A developmental psychology perspective on preschool science learning: Children's exploratory play, naïve theories, and causal learning

van Schijndel, T.J.P.

Publication date
2012

CHAPTER 4

Children's mental models of prenatal development

Van Es, S. E., Van Schijndel, T. J. P., Franse, R. K., & Raijmakers, M. E. J. *Children's mental models of prenatal development.* Manuscript submitted for publication.

ABSTRACT

Children's thinking about prenatal development requires reasoning about change that cannot be observed directly. How do children gain knowledge about this topic? Do children have mental models or is their knowledge fragmented? In Experiment 1, results of a forced-choice questionnaire about prenatal development filled in by 6-13 year olds' (N=317) indicated that children do have coherent, grade-related, theories about early shape of the fetus, but not about bodily functions. Coherence of the mental models was enhanced by a preceding generative task. Children's mental models were in agreement with reasoning about natural transformations (Rosengren et al., 1991) and constraints in representational flexibility (Karmiloff-Smith, 1992). In Experiment 2, an open-question interview was administered to 6-12 year old children (N=37). The interview resulted in incoherent responses that were not grade related. This study contributes to a deeper understanding of the effects of different methodologies being used in the area of mental models.

INTRODUCTION

Children gain a great deal of knowledge about the world around them during childhood (Carey, 1985). For example, children experience gravity with their body, they observe objects falling down, they may be told about gravity as a force, and they may see movies about weightless astronauts in space. To integrate some of this knowledge into a coherent, not necessarily scientifically correct, concept of gravity is known to be an extremely difficult task. The possible sources of this knowledge vary between topics (such as gravity, floating and sinking, the shape of the earth) and between children, depending on age and culture (e.g., Springer, 1999; Harris & Koenig, 2006; Gelman, 2009). A fundamental question regarding each topic is whether children do form coherent ideas, i.e. mental models, or whether their knowledge about the topic remains fragmented until they learn about a scientific concept. To gain a better understanding of knowledge development in general, it is important to gain a better understanding of how the development of knowledge takes place for specific topics.

The aim of the present study is to examine the development of children's knowledge about prenatal development. Children may have multiple sources of knowledge about prenatal development, such as facts about prenatal development that are being told, pictures or movies from a fetus in a pregnancy book, or knowledge about the development of other living things including themselves. In naïve biology, most researchers state that children develop some coherent ideas, although they may disagree about the origin of these ideas. Some researchers link innate predispositions to the early formation of biological theories (Gelman & Hirschfeld, 1990; Keil, 1992; Keleman, 1998; Mandler & McDonough, 1993). Other researchers emphasize the importance of experience (Au & Romo, 1999; Carey 1985; Hatano & Inagaki, 1994; Inagaki, 1997; Springer, 1999) and state that the main principle for theory formation is the acquisition of some factual knowledge combined with some key inferences from this knowledge.

The literature on children's naïve ideas about prenatal development is scarce. Most studies focus on either conception (e.g. Bernstein & Cowan, 1975) or birth models (e.g. Nagy, 1953; Kreitler & Kreitler, 1966). Discussing the creation of babies, Kreitler and Kreitler (1966) argued that children's (aged 4;0 to 5;6) views are based on three theories: (1) the baby is created in the mother's belly from the food she eats; (2) the baby has always existed in the mother's belly; or (3) the baby is swallowed by the mother. The more specific topic of prenatal development has rarely been focused on, however. Zoldosova and Prokop (2007) presented a study with a small population that included a qualitative analysis of drawings and structured interviews. They concluded that four broad categories existed in children's drawings about prenatal development: 1) drawings of a fetus without development, 2) drawings of a pregnant woman with a fetus in her belly, but without development, 3) drawings of a pregnant woman with a developing fetus and 4) drawings of pregnancy development (without further specifying what development means). Furthermore, children showed large variation in the body parts that they drew. For example, 70% (N=20) of the children drew an umbilical cord, equally distributed over age (1st – 4th grade). From the interviews only qualitative observations were reported, e.g. that children had very limited knowledge about the fetus' insides although the older children mentioned the umbilical cord in the context of food uptake.

Two additional topics in naïve biology are of particular interest for the present study: children's understanding of biological growth and transformations, and children's understanding of bodily functions. Preschool-age children already have naïve ideas about biological growth (Carey, 1985; Inagaki & Hatano, 2004; Gottfried, 2005). Using a forced-choice experimental paradigm, Rosengren, Gelman, Kalish, and McCormick (1991) showed that preschoolers' (starting at the age of 3) ideas about natural transformations are coherent. They understand that animals get larger and not smaller with age, but only older children and adults accept rather dramatic changes of color and shape. Reasoning about human insides is an established and larger area of research (e.g., Gellert, 1962; Carey, 1985), which concerns the type of explanations children give for bodily phenomena, for example: intentional, vitalistic, or mechanical causality (Inagaki & Hatano, 2004; Hatano & Inagaki, 1999). In order to reason about the changes in bodily functions during prenatal development, children should understand that some bodily functions are important for staying alive. Several studies (Gellert, 1962; Jaakola, 1997) show, by using interviews, that preschoolers (4-6 years) know that humans need to breath air, to eat, and to drink to stay alive. However, these studies also show that children have very little knowledge about human insides, such as the lungs and the digestive system. Large part of the development of knowledge about the mechanistic causality of people's insides takes place in later childhood until adulthood (Morris, Taplin & Gelman, 2000).

One important way of reasoning about unknown aspects of prenatal development might be by projecting known properties from other living things, that is, from postnatal humans, animals, or plants. Inagaki and Hatano (1996; Hatano & Inagaki, 1994; Inagaki & Hatano, 2004) conclude that at the age of five most children have an integrated category for living things, that is, for animals and plants together. Five-year old children report that growing, taking food/water, and aging/dying are similar phenomena for animals and plants.

The literature shows that children have some factual knowledge about a newborn, about the shape of its body and about the importance of its bodily functions, such as eating, drinking, and breathing. Hence, if a child predicts some properties of prenatal bodily shape and bodily functions the child accepts the transformations of these fetal properties into the properties of a newborn. However, the ability of children to modify their concepts, such as their concept of a newborn, is believed to be dependent on the flexibility of their internal representations. In her representational redescription (RR) theory Karmiloff-Smith (1990) describes the acquisition of knowledge through sequential phases, starting with a representation of knowledge in a procedural, implicit way, followed by re-representations at different levels of abstraction. According to RR theory, the representational flexibility is dependent on the level of representation. This theory has found support in several areas of cognitive development, such as language, mathematics, and physics (Karmiloff-Smith, 1992, Critten, Pine & Steffler, 2007; Hollis & Low, 2005). The sequence of phases is found to be domain general. However, the individual level of representational flexibility is believed to differ between specific topics (i.e., micro domains).

