



# How to reliably diagnose children's concepts in learning science? Using the water cycle as an example

Andreas Louis Imhof<sup>1\*</sup> , Markus Kübler<sup>2</sup> 

<sup>1</sup>University of Teacher Education of Grison, Chur, SWITZERLAND

<sup>2</sup>University of Teacher Education of Schaffhausen, Schaffhausen, SWITZERLAND

\*Corresponding Author: [andreas.imhof@phgr.ch](mailto:andreas.imhof@phgr.ch)

**Citation:** Imhof, A. L., & Kübler, M. (2025). How to reliably diagnose children's concepts in learning science? Using the water cycle as an example. *Interdisciplinary Journal of Environmental and Science Education*, 21(2), e2509. <https://doi.org/10.29333/ijese/15960>

## ARTICLE INFO

Received: 27 Nov. 2024

Accepted: 20 Jan. 2025

## ABSTRACT

The text addresses the question of how methodology in research on children's conceptions about the world possibly affects results. This methodological question has received little attention yet in educational and psychological literature. As part of a preliminary study focusing on the effectiveness of internally differentiated factual texts using the water cycle as an example, the authors examine(d) children's conceptions of it. The data set includes 121 pre- and post-tests of nine- to ten-year-old children. Results show that free recall tasks tend to underestimate children's performance, whereas cued recognition tasks tend to overestimate children's performance. The findings demonstrate that it is worthwhile and important—in terms of reliability and validity of the data—to check the survey methods for their potential biases on the results and to plan for method triangulation from the beginning when surveying preconcepts.

**Keywords:** conceptual change, water cycle, free recall, cued recognition, children

## INTRODUCTION

Information about children's concepts, their knowledge structures about science topics and phenomena of the world, as well as children's reading ability are of great importance for effective learning in natural sciences (Möller, 2018; Relyea & Hwang, 2024) and related subject areas (Karacaoğlu & Kasap, 2023; Koyuncu & Firat, 2020). In educational psychology and subject didactics, children's ideas about the world, or children's concepts, are a significant object of research. From a research perspective, however, the challenge is to measure children's concepts reliably and, above all, validly, independently of their verbalization skills. These psychometric properties, which are central to diagnostics, are given little attention to in many studies of children's concepts, and the literature on subject didactics seldom reflects on the influence that the measuring instruments themselves can have on the results. Depending on how the measurement is implemented, different memory processes are activated in the subjects and this can lead to different results. For the investigation of children's concepts in science education, measurement procedures are common that focus either on information recall or on information recognition.

A precise diagnostics of children's (prior) knowledge or different children's concepts is not only important for research, but also for instruction, especially if internal

differentiation is part of lesson planning. In the discussion about the internal differentiation of instruction, however, prior knowledge and its diagnostic hardly play any role as prerequisites for effective learning in science instruction (Möller, 2018). The following text aims at answering the question of how a reliable and valid diagnostic on children's preconcepts can be achieved, what methodological approaches are needed to support the development of children's knowledge structures in certain factual topics through instruction and how this can be measured.

Certain children know a lot about certain topics, while others know very little and furthermore are even unable to connect the concept/phenomenon (e.g., the water cycle) to their prior knowledge. Research shows that reading skills have a significant influence on how efficiently a child can use factual text to build knowledge in natural science. The correlation between knowledge and reading ability, using history knowledge as an example, is .56\*\* (Kölbl et al., 2006). Of course, this correlation could be bidirectional, as research show that children with domain specific knowledge could better decode information in corresponding texts and thus integrate new knowledge better in their semantic networks (Smith et al., 2021). On the other hand, research shows that skilled readers show greater learning gains in different domains than less skilled readers, beginning with kindergarten (Relyea & Hwang, 2024).

Furthermore, some children are experienced readers, others still decode individual letters. Nevertheless, the lessons are the same for all the children; there is hardly any differentiation with respect to either the previous knowledge of the children or their reading competence. There are also hardly any teaching materials that take these differences into account. Thus, some children are bored (great prior knowledge; experienced readers) and others are acutely overtaxed (hardly any prior knowledge; inexperienced readers).

The water cycle is an example of a subject for which research on children's concepts has existed for some time. The topic is of interest in subject didactics from two perspectives: firstly, the water cycle is a content that is frequently taught by many teachers in the second to fourth elementary school. Consequently, the subject matter is relevant to the teaching of science in elementary school. Secondly, the water cycle contains both observable (e.g., rain and rivers) and non-observable phenomena (e.g., evaporation, condensation, and groundwater).

In the case of the latter, this poses significant challenges for the children to develop a correct understanding of the natural sciences: the content cannot be explored perceptually or experientially but only analytically (e.g., through description in the form of a factual text). So, the water cycle is accessible to most children (numerous perceptual elements to which children's prior knowledge can be linked) and yet also contains more complex elements where children can show in-depth knowledge and thus is well suited for a comparative study of survey instruments for diagnosing children's concepts.

The present paper draws on data from a preliminary investigation for a larger field study on the question of how factual texts must be designed to enable children to learn about natural sciences. In this article, we focus on the question of measurement methodology for children's concepts—using the example of the water cycle—and on the question of the structure of children's concepts:

1. How do measurement methods based on information retrieval and recognition agree for the elicitation of children's concepts?
2. What concepts about the water cycle are found in nine- to ten-year-old children (third and fourth grade)?

