index.php/iejee/article/view/279/302

Zilverschoon, M., Custers, E. J., Cate, O. ten, Kruitwagen, C. L. J. J., & Bleys, R. L. A. W. (2021). Support for using a three-dimensional anatomy application over anatomical atlases in a randomized comparison. *Anatomical Sciences Education.* Advance online publication. https://doi.org/10.1002/ase.2110