Karmiloff-Smith (1990; Spensley & Taylor, 1999) used children's drawings to explore representational flexibility. She focused on the constraints children have in their representational flexibility by looking at what changes children make in their representation of a concept (e.g., a house or a human being), when asked to draw a non-existing concept (e.g. a fake house or a fake human being). Karmiloff-Smith's results of the 'non-existing human being' task, show that changes introduced by the younger children (aged 4-6) mainly involved deletions and changes in size and shape, whereas older children (aged 8-10) often changed position and orientation of elements and added elements from other conceptual categories. Recent discussions on the RR theory have mainly focused on whether the scores on the drawing task are directly related to procedural rigidity of drawing skills and/or due to inflexibility of mental representations (Barlow, Jolley, White & Galbraith, 2003; Picard & Vinter, 2007).

Based on former research in naïve biology and taking into consideration RR theory (Karmiloff-Smith, 1990), we predict children to have the following mental models of prenatal development. I) A null model of prenatal development would be *no change*, that is, a fetus has the same shape as a newborn and has equivalent bodily functions. II) A *growth* model could be growth of the bodily shape and gradual increase of bodily functions. However, what children observe in postnatal development is that the shape grows and that bodily functions stay equal. Both constant and increasing bodily functions would agree with children having limited representational flexibility but have the understanding of *growth*. III) An alternative naïve idea about prenatal development could be the *outgrowth* of body elements and the *emergence* of functions, analogous to a growing plant with outgrowing branches. Some children might have directly observed how a seed grows into a plant and take that as an analogy for how an egg grows into a newborn. A transformation that consists of the addition of elements agrees with a higher level of the RR theory. IV) A child in an even later representational stage of its concept of a newborn would allow for more dramatic changes. It would accept that the shape of a hand of a fetus looks *different* from the hand of a newborn and that some bodily functions are implemented differently.

Mental models versus fragmental knowledge

Most authors claim that children have coherent ideas about biological topics, but only in a few cases is this claim tested experimentally (e.g., Rosengren et al., 1991). The question of whether children's knowledge about specific topics is coherent and represented in mental models or whether such knowledge is fragmented has been a topic of heated debate in the literature. Theorists on the coherence side of this debate maintain that children's naïve ideas are organized into coherent and consistent theories, which structure everyday thinking and are more or less resistant to change (Wellman & Gelman, 1998; Johnson-Laird, 1983). Several different domains of knowledge have been studied within this approach: children's mental models of the shape of the earth (Vosniadou & Brewer, 1992); the motions of the earth, sun, and moon; the relative location of the earth, sun, and moon in space and the day-night cycle (Samarapungavan, Vosniadou, & Brewer, 1996); and, children's frameworks on evolution and speciation (Samarapungavan & Wiers, 1997). A more specific issue that is central in the debate is whether children form mental models 'on the spot'. According to Vosniadou, Skopeliti, and Ikospentaki (2004) children form on-the-spot dynamic situation-specific representations for the specific purpose of answering questions that are being posed to them. They argue that tasks like drawing, making clay models or responding to open-ended questions encourage children to make generative use of the scientific information that they have at their disposal, which encourage on-the-spot formation of mental models, thereby increasing the coherence of these models.

In contrast to the coherence theorists, researchers on the other side of the debate claim that children's naïve knowledge is fragmented. DiSessa (1988), for instance, claims that intuitive physics stems from fragmented knowledge, represented like a set of loosely connected ideas. In line with this view, some researchers have concluded that children's knowledge of the earth is fragmented (Straatemeier, Van der Maas & Jansen, 2008; Nobes et al., 2003).

A complicated aspect of this discussion is that differences in conclusions are related to the used experimental methodologies. Vosniadou et al. (2004) argued that the best way to investigate children's knowledge is by making use of generative methods, such as drawing. Several points of criticism exist concerning the use of drawings in the study of mental models. Firstly, the constraints on the planning of drawings may hamper children in divulging mental models (Nobes, Martin & Panagiotaki, 2005). A second criticism on the use of drawings lies in the difficulty of establishing objective scoring methods (Brainerd, 1973, in Straatemeier et al., 2008). However, results of drawings can be valuable when acceptable levels of inter-rater reliability are maintained. Thirdly, the consequence of assigning one score from a predefined range of models is that no unexpected mental models can be observed from the data. Finally, when aiming to test the existence of mental models, a measure of coherence is needed. Coherence can only be measured when different comparable elements or items are used, such as in a questionnaire or in an interview, but not in one drawing.

Forced-choice questionnaires have been one of the methods of choice within the area of mental models of the earth (Nobes et al., 2003; Straatemeier et al., 2008; Frède et al., 2011;), leading to the conclusion that children do not have mental models of the shape of the earth. Vosniadou et al. (2004) criticized this method for several reasons. Firstly, responses to the forced-choice questionnaire may be biased because the choice of response options is limited. This problem can be overcome if the construction of the test is based on extensive piloting on possible alternative models. Secondly, because children only need to recognize correct information, they might perform better on forced-choice questionnaires. A solution to this problem could be the use of an additional open-ended interview.

The use of open-ended interviews is not without problems, either. In order to arrive at a full understanding of the underlying conceptual structures, similar questions addressing the same issues are asked, resulting in a prolonged method of repeated questioning. This unnatural situation may confuse children, because everyday conversational rules do not apply (Siegal, 1997). In order to circumvent this problem, the structure of the interview should be designed in such a way that sufficient responses can be collected, but a normal communication situation can exist. Reducing the amount of follow-up questions is a way of doing this. Another criticism Siegal emphasizes in the use of interviews, which is applicable to all verbal methods, is that children might not be familiar with the words used in the interview. This bias can be reduced by using a model of the topic at hand to facilitate the understanding of the child.

To conclude, the most favorable methodology to examine the existence of mental models of prenatal development is either a forced-choice questionnaire in which all perceptible mental models are represented or an interview set up in such a way that conversational rules apply and a model (a doll of a newborn) is used to facilitate the child's understanding of the question. With the former method coherence can be measured best because we can apply a larger number of comparable items and we can study a larger number of children. In the current study, both methodologies were used. Furthermore, these methodologies were compared to each other. In experiment 1 a questionnaire was designed in which pictures were used instead of written response categories, to avoid problems related to verbal methodologies. In experiment 2, in addition to the forced-choice questionnaire, an interview was administered. We examined whether children revealed the same mental models in an interview as in the questionnaire. In the questionnaire and the interview we asked similar questions. In addition, a drawing task on prenatal development was included in the design of both experiments. First, we examined the effect of a preceding generative task on the coherence of mental models with a drawing task. Moreover, by assigning the drawings to a mental model according to well-described scoring rules, we studied the relation between the results of the three assessments methodologies.