Further research questions of the preliminary investigation, such as the influence of reading literacy on conceptual learning or the effects of the use of bi-differentiated factual texts will be discussed in more depth in a separate paper (in planning).

The paper begins with a description of the main lines of discussion on the concept change paradigm used. Based on this, common procedures for measuring children's concepts and their epistemological implications as well as their concepts about the water cycle are described. Subsequently, the methodology of the present preliminary experiment is presented. The results are described and discussed.

## THEORY

### Conceptual Change

Before entering school, children already have conceptions about the world. In psychological theory, there are different theoretical assumptions on the structure of these pre-concepts. Vosniadou and Brewer (1992) and Carey (1985) refer to (pre)concepts as coherent, theory-like structures. di Sessa (2008), on the other hand, speaks of only loosely connected, partly isolated knowledge components. In our own research, we were able to find evidence for both forms of children's concepts. Meanwhile, in the discussion about conceptual change, it is agreed that children's conceptions show a high resistance to change. So, very often, incomplete or wrong conceptions (misconceptions) are active in the classroom, which are maintained or even strengthened with arguments. Thus, enrichment rather than concept change takes place in the classroom (Aleknavičiūtė et al., 2023; Möller, 2015, 2018; Pacaci et al., 2024). From a developmental psychological point of view, this circumstance is nowadays justified less by maturation theory than by the fact that children have incomplete knowledge as well as less technical terms and a limited working memory capacity (Reynolds et al., 2022; Sodian, 2008; Ullman et al., 2014). Children construct ideas about the world by using analogies and inferential thinking to supplement their knowledge gaps into theory-like constructions. In doing so, they do frequently generate their own concepts for real world phenomena (for an overview, see Adamina et al., 2018).

### Recall and Recognition in the Measurement of Children's Concepts

Empirical studies of children's concepts often use qualitative procedures: Children are asked about a factual topic or phenomenon directly or through the presentation of an auxiliary stimulus. This approach is usually associated with rather small samples. Studies with larger samples often use questionnaire formats with more or less sophisticated items. Based on the responses to these items, children's concepts are inferred. However, measurement methods differ not only in their conception—some being more qualitatively and others more quantitatively oriented, but also in the cognitive processes they activate: some may require information retrieval ("recall") whilst others are set to trigger information recognition (cf. Anderson & Bower, 1973; Brown, 1976; Kintsch, 1970). The former requires productive performance from the child. Children typically receive no or little help to answer the task. Based on a keyword or an (open) question, they must search for suitable information in their own knowledge networks or long-term memory, compare it with the task and then reproduce it. An example of this is an open request for a child to tell what they know about a certain topic.

In contrast, presenting the children with several answer options (as in a questionnaire) or allocating helping stimuli enables a child to recognize what he or she has once learned in the test situation. Consequently, information stored in memory does not have to be searched for but can directly be compared with the information at hand. This approach can activate "hidden knowledge" that cannot be reproduced in an

**Table 1.** Methodical setting “bi-differentiated factual texts”

Pre-testing		Intervention		Post-testing	
1	2	3	4	5	
Drawing and children’s explanation	SLT and children’s explanation	Reading the text: EG1: Level adapted texts & EG2: Non-adapted text	Interview repeating text and children’s explanation	SLT and children’s explanation	

open prompt. Knowledge is stored in memory in this case but is not accessible to the child.

However, both procedures can potentially distort the measurement of children’s knowledge: On the one hand, helping helps the child to activate existing associations from memory that he or she would otherwise not associate with the phenomenon. On the other hand, from a diagnostic point of view, there is a risk that no prior knowledge, but “current construction” is produced due to the assistance. These current constructions, in contrast to the “deep structures”, are not connected to children’s concepts in their memory (Hartinger & Murmann, 2018). In unsupported, free recall, this is a minor risk, as children cannot rely on external support (provided that implementation objectivity is given). However, with measurement procedures in the retrieval paradigm, there is a different risk of bias: Especially those children still showing less connected knowledge, may not exhibit available knowledge due to the lack of situational accessibility (since no connections/associations are drawn to actually existing prior knowledge). In addition, especially with younger children, it can be assumed that—depending on the chosen form of expression (oral, drawing, writing)—not all children are equally competent in their expression (lack of drawing/writing/telling skills or technical terms, limited working memory or lacking confidence in testing situations). It can therefore be assumed that measurement procedures with a stronger focus on information retrieval have a potential bias because they underestimate children’s knowledge. In contrast, surveys with the possibility of recall are at risk of overestimating or biasing child knowledge.

### State of Research on Children’s Concepts About the Water Cycle

In studies based on the theorem of conceptual change, it is common that children’s conceptions are described by means of concept levels. Of the 30 studies concerning children’s knowledge of water cycle we found, six of them use differently designed concept levels as a measuring instrument. The structure of the levels mostly follows a progression logic in the sense that the children’s concepts are ordered one-dimensionally with respect to their technical correctness. However, we found no studies that organize children’s concepts with respect to their qualitative content. For example, Miner (1992;  $n = 56$ ) has four levels: Level 1: Complete confusion; Level 2: Partial confusion; Level 3: Partial understanding; and Level 4: Complete understanding. Cardak (2009;  $n = 156$ ) uses six levels: Level 1: No drawing; Level 2: Drawing without reference to the topic; Level 3: Drawing with some correct approaches and with misconceptions; Level 4: Drawings with partially correct conceptions; and Level 5: Drawings with understandable and correct conceptions. Heng and Karpudewan (2017;  $n = 53$ ) and Koomson and Owusu-Fordjour (2018;  $n = 86$ ) use(d) similar levels: Level 1: No drawing; Level 2: Nonrepresentational