EXPERIMENT 1

METHOD

Participants

We tested 317 children (age: *M*=9.35, *SD*=1.84) from three different primary schools in different parts of the Netherlands, all providing regular education. An opt-out consent procedure was used, parents being required to sign and return a form if they did not want their child to participate in the study. The sample consisted of children between 6 and 12 years old: participants included 43 children from Grade 1, age 6-7 (*M*=6.56, *SD*=0.59); 50 children from Grade 2, age 7-8 (*M*=7.74, *SD*=0.60); 68 children from Grade 3, age 8-9 (*M*=8.75, *SD*=0.63); 47 children from Grade 4, age 9-10 (*M*=9.91, *SD*=0.67); 51 children from Grade 5, age 10-11 (*M*=10.84, *SD*=0.51); and 58 children from Grade 6, age 11-12 (*M*=11.76, *SD*=0.60).

Materials

Questionnaire. The questionnaire was designed for children aged six and older. In the process of designing the questionnaire, Siegler's (1981) rule assessment methodology (RAM) was used. Rules in Siegler's RAM framework are equivalent to mental models in this study. It was expected that a child with a specific mental model would generate a specific response pattern for a set of items. Items were constructed such that maximum differentiation was possible between the expected mental models: 1) the no-change model, 2) the growth model, 3) the outgrowth model, and 4) the different model. The themes of the questions used in the questionnaire were based on children's misconceptions of prenatal development reported in earlier studies (Zoldosova & Prokop, 2007; Gellert, 1962; Jaakola, 1997) in group discussions and children's drawings at an early science education workshop.[1]

The questionnaire started with a trivial example item, followed by fifteen questions on prenatal development. These questions were divided to cover an early stage and a late stage of prenatal development. In turn, the items within this division concerned either shape of the body or the bodily functions of the fetus. Topics covered in the shape-related items were the hand, the leg, the ear, the foot, the arm, and the eye. There were 6 early-shape items (ES) and 3 late-shape items (LS: hand, leg, ear). Topics covered in the function-related items were breathing, drinking and maintaining body temperature. There were 3 early-function items (EF) and 3 late-function items (LF). In the booklet the response options of the shape-related items were depicted with adapted photographs. The response options of the function-related items were given in a few short, easy to read words. The order of the response options was randomized. A translated example of 2 items, an ES and a LF item, are depicted in Fig. 1. The experimenter said the following while presenting the ES item in front of the classroom:

This is what the hand of a baby looks like, when it has just been born. Which picture resembles the hand of a baby most, when the baby still has to stay in the belly of the mother for a long, long time?

a) A hand looking exactly the same (no-change response*)*
b) A hand looking exactly the same, but a bit smaller (growth response*)*
c) A hand looking like the hand of a newborn, but it does not have all the fingers yet (outgrowth response*)*
d) A hand that looks different from the hand of a newborn with a different shape (different response*)*

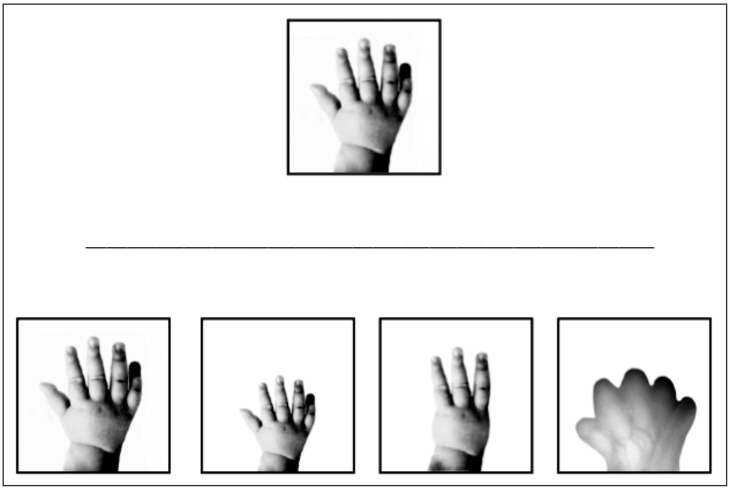
The experimenter said the following while presenting the LF item in front of the classroom:

When a baby has just been born it needs to breath air to stay alive. How does the baby get air just before it is born?

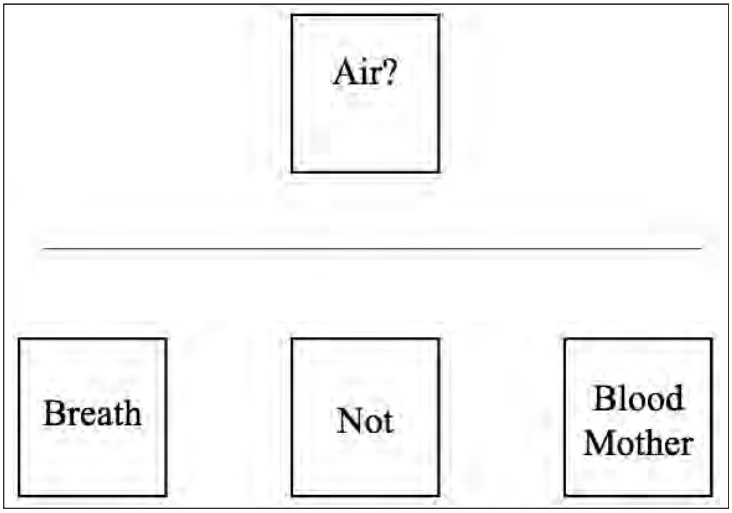
a) The baby breaths air just like a newborn (no-change response/ growth response)
b) The baby does not need anything from the air (absence response*)*
c) The baby gets something from the air through the blood of the mother (different response*)*

Drawing task. Participants were asked to make two drawings of a fetus: one in an early stage of development and one in a late stage of development. The assignment for the first drawing was 'draw a baby in the belly of the mother, for the case when the baby has to stay in the belly of the mother for a long, long time'. The assignment for the second drawing was 'draw a baby in the belly of the mother, right before it is born'. On A3 sized paper children received a format depicting three pictures of a mother, the first two with an empty square on the belly of the mother where the baby had to be drawn. The third picture showed a mother together with her newborn.

The drawings were scored based on the types of changes that appeared between the first and the second drawing. The main elements of change were scored, presented by growth from early to late stage of fetal development, deletion in the early stage as compared to the late stage, and changes in shape from early to late stage. Drawings were sorted into four



a)



b)

FIGURE 1. a) Shape item about the hand used in the questionnaire. b) Function item about breathing. Of each item the upper figure depict the shape (function) of a newborn; the lower figures are the four response options for shape items (no-change, growth, outgrowth, and different) or three response options for function items (different, no-change, absence). In the questionnaire the response options were presented in random order. Figure b) is a translation of the original Dutch item.

categories related to the predicted mental models; 1) *no-change*, 2) *growth*, 3) *outgrowth*, and 4) *different*. The change between the drawings representing the highest level of conceptual change was the criterion for deciding to which category the drawing belonged. For example, if both growth and a different shape were apparent in a drawing, the child was classified in category 4.

Design and procedure

The experimental procedure consisted of two parts: completing the questionnaire and the drawing task. These two tasks were administered group wise in a classroom setting. The children were tested under two conditions: either they first filled out the questionnaire followed by the drawing of the fetus (151 participants), or they first drew the fetus then completed the questionnaire (166 participants). The instructions of all tasks were read out loud by the test leader, as well as the questions of the questionnaire. Additionally, an enlarged version of all questions was projected on a screen. The subtests took about twenty minutes each, so the total procedure lasted for an hour after which children were given the opportunity to ask questions.

Statistical analyses

To detect mental models from responses of forced-choice questionnaires a widely used methodology is Siegler's rule-assessment methodology (Siegler, 1981). For detection of strategies from series of responses, Siegler matched patterns of observed responses to theoretically expected response patterns. Subjects were assigned to a mental model (i.e., rules for solving the balance scale task in Siegler 1981) according to the best matching expected pattern, provided that the match exceeded a predefined threshold (and possibly additional criteria). The number of successfully classified subjects is taken as an index of the fit of the rule system.

However, the use of pattern matching in the analysis of mental models has several shortcomings (Van der Maas & Straatemeier, 2008). A considerable problem is that the assignment of subjects to mental models takes place by an arbitrary criterion of correspondence between observed and expected responses. Moreover, only mental models that are a priori defined can be detected, meaning that the detection of alternative, unexpected, models is not possible. Finally, pattern matching does not provide statistically motivated criteria for the selection of the most parsimonious, best fitting model (e.g., to decide whether additional mental models should be included into the descriptive model of the data).

Latent class analysis (LCA; McCutcheon, 1987; Visser, 2011) provides a statistically more reliable method to detect rules (Jansen & Van der Maas, 1997) or mental models (i.e., coherent ideas; Straatemeier et al., 2008) from nonverbal responses. LCA is a standard statistical technique for models with categorical manifest variables and categorical latent variable. Latent class models define categorically different response patterns (latent classes). In our study a latent class can be interpreted as a response pattern related to a specific mental model. A latent class model is defined by unconditional probabilities and conditional probabilities, which are estimated from the data. Unconditional parameters define the probability of belonging to a latent class, i.e. the size of the class. Conditional probabilities are defined for each class and represent the probability of each answer option, given that the subject belongs to the class. To calculate maximum likelihood estimates of the parameters, we used the DepmixS4 package (Visser & Speekenbrink, 2010) for the R program for statistical computing (R Development Core Team, 2010). Model selection criteria, especially the Bayesian Information Criterion (BIC; Schwartz, 1978), were used to choose the optimal, most parsimonious model, i.e. the optimal number of classes that described the data. If children have a mental model of prenatal development, we should find a limited number of latent classes with a pattern of conditional probabilities that can be interpreted as mental models. Because response patterns of classes do not need to be formulated beforehand, it is possible to detect alternative mental models. Obviously, the design of the questionnaire restricted the models we could find, but this is inherent in all closed form assessment tools.

RESULTS

General

The internal consistency of the entire questionnaire as expressed by Cronbach's Alpha was .82 with the response options taken as an ordinal variable (*no-change*, *growth*, *outgrowth*, *different*). Nonparametric (Spearman's rho) correlations between items varied between .49 and .79 for the early-shape items (ES), between .05 and .18 for the early-function (EF) items, were between .90 and .94 for the late-shape (LS) items, and were .01 and .20 for the late-function (LF). The reason for the low correlations between functions items will be explained through the latent class analyses in the next section. Table 1 shows the distributions of responses. For the ES items, the response frequency of the *no-change* response was very low. Hence, in further analysis the *no-change* response was combined with the *growth* response for the ES items. We model the three LS items separately.

The function items (EF and LF) showed little variance in responses and there was no striking difference in frequencies between EF and LF. Correlations between items on EF and LF were between .67 and -.14. We modeled these six function items together in a separate latent class analysis, because we wanted to observe whether children responded coherently between the early and late stage. For the latent class analyses, the data of the two conditions, that is the two task sequences, were taken together.

TABLE 1. Distributions of responses for the questionnaire items in Experiment 1.

Item	No-change	Growth	Outgrowth	Different
Early Shape				
Hand	2.5	36.6	2.8	58.0
Leg	1.9	39.6	13.0	45.6
Ear	0.3	46.4	26.8	26.5
Foot	0.3	49.5	2.8	47.3
Arm	0	37.9	11.0	51.1
Eye	0.9	36.0	14.5	48.6
Early Function				
Drinking	2.2		1.6	96.2
Keeping Warm	0.3		3.2	96.5
Breathing	15.5		11.0	73.5
Late Shape				
Hand	52.1	46.7	.9	.3
Leg	51.4	47.6	.6	.3
Ear	51.3	48.1	0	.6
Late Function				
Drinking	6.3		1.9	91.8
Keeping Warm	1.3		1.9	96.8
Breathing	15.5		9.5	75.1

Note. Frequencies are given as percentages of children (N=317).

Mental models of bodily shape

To examine whether children showed mental models about bodily shape in the early stage, latent class analysis was first performed on the responses of the six ES items. If children had mental models, a limited number of latent classes should be found with conditional probabilities consistent with mental models, such as those described in the introduction. Table 2 (upper section) shows the fit statistics of the explorative latent class models with increasing number of classes. Based on the BIC, a four-class model yielded the most parsimonious best-fitting model. The response patterns for each class are graphically depicted in Figure 2. 30% of the children consistently chose the *growth* response, implying a growth-only mental model of prenatal development. 42% of the children chose the *different* responses consistently for most items, implying a mental model that includes important changes of shape. 6% of the children chose the *outgrowth* response option for most items, although less so for the hand and the foot, implying a mental model that allows for outgrowth of body parts. Although the responses of only a small subgroup of children was best described by this class, it does contribute to the most parsimonious best-fitting description of the data. Finally, 22% of the children chose a mixed set of responses, *growth* and *different* response in an unsystematic combination.