drawings; Level 3: Drawing with misconceptions; Level 4: Partial drawings; and Level 5: Comprehension representation drawings. Suryanti et al. (2018;  $n = 23$ ) use(d) six levels but did/do not describe them in detail. Ursavas and Genç (2021, p. 244) use content analysis to define four level as function of student’s test results: “insufficient” 0-25 points; “limited” 26-50 points; “sufficient” 51-75 points; “excellent” 76-100 points. They notice that data collection procedure had a major impact on their results, as younger children showed superior achievements in playful test environments using drawings than older children (Ursavas & Genç, 2021, p. 251). Similarly, Aleknavičiūtė et al. (2023) argue that many research studies mainly examine knowledge enrichment, but few focusses knowledge reconstruction (e.g., conceptual change).

As mentioned already, these concept-level models are summatively conceptualized but make little reference to the content of students’ conceptions.

Therefore, a concept-level model is iteratively developed using children’s data. In the first step, different concepts about the water cycle are identified by analyzing children’s drawings. In a second step, these concepts are grouped into levels with respect to commonalities of the sequential flow of the water cycle (e.g., closedness of the cycle). Within these levels, in a third step, different sub-models are identified, which differ mainly in their degree of differentiation (same concepts, but different degrees of differentiation, or “enrichment”). The following concept levels are postulated: Level 0: No knowledge or concepts; Level 1: Unconnected elements of the water cycle; Level 2: Linear concepts and “reservoir model”; and Level 3: Complete cycle concepts.

## METHODOLOGICAL APPROACHES

### Children’s Factual Knowledge About the Water Cycle

Factual knowledge about the water cycle was measured in two ways: based on drawings and structure-laying techniques (Scholl, 2014). In each case, measurement began by asking children to draw a water cycle. Following this, children were asked to explain their drawing in more detail to the investigators (“recall”, information retrieval pre-test). Subsequently, the children were given cards with elements of the water cycle and asked to put these together in an order that made sense to them (“recognition”, pre-test). Children were then assigned to a concept level depending on their utterances to their drawings and their card structures. The preliminary study is designed as experimental study. Children were tested two times, once before (pre-test) and once after (post-test) an intervention with the same testing material (see **Table 1**). The intervention consisted in reading a factual text that is either level-adapted (experimental group 1) or not (experimental group 2). The experiment itself is described in detail in another article (in planning).

After the intervention, the children were first asked to report on what they had learned/read about the water cycle ("recall", information retrieval post-test) and then to make a card structure again ("recognition", recognition post-test). For the present article, the children's utterances recorded during the structure-laying technique and during the narration about the drawing/text were classified into the category system which had been developed inductively beforehand and subsequently were examined inferentially<sup>1</sup>.

From a statistical perspective, however, the classification into a category system poses some challenges. For example, with respect to the category system, two prerequisites must be met in order to analyze the data on an interval scale level: 1. equal distances between the individual concept levels 2. higher concept levels must represent "more correct" or "more scientific" concepts<sup>2</sup>.

We assume that fulfilling these prerequisites is also important from a didactical perspective. In particular for teaching, the question is crucial how much "cognitive effort" must be invested to get from one level to the next, whether this effort is the same for all levels and whether the knowledge progression runs along the postulated sequence or whether certain levels can be omitted. For example, it can be assumed that it is probably easier to get from level 1 to 2 (necessary cognition: connection or sequencing of individual elements) than from level 2 to level 3 (necessary cognition: conceptual understanding of latent phenomena such as condensation and evaporation).

### Development of Concept Levels and Experimental Design

The concept levels were created inductively based on the children's products in several iterations. In the first step, the drawings and the structure-laying techniques of children (pre-testing) were analyzed regarding the arrangement of the cards of the structure-laying technique. Subsequently, the order of the cards was determined using the card "sea" as point of reference. Based on these data, students' concepts were grouped preliminarily. The student concepts found were then ranked according to their scientific correctness. A distinction was made between concept change and mere enrichment.

### Hypotheses

The different survey methods on recall and recognition may lead to different estimates of concept levels among children. In the present research, the following hypotheses are tested:

#### Hypotheses on RQ1

1. Measurement of concept levels on the water cycle via the structure-laying technique (recognition) and the drawing narration (recall) correlate positively (pre- and post-test).

2. The measurement via the structure-laying technique results in a higher mean concept level than the measurement via the drawing/narration.

The first hypothesis relates to the convergent validity of the two types of measurements. If they correlate only slightly, it cannot be assumed that the same underlying construction is measured. However, it is postulated that both methods are appropriate measures of children's knowledge; accordingly, a high positive correlation is expected.

Information retrieval requires more cognitive processes than information recognition, which is why higher conceptual levels are expected to be measured by recognition than by information retrieval.

Preliminary work shows that children's concepts of the water cycle can vary widely in terms of their scientific correctness. However, it is expected that combining the measurement procedures on information retrieval and recognition through an inductive approach facilitates the identification/definition of content-distinct concepts. Furthermore, we also anticipate that this procedure helps to rank these concepts in terms of their scientific correctness. This classification is based on previous work by Hardy et al. (2006) and Vosniadou (2008).