A second series of models were fit to the responses of the late shape items. Table 2 (second section) shows the fit statistics. Based on the BIC, we selected the two-class model as the best, most parsimonious model. 51% of the children chose the *no-change* response with high probability for all 3 items (estimated conditional parameters are .99, .99, .98 for the hand, leg ear items, respectively). 49% of the children were consistent in choosing the *growth* response (estimated conditional parameters are .95, .96, .95 for the hand, leg ear items respectively).

TABLE 2. Fit Statistics for Latent Class Models.

Model	# classes	LogL	df	BIC
Early	2 class	-1300.8	25	2745.5
Shape	3 class	-1242.1	38	2703,0
	4 class*	-1201.7	51	2697,0
	5 class	-1184.9	61	2738.1
Late	1 class	-704.8	8	1455.6
Shape	2 class	-357.6	17	813.1
	3 class	-343.9	26	837.4
Function	1 class	-735.4	12	1539.9
6 items	2 class	-648.2	25	1440.2
	3 class*	-609.0	38	1436.7
	4 class	-592.4	51	1478.4
Function	1 class*	-265.2	8	576.5
4 items	2 class	-247.6	17	593.0

Note. *Indicates the most parsimonious, best fitting model; LogL, Loglikelihood; df, degrees of freedom; BIC, Bayesian Information Criterion; no restrictions were added. N=316.

Mental models of bodily functions

To examine whether children showed mental models about bodily functions, latent class analysis was performed on the responses of the six function items: 3 EF items and 3 LF items. Table 2 (third section) shows the fit statistics of the explorative latent-class models with increasing number of classes. Based on the BIC the three-class model was selected as the best-fitting, most parsimonious model. The response patterns for each class are graphically depicted in Figure 2, lower panels. 23% of the children chose a *no-change* response for the breathing items, combined with a *different* response for the other items. 11% of the children chose the *absence* response for the breathing items, again combined with a *different* response for the other items. 66% of the children chose a *different* response for all items. This implies that children systematically chose responses other than *different* only for the breathing items. We checked this interpretation by modeling the function items without the breathing items. This yielded a 1-class model as the best, most parsimonious model (see Table 2, lower section).

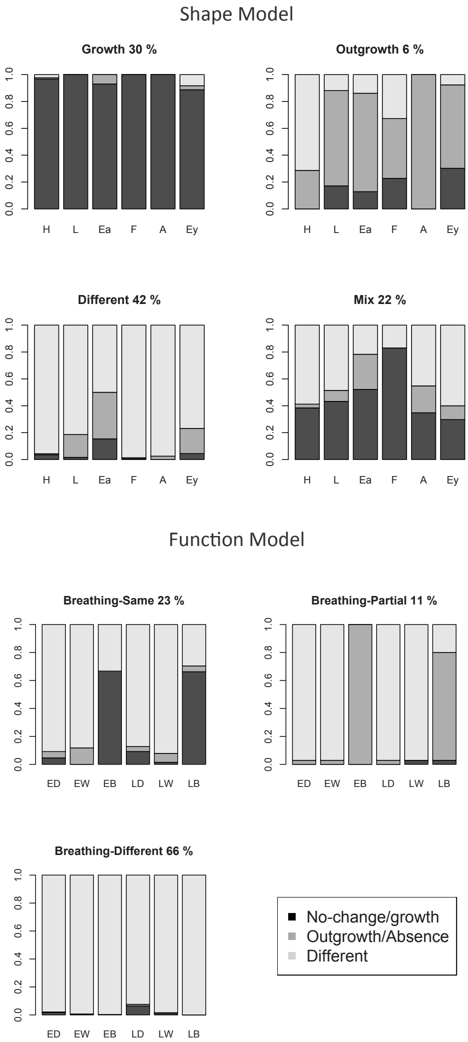


FIGURE 2. Graphical depiction of the latent class models of early shape items (upper part) and early and late function items (lower part). Each figure shows the conditional response probabilities for all items in one of the latent classes. Black refers to the conditional probability of the *no-change/growth* response, dark grey refers to the *outgrowth* response, and light grey refers to the *different* response. For the shape model the respective items are Hand (H), Leg (L), Ear (Ea), Foot (F), Arm (A), and Eye (Ey). For the function model the respective items are Early Drinking (ED), Early Warm (EW), Early Breathing (EB), Late Drinking (LD), Late Warm (LW), Late Breathing (LB).

Grade-related differences

For both models, the shape and the function model, each individual was assigned to a latent class based on the posterior probabilities of their responses, given the model. For the shape model, the Wald criterion demonstrated that grade made a significant contribution to predicting class membership (z=4.45, $p < .001$). Fig. 3 shows the relationship between grade and class membership: The *growth* class is most frequent for grade 1, whereas the *different* class is most frequent for grades 3 to 6. The *incomplete* and *mix* classes appeared in all age groups. For the function model, however, the Wald criterion did not clearly demonstrate a relation between grade and class membership (z=-1.89, p=0.06). In spite of the low variation in responses, grade is a significant predictor of the sum score of the four function items about drinking and keeping warm ($t(315)$=5.41, $p < .001$; adjusted R^2=.08). Accuracy of responding increased with grade.

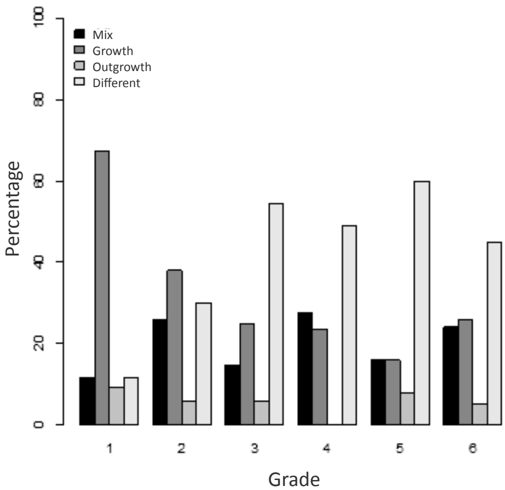


FIGURE 3. Graphical depiction of the relation between grade and mental model based on the early-shape items. For each grade the distribution of the children over mental models is shown (in percentages).

Mental models 'on the spot'?

To examine whether a preceding generative drawing task increased the coherence of the responses for the early-shape (ES) items, we compared the two experimental groups: children who first completed the drawing task then the questionnaire and children for whom this sequence was reversed.

Coherence was defined as the proportion of responses for ES items that were equivalent (i.e., either *no-change/growth* or *out-growth*, or *different*; M=.81, SD=.18). A t-test comparing the two groups showed a small, but significant difference for coherence ($t(302)$=2.49, p=.01). Children who completed the drawing task first had a significantly higher coherence score (M=.84, SD=.17) than children who completed the questionnaire first (M=.79, SD=.19).