#### Hypothesis on RQ2

3. Student concepts can be grouped by content and assigned in an ascending concept level system (related to scientific correctness).

### Sample

The surveys took place in two settings.

#### Setting 1

Modulated text difficulty levels in six primary school classes are used (121 primary school children). All children were in the third or fourth grade at the time of the study (age: mean [M] = 10.05 years, standard deviation [SD] = 0.64 years). For the present analysis, the data from four classes were analyzed, leaving 62 children in the study sample. In the remaining sample, the average age was 10.0 years (SD = 0.69 years). The sample consisted of 54.8% female and 45.2% male children. The children were tested in the years 2018-2020.

#### Setting 2

The concept level in factual texts was modulated in two 3rd grade primary classes with 37 children age: M = 9.38 years, SD = 0.60 years. This means that after the pre-test, a factual text adapted to the children's level of knowledge was offered. One class with a good knowledge of German and one class with a rather modest knowledge of German were tested. The results of the second setting will be discussed elsewhere (in planning).

<sup>1</sup> Due to limited resources, the ratings are carried out by the authors in an iterative and consensual setting. Reliability testing by calculating interrater statistics is planned in future. For this preliminary study, the ratings of two raters were compared and differences are solved by discussion.

<sup>2</sup> We analysed our data both on ordinal and interval level. Because there are only marginal differences in test results, only standard test results are reported.



**Table 2.** Model of concept levels according to children's concepts

Concept level	Cl 2	Rating	Concept level description
3 Cycle concept	3c	7	Complete cycle of water. The major cycle and the sub-cycles are explicitly mentioned and justified by the children.
	3b	6	Complete cycle of water. The great cycle and at least one sub-cycle are explicitly mentioned and justified by the children's utterances.
	3a	5	Complete cycle of water. The circularity is explicitly mentioned and justified by the children, e.g., with the statement: "and it starts all over again".
2 Linear concept (reservoir model)	2c	4	Linear concept of water cycle. At least one complete, linear reference is constructed graphically or linguistically, such as from rain (or spring) to the sea; or stream-sea.
	2b	3	Linear concept of the water cycle. Few linear references are constructed, or an incomplete course is constructed.
	2a	2	Linear concept of water cycle. A single linear reference is named or is identifiable.
1 Isolated concept	1	1	Isolated and unrelated elements of the water cycle (e.g., fresh water, salt water, earth as a water planet; a water faucet).
0 No concept	0	0	Drawing without technical reference to the water cycle (e.g., water conservation; drain vortex).

## RESULTS

### Children's Concepts of the Water Cycle

Based on our previous investigations, we made use of 121 children's drawings from six 3<sup>rd</sup> grades of elementary classes in Switzerland (setting 1). Test criteria focused on the factual correctness and completeness of the children's drawings and their statements (complete or incorrect). For concept level construction, we used complexity (simple or complex drawings/card orders), the coherence of the mental models (loose or connected elements), the closedness of the circuit (closed or linear) as well as the process character of the mental models (static or dynamic).

- **Level 0:** The drawing content has no or little relation to the water cycle (no concept)<sup>3</sup>.
- **Level 1:** The drawing contains correct, but only isolated and disconnected elements of the water cycle (isolated concept).
- **Level 2:** The drawing depicts linearly connected elements of the water cycle (reservoir model: water runs from the source to the sea; linear concept).
- **Level 3:** The drawing depicts a cycle of water (circulatory concept).

Besides the four basic levels, we made additional differentiations within the levels. Therefore, levels two and three were divided into three internal levels, which elaborate on completeness, complexity and coherence (see **Table 2**).

We can call this differentiation within levels two and three "enrichment", i.e., a progression within the same student concept. So, there is no actual concept change, rather, children keep their concepts and enrich them additively with more information (Carey, 1991).

For each concept level, a child's drawing is attached as an example for illustration.

#### **Level 0. No concept: Drawing with no technical reference to the water cycle**

The girl (see **Figure 1**) explained that she drew a water vortex. In doing so, she described that she imagined how the "water runs in circles" when it disappears into a sinkhole or a



**Figure 1.** Girl: 9.3 years old and 3<sup>rd</sup> grade (Source: Authors own collection, Chur/Schaffhausen)

drain. The girl is not German-speaking by origin and has broken down the compound noun (German: *Wasser-Kreis-Lauf*; English: *water cycle*) and linked it to a phenomenon she knows. In this respect, we encounter a "current-construction": The child's ignorance of the concept makes her establishing a link with something she already knows and as a result of this to generate meaningful (but technically wrong) own ideas.

#### **Level 1. Isolated concept: Isolated and unconnected elements of the water cycle**

The boy (see **Figure 2**) draws various technically correct elements related to water and its cycle: Earth as a water planet, groundwater, freshwater, and saltwater; glaciers and rain that generates freshwater. Overall, his knowledge is isolated and disconnected. How the individual elements relate to each other remains unexplained. His knowledge is static; processes are not represented.