The same effect was present when analyzing whether the sequence of tasks affected the model children adhered to (the Wald statistic in a logistic regression: z=2.41, p=.02). If the drawing was made before the questionnaire was filled out, children were assigned to a class with greater coherency. That is, a growth model was more frequent (37% versus 25%) and a mixed model was less frequent (16% versus 24%).

Drawing task & comparison of methodologies

We classified the drawings into three categories. The *no-change* and the *growth* category were merged, because in the analysis of ES items of the questionnaire the four response options were analogously reduced to three response options. Moreover, scoring *growth* in the sometimes miniature drawings was difficult. Inter-rater reliability of classifications based on the three categories (1- no change or growth, 2- outgrowth, and 3- different) of the drawing task ranged from 92% to 96%. Cohen´s Kappa ranged from .84 to .89.

There was no significant relation between grade and the classification of the drawings. To see whether agreement existed between the different tasks, we compared the classification of the drawings with the classification based on the questionnaire. We did not find a significant relation.

Conclusion

Latent class analysis of forced-choice questionnaire responses, revealed mental models for children's ideas about early fetal shape: Three subgroups of children showed coherent responses, which can be interpreted as mental models. One model, the *growth* model, which was most frequently found in the early grades, assumes that the early fetus is only smaller than a newborn baby. The second model, the *outgrowth* model, which was present in all age groups in small percentages, assumes that body parts are only partially present during early fetal development. The third model, the *different* model, assumes that the early fetus looks very different from a newborn baby. This model was dominant in the highest grades. A fourth group of children gave mixed responses, which were not correlated. In contrast, the LCA of early and late-function items did not reveal multiple coherent naïve ideas. Most children responded that the functions are different from post-natal functions in both early and late fetal development. Only in concern to breathing did we find multiple ideas, which implies that there were no mental models about bodily functions in general, other than the accurate idea that functions are different during fetal development. Moreover, the accuracy of responding increased somewhat with grade. This developmental pattern is consistent with the theoretical stance of fragmented knowledge.

The hypothesis that the coherence of ideas increases by doing a generative task, such as drawing, was confirmed for the shape items. Straatemeier et al. (2008) did not find an effect of a generative task on the coherency of questionnaire responses. However, they did not find mental models as we do for the early shape items. Coherence of responses for the questionnaire was larger for the children who first drew two stages of fetal development. 8% more children yielded a mental model in that case, mostly a *growth* model. We did not find a clear correspondence between the models resulting from the different measurement methods. The result of drawing method also did not relate to age suggesting that this method is less useful in detecting mental models.

EXPERIMENT 2

In Experiment 2 we compare the forced-choice questionnaire with structured interviews, which is another popular method in mental model studies. In contrast to forced-choice questions, open-ended interviews allow for generating responses that were not anticipated during the design of the task. The main research question is whether the mental models about early shape that we found with the questionnaire, can be replicated when using the interview. The same procedure was used as in Experiment 1 (questionnaire and drawings), but interviews were added to the design. Asking open-ended questions is argued to have a generative effect on mental models (Vosniadou et al., 2004). Hence, Experiment 2 also examines the effect of open-ended questions on the questionnaire.

METHOD

Participants

Thirty-seven children were tested (age: *M*=8.84, *SD*=1.75) from a primary school providing regular education. An active consent procedure was used with parents being required to sign and return a form if they agreed to let their child participate. The sample consisted of children between 6 and 12 years old: participants included 6 children from Grade 1, age 6-7 (*M*=6.33, *SD*=.52); 6 children from Grade 2, age 7-8 (*M*=7.33, *SD*=.60); 8 children from Grade 3, age 8-9 (*M*=8.50, *SD*=.63); 8 children from Grade 4, age 9-10 (*M*=9.38, *SD*=.67); 7 children from Grade 5, age 10-11 (*M*=11.00, *SD*=.51); and 3 children from Grade 6, age 11-12 (*M*=11.33, *SD*=.58).

Materials

In Experiment 2, the same materials were used as in Experiment 1: the forced-choice questionnaire and the drawing task on prenatal development. Additionally an interview was conducted individually. The structure of the interview was similar to the structure of the questionnaire, with children being asked questions about the early and late stages of prenatal development. Six of the questions concerned shape (3 ES and 3LS items) and four concerned bodily functions of the fetus (2 EF and 2 LF items). In addition, two questions about the fetus as a whole were added, so children were asked 12 questions. Questions were formulated to be open-ended, follow-up questions were defined for anticipated responses. A maximum of two follow-up questions was used to avoid repeated questioning and to enhance rules of normal communication. A doll was used as a model to represent the newborn, to facilitate the child's understanding of the question. A translated example of an early stage item on shape is: '*Look* [interviewer points at the hand of the doll], *this is what the hand of a baby looks like, when it has just been born. Does the baby have hands yet when the baby still has to stay in the belly of the mother for a long, long time?'* A confirming answer would lead to the question: *'What does the hand look like when the baby still has to stay in the belly of the mother for a long, long time?'* If the child answered that the fetus does not have hands yet, the interviewer would ask: *'So what does the baby have?'* The interview took about ten minutes.

All interviews were audio-visually recorded and scored afterwards. All questions were scored separately and independently. It was assumed that children's responses would be consistent with one of 4 expected models. A list of responses was specified for each question, representing the core characteristics of the models. Next, shape responses children gave were scored as 1) *no-change* response, 2) *growth* response, 3) *outgrowth* response, and 4) *different* response. For the function items we scored again in three categories: 1) *no-change* response, 2) *absence* response, and 3) *different* response.

Design and procedure

We tested children in Experiment 2 with two sequences of the tasks. There was a fixed sequence for the tests administered in a classroom setting; all children first completed the drawing of the baby and then filled out the questionnaire. Children were either interviewed individually before the classroom tests (18 participants) or after the classroom tests (20 participants). The procedure of the classroom tests was the same as in Experiment 1.

RESULTS

General

Children's interview responses were scored in three categories for each item. The inter-rater reliability of these ten items was 84% agreement on average (63% to 100%), which resulted in a Kappa of .73 (range .54 to 1.0). The internal consistency of the Shape and Function items in the interview with the response options taken as an ordinal variable (*no-change*, *growth*, *incomplete*, *different*) was .71, as expressed by Cronbach's Alpha. Nonparametric (Spearman's rho) correlations between items varied between .35 and .61 for the early-shape (ES) items, was .51 for the early-function (EF) items, were between .44 and .53 for the late-shape (LS) items, and was .48 for the late-function (LF) items. Table 3 shows the distributions of responses.