#### **Level 2. Linear concept of the water cycle**

The girl (see **Figure 3**) draws a complete and technically correct pathway of water from the source to the sea. She graphically visualizes the increasing width of flowing water to its river mouth. She writes the individual elements in the drawing for clarification: Source, brook, pond, stream, lake,

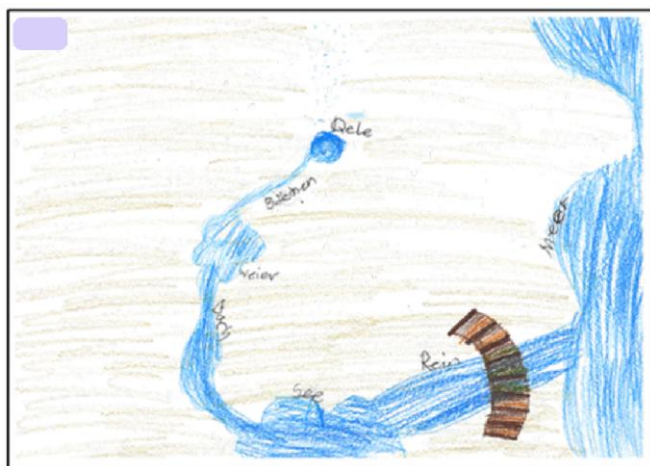
<sup>3</sup> Different student concepts or misconceptions, that are not connected with the topic, are identified. The ranging of different misconceptions is not within the scope of this investigation. For this study, they all are rated as "level 0".



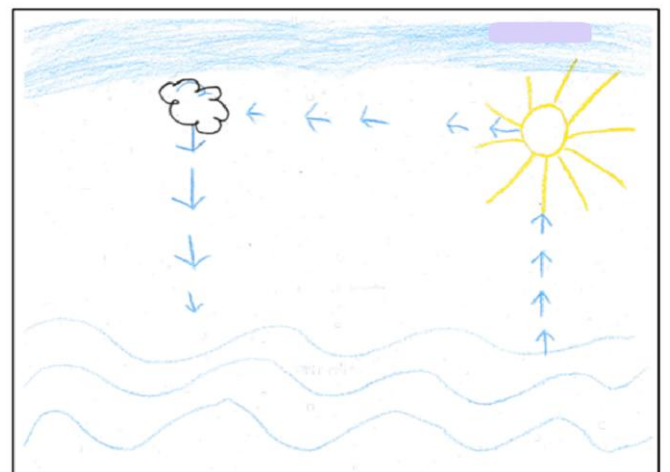
**Figure 2.** Boy: 9.2 years old & 3<sup>rd</sup> grade (Source: Authors own collection, Chur/Schaffhausen)



**Figure 4.** Girl: 10.0 years old & 3<sup>rd</sup> grade (Source: Authors own collection, Chur/Schaffhausen)



**Figure 3.** Girl: 9.8 years old & 3<sup>rd</sup> grade (Source: Authors own collection, Chur/Schaffhausen)



**Figure 5.** Boy: 10.5 years old & 3<sup>rd</sup> grade (Source: Authors own collection, Chur/Schaffhausen)

the stream “Rhine” and sea. The decisive step of evaporation, the returning transport of the moisture and the rain, which becomes the source, are however completely missing. We call this (incomplete) concept “reservoir model”, a subcategory of the linear concept, because the path of water has a beginning and an end. The child had no answer to the question whether the sea would not eventually overflow.

### Level 3. Circulation concept: Complete circulation of water

In level 3, a child (see **Figure 4**) draws the complete, or technically correct, cycle of water. Complete in this case means that the representation showed that the idea of the cycle, i.e., the endless repetition of the path of water was clearly represented graphically. However, complete also means that the essential and indispensable elements of the water cycle were identifiable and named by children, for example with the key phrase: “and then everything starts all over again.” The drawing also contains process elements (arrows) and tries to show parts of the water cycle that are not visible (evaporation).

Variants are also possible: A minimalist representation is present in the drawing below (see **Figure 5**): the boy shows a cycle in its reduced variant. However, whether he was able to fully conceptualize the idea of the cycle only got apparent by

help of his utterances and in particular, when he formulated the anchor sentence “starts over again”.

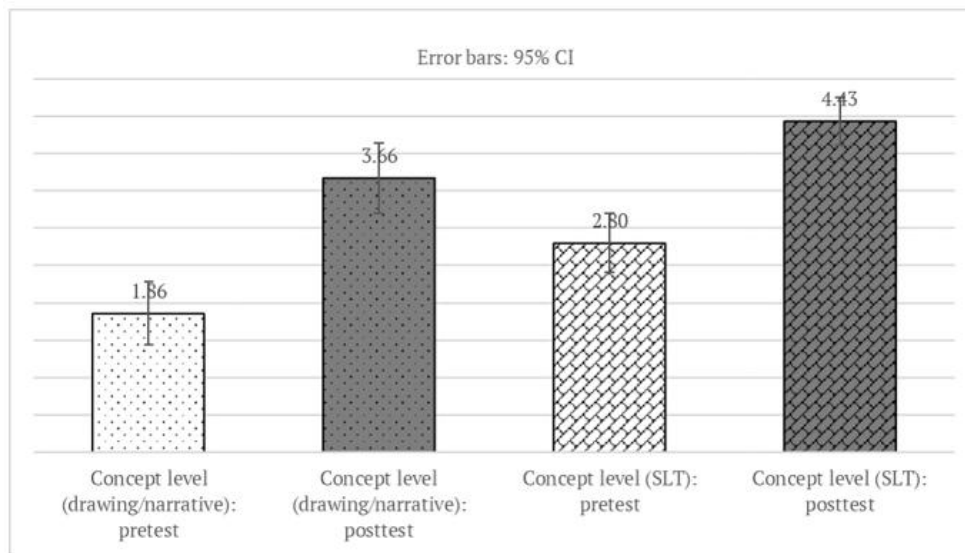
In our opinion, the iterative procedure in the development of the concept levels has proven successful. The tests after the development of the concept levels showed that the rating of the drawings or the classification of the drawings into the four concept levels was distinct and relatively easy to make. It was also found that different raters generated the same results independently. Only children who had problems with linguistic expression caused some difficulties for the raters; this was due to the lack of linguistic material during testing (children do not talk much).