TABLE 3. Distributions of responses for the interview items in Experiment 2.

Item	No-change	Growth	Outgrowth	Different
Early Shape				
Hand	0	47.4	31.6	21.1
Leg	2.6	50.0	34.2	13.2
Ear	2.6	47.4	28.9	21.1
Early Function				
Drinking	19.4	0	5.6	75.0
Breathing	51.4	0	10.8	37.8
Late Shape				
Hand	39.5	50.5	0	0
Leg	32.4	62.2	5.4	0
Ear	28.9	68.4	0	2.6
Late Function				
Drinking	19.4	0	8.3	72.2
Breathing	44.4	2.8	13.9	38.9

Note*.* Frequencies are given as percentages of children (N=38). From the 380 data points, 8 were missing.

Mental models of bodily shape and function

Due to the small number of participants in this study it was not possible to apply LCA to the scored responses of the interview. To examine whether the interview revealed mental models, we analyzed the coherency of the scored responses. The shape model based on the questionnaire suggests that we could expect three mental models for the early-shape items: the *growth* model, the *outgrowth* model, and the *different* model. Hence, to examine coherency we counted the number of responses from the same mental model (*no-change/growth*, *outgrowth*, *different*). For the interview, 50% of the children responded to all three early-shape items in the same mental model, compared to 72% for the corresponding items in the questionnaire. For all four early and late-function items, for which we did not find mental models in the questionnaire, the percentage of children with coherent responses was 44% of the children for the interview versus 72% for the corresponding questions in the questionnaire. In summary, responses in the interview were not very coherent for both the shape and function items. Hence, the function items were difficult to classify into mental models.

Comparing methodologies

To examine whether children gave evidence for adhering to the same mental model based on the interview and the questionnaire, we classified children's questionnaire responses. We based this classification on the Experiment 1 early-shape latent class model, by means of the posterior probabilities of the responses given the four classes of this model.[2] There was no significant effect of grade on the model classification of questionnaire responses.

Based on the three ES items of the interview children were also classified to one of the four classes. We used the Experiment 1 early-shape latent class model to classify responses instead of pattern matching to include the possibility that a response pattern was assigned to the mixed class. There was no significant effect of grade on the model classification of interview scores.

Comparing the classifications of children based on the questionnaire and based on the interview, a significant dependency was found ($\chi^2(2)=20.68$, bootstrapped $p < .01$; see Table 4). However, the dependency between the two classifications resulted from the large growth class only. For the other cells in the cross-table we do not see overlap.

As in experiment 1, children's drawings were classified as agreeing with one of three models. The classification based on the drawings was not related to the questionnaire classification or to the interview classification.

TABLE 4. Cross-table of the classification of children's questionnaire and interview responses.

Questionnaire	Interview				
	Growth	Outgrowth	Different	Mix	N
Growth	14	2	1	5	22 (61%)
Outgrowth	0	0	0	0	0 (0%)
Different	0	4	1	0	5 (14%)
Mix	1	4	3	1	9 (25%)
N	15 (42%)	10 (28%)	5 (14%)	6 (17%)	36

Mental models 'on the spot'?

For each child, the coherence of the questionnaire responses was expressed as the proportion of early-shape responses that were from the same category (M=.83, SD=.19). There was no effect of the sequence of tasks (interview before or after the drawing and questionnaire) on the coherence score. There was also no effect of sequence of tasks on the mental models (based on the questionnaire) to which the children adhered.

Conclusion

Interview responses were less coherent than questionnaire responses. Considering the difficulty found in categorizing the response patterns to the expected models, it is concluded that the responses in the interview did not support the existence of the expected mental models. Age related differences were not found in the questionnaire classifications and also not in the interview classifications, due to the high percentage of children in the *growth* class. A significant dependency was found between the classification based on the interview and the classification based on the questionnaire. The correspondence, however, was only due to the large *growth* class. No dependency was found between the other classifications.

Vosniadou et al. (2004) stated that asking open-ended questions would have a generative effect on the construction of mental models, just as drawings do. In Experiment 2 the questionnaire was always preceded by the drawings, which was shown to have a generative effect on the questionnaire in Experiment 1. Coherency of the early-shape items of the questionnaire in Experiment 2 (M=.83, SD=.19) was indeed comparable to the coherency of these items in Experiment 1 for the case when it was preceded by drawing (M=.84, SD=.17). Administering the interview did not have an additional generative effect on the coherence of children's questionnaire responses for the early-shape items.

DISCUSSION

Mental models of prenatal development

One of the main questions in the literature about children's naïve ideas about the world is whether children form coherent ideas about a specific topic, that is, mental models. We studied this issue for the case of mental models about prenatal development. The position that we took in this study is a methodological one: the measurement method needs to allow for testing the coherence between responses in order to decide on this issue. In this respect, the forced-choice questionnaire is the most appropriate, especially when combined with a statistical analysis technique that tests the existence of coherent response patterns in the data, such as latent-class analysis. Experiment 1 shows that children's naïve ideas about changes in the shape of a fetus' body were coherent for most children (78%). These children believed for all body parts that they only *grow* between early and late stage of prenatal development (30%), that most body parts *grow out* of the body (6%), or that the shape of all body parts undergo *important changes* (42%). Note that the outgrowth model was not found in Experiment 2, but given that this group was small in Experiment 1, this is not completely unexpected. 22% of the children showed an unsystematic mix of two types of responses: body parts grow or undergo important changes. The reasons for this incoherence can be various: some pictures of the test seemed implausible to them; their ideas were instable; or they guessed between the two most plausible response options. The finding of mental models of prenatal body shapes is consistent with what was found by Rosengren et al. (1991) about the natural transformations that children accept. According to Rosengren et al., young children know that animals get larger with age, but only older children and adults accept rather dramatic changes in shape.

For bodily functions the results were different. Most children believed that fetal bodily functions were different from the functions of a newborn. For breathing only, some children (23%) believed that this function is the same in the fetus and the newborn and some children (11%) believed that the fetus does not need something from the air at all. Hence, children did not have general ideas about changes of bodily functions during prenatal development other than that they are different from the functions of newborns. Although in reality, bodily functions are different in a fetus compared to a newborn, the *different* responses do not necessarily indicate that most children had a scientifically acceptable idea of this difference. In the questionnaire we did not test children's ideas about bodily functions in more detail because a coherent detailed idea about multiple functions is not very likely (or thinkable). Hence, we conclude that children do not have mental models about change of bodily functions (drinking, breathing, keeping warm) other than that these functions are different in the fetus.