### Test Congruence, Concept Levels, and Intervention Effects

In the pre-test data, significant correlations are found between the measurements related to drawing and those related to the structure-laying technique ( $r = .755^{**}$ ). Furthermore, significant correlations are found for the post-test measurements between these two measures:  $r = .533^{**}$ . In general, the correlations between the individual measures at the same test time are very large (strong effect according to Cohen), indicating good convergent validity.

**Table 3.** Classification of concept levels by drawing/narration at pre- and post-test (the number of children is indicated in the cells)

Concept level	Drawing/recall pre-test	Structure-laying technique:	
		pre-test	post-test
0	11	2	0
1	21	9	6
2a	15	19	7
2b	6	14	13
2c	1	8	4
3a	6	4	12
3b	1	5	8
3c	0	0	0

**Figure 6.** Mean concept level for SLT and drawing/narration: Pre- and post-test (Source: Authors' own elaboration)

The results in **Table 3** show that at pre-test, most ( $n = 32$ ) children's drawings/narrations (information retrieval) are classified into level 1 (drawing unrelated elements) or Level 0 (have no concept).

21 subjects are classified at level 2 (linear concepts, 15 level 2a, respectively six level 2b). In contrast, based on the structure-laying technique (SLT; information recognition), most subjects are classified at level 2 (33 subjects, 19 at level 2a and 14 at level 2b). The mean of concept levels measured by SLT is significantly higher than the mean measured by drawing/narration ( $t [60] = -7.403, p < 0.001, r = .755; M_{\text{drawing}} = 1.79, M_{\text{SLT}} = 2.80$ ). Only a few or no children reach higher concept levels with both types of measurement (level 3c, see **Table 2**).

At post-test (after reading the text), children are rated significantly higher compared to pre-test, both for SLT ( $t [60] = -9.613, p < 0.001, r = .563; M_{\text{pre-test}} = 2.80, M_{\text{post-test}} = 4.43$ ) and drawing/narrating ( $t [49] = -6.604, p < 0.001, r = .247; M_{\text{pre-test}} = 1.86, M_{\text{post-test}} = 3.66$ ). The difference between the test remains stable, with SLT indicating higher concept levels than drawing/narration ( $t [49] = -3.812, p < 0.001, r = .533; M_{\text{drawing}} = 3.66, M_{\text{SLT}} = 4.44$ , see **Figure 6**).

In the drawing/narration, children are distributed across all concept levels; in the SLT, more than half of the children are now assessed at concept level 3a/3b (see **Table 3**).

At pre-test, 27.9% of the children are assessed at the same concept level, independently of test (SLT or drawing; post-test: 40%). 67.2% of children are classified higher in the SLT than in the drawing (post-test: 48%) and only occasionally (4.9%), the concept level in the drawing is higher than in the SLT (post-test: 12%).

## CONCLUSIONS AND DISCUSSION

The results of the study are of particular interest for two reasons. Firstly, it offers a diagnostic perspective with respect to the measurement of children's concepts using the specific example of the water cycle. Second inferences can be drawn regarding the reliability and validity of the different types of measurement.

The results show that children are rated higher in terms of their concept level when they receive support (SLT: recognition) than when they are asked to recall their knowledge from memory (both at pre-test and post-test, cf. hypothesis 2). Children show a significantly higher mean concept level in recognition than in the recall paradigm. On the individual level, only a small percentage of children (4.8% resp. 12%) were rated higher in recall compared to recognition. In contrast, and in line with hypothesis 1.2, 67.2% resp. 48% are rated higher in the recognition paradigm.



The results thus support the assumption that measurement can have a significant influence on diagnostics. Advantages and disadvantages of the respective elicitation methods must therefore be well weighed in studies of children's concepts: Survey methods that require free recall from memory tend to underestimate conceptual knowledge in children. Prior to engaging with the water cycle topic, children exhibited lower concept levels than in the recall paradigm.

Most often, children measured with recall are classified at the lowest concept levels, indicating no or less specific topical knowledge. This means that they cannot associate the subject matter with known knowledge structures in their memory (level 0: no concept) or do not connect their knowledge (level 1: isolated knowledge/concept). Thus, children may misunderstand the task and, despite existing subject knowledge, fail to recall it or recall concepts unrelated to the subject due to incorrect associations: conceptual knowledge would be present but is not recorded due to the survey methodology.

In contrast, in the recognition paradigm, few children are assigned to the lowest two levels. Accordingly, the stimuli allow them to better demonstrate or express their factual knowledge. As an example, the data of the girl in drawing "level 0" can be used. She drew a bathroom drain in the drawing task (and named it). However, by help of appropriate assistance in the SLT task, it could be shown that the girl indeed has conceptual ideas about the water cycle, which could be categorized as linear (level 2a).

A second challenge with eliciting children's concepts with more open-ended tasks is the lack of expressive ability or social cognition (e.g., increased importance of experimenter effects) of some children. Guided (measurement with aids) elicitation of children's concepts may have an opposite tendency: there is a possibility of overestimating children's knowledge by diagnosing ad-hoc or "on-the spot" constructs (cf. Deutsch, et al., 2016), which do not necessarily represent conceptual knowledge as such. An example of this can be found in the data of a subject who was diagnosed with a high conceptual level based on her solution of the structure-laying technique only. However, this finding could not be verified based on her explanations of the SLT. Further indication of this can be found in the data of the post-test: here, the proportion of children at the highest achieved concept levels 3a/3b is very high, whereas in the retrieval paradigm, all concept levels are equally represented. Therefore, it remains to be seen whether the exposure to the topic (after intervention by reading a factual text; post-test) allows all children to acquire knowledge as high as concept level 2b (or higher).

Consequently, when measuring children's concepts, the influence of the measurement instrument on the measurement must be considered. Rather few studies on children's concepts so far have taken this aspect into account in their research design. If adequate knowledge diagnostics is an important aspect of research, e.g., in the survey of children's concepts in content areas where little empirical evidence is available, it seems reasonable to combine several survey methods. Triangulation has the potential to mutually cancel or mitigate respective weaknesses of specific methods. In the present analysis, this seems to make sense especially because the two types of measurement show very high

correlations ( $r = .755$  resp.  $r = .559$ ) despite significant differences in mean values and thus presumably a high construct validity (hypothesis 1 can be confirmed).

However, in the context of experimental settings, e.g., when testing the effectiveness of different instruction methods, it seems to us that procedures are particularly useful which allow children to demonstrate their knowledge in the test task and support them accordingly by providing appropriate cues. The present data show that changes, independently of the data collection method, yield comparable results with respect to the learning progress (very high correlations between the two methods in both pre- and post-test and change scores). So, when doing quantitative research, which requires relatively large samples with good power statistical planning, those methods can be used for which data can be collected with less effort. Before being used in experimental settings, the corresponding methods should be checked regarding their diagnostic validity by means of a method combination, e.g., as in the case presented here.

The iterative construction of our proposed concept-level model that not only focusses on progression but also considers and ranges specific children's concepts, e. g. the reservoir model or linear vs. circumplex Ideas. In subject didactics, it seems essential to not only range children's concepts according to their correspondence with scientific concepts, but also regarding their subjective cognitive structure. For example, we identified the importance of understanding latent phenomena such as evaporation as key factors to reach level 3 in our model. Models like this are therefore suitable for more precise lesson planning concerning internal differentiation. Regarding hypothesis 3, it can be noted that in most cases, the drawings and children's narrations did not cause any difficulties to classify them into the corresponding concept levels. Of course, further evidence of the validity of our proposed level-system on the water cycle is needed.

At this point, it should be noted that the present experiment was a preliminary investigation. The data can only be interpreted cautiously due to partly slightly different procedures in the execution of the experiment (optimizing the experimental procedures in the field), limited resources regarding the evaluation (not all drawings and conversations were evaluated, respectively only limited possibilities to test the interrater reliability) and the only partly validated test instruments. A corresponding goal of the research was to understand better this last point for the present methodology, i.e., the measurement of children's concepts about different processes of retrieval (recall) and recognition.

Overall, however, the work shows that it is worthwhile to pay more attention to the connection between survey methods and results. Future studies could start here and look at two perspectives in more depth: Firstly, it is profitable to continue to question factual topics oriented to preconcepts. Secondly, it is important to develop a repertoire of methods to determine whether they tend to underestimate or overestimate children's knowledge. Based on our analyses, we recommend that researchers who are interested in diagnosing children's concepts use more than one measurement instrument ideally based on different cognitive processes (recall and recognition) or to give children different possibilities to express their knowledge (e.g., drawing, talking, and writing). We argue that



combining methods can increase diagnostic quality and lay the foundation for a better understanding of children's learning processes in science education.

**Author contributions:** ALI & MK: study conception and design, material preparation, and data collection and analysis & ALI: writing the first draft of the manuscript. Both authors agreed with the results and conclusions.

**Funding:** No funding source is reported for this study.

**Ethical statement:** The authors stated that the study adhered to the highest ethical practices applicable in scientific research. Participating children, their legal guardians and teachers were fully informed about the aims and procedures of the project. They were able to refuse data collection and intervention without giving reasons.

**Declaration of interest:** No conflict of interest is declared by the authors.

**Data sharing statement:** Data supporting the findings and conclusions are available upon request from the corresponding author.