Whether the interviews gave us more insight in children's naïve ideas is questionable. Firstly, the coherence of the responses in the interviews was considerably less than for the questionnaires. Secondly, the relation that was found between the classifications based on the questionnaire and on the interview was mainly due to a large number of children adhering to a *growth* model. Hence, the reliability of the interview responses as reflections of children's mental models is questionable. It is very likely that children hardly thought about bodily functions of a fetus before (Zoldosova & Prokop, 2007; Morris et al., 2000) and that their sometimes-detailed responses were constructed on the spot. It almost never occurred that children responded with "I don't know" (a maximum of 2% of the responses were omitted), which is remarkable given that young children appear to know very little details about human insides (Morris et al., 2000). If children were constructing the responses on the spot instead of basing them on their mental model, the interview did not test children's knowledge about prenatal development, but children's abilities to construct narratives from fragmented knowledge (DiSessa, 2008).

Although we conclude that interviews and drawings are not very reliable methods for testing the presence of mental models, these formative assessment methodologies might be interesting from an educational perspective. In Experiment 1 we see that children's responses on the questionnaire were more coherent after drawing different phases of prenatal development. Hence, these methods seem to stimulate children's thinking about prenatal development, creating good opportunities for teachers to present new knowledge about the topic (Schwartz & Sadler, 2007).

Representational flexibility

The mental models that we found for early shape development are consistent with Karmiloff-Smith's (1992) ideas about representational flexibility. Representational flexibility is assumed to occur within specific topics (i.e., micro-domains) throughout development, but the overall sequence of introduced modifications is believed to be domain general. Indeed, the grade-related results on the questionnaires are in agreement with this specific sequence of phases in representational flexibility. Karmiloff-Smith's original ideas about representational redescription were based on drawings of human beings. Remarkably, our results based on children's drawings did not reveal a significant relation between representational flexibility and grade.

Whether representational flexibility is dependent on procedural constraints or is more dependent on conceptual rigidity is a debated issue in literature with both positive (Picard & Vinter, 2007) and negative evidence (Spensley & Taylor, 1999, Barlow et al., 2003, Hollis & Low, 2005). According to Karmiloff-Smith (1990) the procedural constraint in the first phase of representational redescription is that young children's internal representations are dependent on a sequentially fixed list. Representational flexibility increases in subsequent phases because the constraints are relaxed. Results of the study by Picard and Vinter (2007) are in line with Karmiloff-Smith's theory. They find that rigidity in routine development constitutes a sequential constraint that limits inter-representational change. When this sequential constraint is relaxed, it is likely to be one of the factors leading to inter-representational flexibility. In the present study we found evidence that the representational redescription was more dependent on conceptual rigidity, and we could not reveal it was dependent on procedural constraints, i.e. drawing skills.

Methodological issues

The reason for the use of a forced-choice questionnaire was that this methodology makes it possible to measure coherence in responses, which is a criterion for mental models. A point of criticism of this methodology could be that due to the limited response options children are more likely to respond according to a specific model, increasing the chance of the researcher finding a coherent model. However, it appears that the use of forced-choice questionnaires in investigating children's mental models does not necessarily lead to finding coherent models (Straatemeier et al., 2008). Moreover, unlike the shape related items, responses on the function items did not reveal multiple coherent models. A second point of criticism could be that the questionnaire was composed of adapted photographs. This could have lead to a response bias; children might have just picked the most unusual picture. The lack of variation in the responses to the late-shape items provides evidence against the existence of such a bias. Here, children consistently chose the *no-change* and *growth* response, which photographs look like the newborn.

This study shows that more specific conclusions can be drawn using LCA as compared to pattern matching. The use of LCA made it possible to detect a subgroup of children with mixed, incoherent responses. Moreover, LCA provides a measure of how many mental models are present in the data. For the early-shape items only a few children showed a more or less coherent pattern of outgrowth responses. Nevertheless, identifying this subclass of children contributes to the best describing, most parsimonious model.

Conclusion

Taking into account the scarcity of previous experimental work on mental models of prenatal development, our study adds substantial insights on the topic. Children have a mental model of prenatal shape development, which is grade related. Knowledge about bodily functions appears to be more fragmented. That is, dependent on the specific topic, children reason from separate facts or from general principles. About bodily functions, children reason form separate facts (cf., Zoldosova & Prokop, 2007). About bodily shape, children reason from general principles, i.e. transformations that are acceptable for shapes or animals (Rosengren et al., 1991). More systematic studies on the relation between available sources of knowledge and the type of knowledge that children construct from these sources could provide more general insight in the development of knowledge.

The findings of this study are of interest to both developmental psychology and education. Teachers in primary schools will derive from this study a better understanding of the development of children's mental models of prenatal development, in particular and of knowledge development, in general. Indeed, it can be difficult for children to acquire scientific theories because these theories are often inconsistent with their naïve ideas (Karmiloff-Smith & Inhelder, 1975; Samarapungavan et al., 1996). Restructuring a mental model is even more complex considering the fact that adults, such as teachers, have different models than children and are usually unaware of children's initial models. When teachers take into account children's initial assumptions, they can guide children's restructuring of mental models in an optimal manner (Schwartz & Sadler, 2007).

Our study contributes to a deeper understanding of the effects of different methodologies being used in investigating mental models. A statistically valid check for coherence as opposed to the assumption of coherence should be part of each methodology.

ACKNOWLEDGEMENTS

This work was funded by the Dutch Organization for Scientific Research (NWO; VIDI and ASPASIA grants from M.E.J. Raijmakers) and the Dutch Curious Minds program, which is supported by the Ministry of Education, Culture and Science and Platform Bèta Techniek. The authors thank the teachers and pupils who participated in this research. Also, we thank Jolande Klijn for assisting in this study. We thank A. Karmiloff-Smith for her valuable comments on an earlier version of this paper.

NOTES

1. The workshop was LOOL, which stands for 'leren onderzoekend en ontwerpend leren', Dutch for 'enquiry based learning and learning by design'. The project was lead by Marja van Graft and executed by Rooske Franse in 2005 and 2006.

2. The classification of individuals depends both on conditional and unconditional parameters of the model. We assume that the estimates of conditional parameters are the same for Experiment 1 and Experiment 2 children. However, the unconditional parameters depend on the distribution of age within the group, which we do not assume to be equal across experiments. Therefore, we estimated the latent class model again with grade added as a covariate on the posterior probabilities. Grade was significantly related to the model classification of the questionnaire (likelihood difference between models is: $\chi^2(2)=7.07$, $p < .03$). Almost the same classification of Experiment 2 children was found on basis of this model. Only one subject was classified as having a *growth* model instead of *mixed* ideas. For reasons of simplicity, we worked with the simple model that was explained in the text.