## REFERENCES

- Adamina, M., Kübler, M., Kalcsics, K., Bietenhard, S., & Engeli, E. (Hrsg.). (2018). "Wie ich mir das denke und vorstelle ..." *Vorstellungen von Schülerinnen und Schülern zu Lerngegenständen des Sachunterrichts und des Fachbereichs Natur, Mensch und Gesellschaft* ["How I think and imagine it ..." Pupils' ideas about learning subjects in general education and the subject area of nature, man and society]. Klinkhardt.
- Aleknavičiūtė, V., Lehtinen, E., & Södervik, I. (2023). Thirty years of conceptual change research in biology—A review and meta-analysis of intervention studies. *Educational Research Review*, 41, Article 100556. <https://doi.org/10.1016/j.edurev.2023.100556>
- Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Psychology Press. <https://doi.org/10.4324/9781315802886>
- Brown, J. (Ed.). (1976). *Recall and recognition*. Wiley. <https://doi.org/10.1037/11314-016>
- Cardak, O. (2009). Science students' misconceptions of the water cycle according to their drawings. *Journal of Applied Sciences*, 9(5), 865-873. <https://doi.org/10.3923/jas.2009.865.873>
- Carey, S. (1985). *Conceptual change in childhood*. MIT Press.
- Carey, S. (1991). Knowledge acquisition: Enrichment or conceptual change. In S. Carey, & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition* (pp. 257-291). Lawrence Erlbaum.
- Deutsch, R., Gawronski, B., & Hofmann, W. (2016). *Reflective and impulsive determinants of human behavior*. Psychology Press. <https://doi.org/10.4324/9781315523095>
- di Sessa, A. A. (2008). A bird's-eye view of the "pieces" vs "coherence" controversy (from the "pieces" side of the "fence"). In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 35-60). Routledge. <https://doi.org/10.4324/9780203154472>
- Hardy, I., Jonen, A., Möller, K., & Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of "floating and sinking". *Journal of Educational Psychology*, 98(2), 307-326. <https://doi.org/10.1037/0022-0663.98.2.307>
- Harteringer, A., & Murmann, L. (2018). Schülervorstellungen erschliessen—Methoden, Analyse, Diagnose [Exploring student ideas—Methods, analysis, diagnosis]. 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 Lerngegenständen des Sachunterrichts und des Fachbereichs Natur, Mensch und Gesellschaft* (pp. 51-62). Klinkhardt.
- Heng, C. K., & Karpudewan, M. (2017). Facilitating primary school students' understanding of water cycle through guided inquiry-based learning. In K. Mageswary, Z. Ahmad Nurulazam, & A. L. Chandrasegaran (Eds.), *Overcoming students' misconceptions in science* (pp. 29-49). Springer. [https://doi.org/10.1007/978-981-10-3437-4\\_3](https://doi.org/10.1007/978-981-10-3437-4_3)
- Karacaoğlu, Ö. C., & Kasap, Y. (2023). The effects of reading comprehension skills on mathematics and science according to PISA data. *International Journal of Educational Research Review*, 8(3), 623-637. <https://doi.org/10.24331/ijere.1246885>
- Kintsch, W. (1970). Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory* (pp. 331-373). Academic Press. <https://doi.org/10.1016/B978-0-12-521350-9.50016-4>
- Kölbl, C., Tiedemann, J., & Billmann-Mahecha, E. (2006). Die Bedeutung der Lesekompetenz für Sachfächer [The importance of reading competence for subjects]. *Psychologie in Erziehung und Unterricht*, 53, 201-212.
- Koomson, C. K., & Owusu-Fordjour, C. (2018). Misconceptions of senior high school science students on evaporation and water cycle. *European Journal of Research and Reflection in Educational Sciences*, 6(5), 13-28.
- Koyuncu, I., & Firat, T. (2020). Investigating reading literacy in PISA 2018 assessment. *International Electronic Journal of Elementary Education*, 13(2), 263-275. <https://doi.org/10.26822/iejee.2021.189>
- Miner, J. T. (1992). An early childhood study of the water cycle [PhD dissertation, University of Nevada, Las Vegas]. <https://doi.org/10.25669/e7ih-h2ja>
- Möller, K. (2015). Genetisches lernen und conceptual change [Genetic learning and conceptual change]. In J. Kahlert, M. Fölling-Albers, M. Götz, A. Hartinger, S. Miller, & S. Wittkowske (Eds.), *Handbuch Didaktik des Sachunterrichts* (pp. 243-248). Klinkhardt.
- Möller, K. (2018). Die Bedeutung von Schülervorstellungen für das Lernen im Sachunterricht [The importance of student concepts for learning in science classes]. 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 Lerngegenständen des Sachunterrichts und des Fachbereichs Natur, Mensch und Gesellschaft* (pp. 35-50). Klinkhardt.

- Pacaci, C., Ustun, U., & Ozdemir, O. F. (2024). Effectiveness of conceptual change strategies in science education: A meta-analysis. *Journal of Research in Science Teaching*, 61(6), 1263-1325. <https://doi.org/10.1002/tea.21887>
- Relyea, J. E., & Hwang, H. (2024). Transactional development of science and mathematics knowledge and reading proficiency for multilingual students across languages of instruction. *Developmental Psychology*. <https://doi.org/10.1037/dev0001858>
- Reynolds, M. R., Niileksela, C. R., Gignac, G. E., & Sevillano, C. N. (2022). Working memory capacity development through childhood: A longitudinal analysis. *Developmental Psychology*, 58(7), 1254-1263. <https://doi.org/10.1037/dev0001360>
- Scholl, A. (2014). *Die Befragung* [The survey]. UVK.
- Smith, R., Snow, P., Serry, T., & Hammond, L. (2021). The role of background knowledge in reading comprehension: A critical review. *Reading Psychology*, 42(3), 214-240. <https://doi.org/10.1080/02702711.2021.1888348>
- Sodian, B. (2008). Entwicklung des Denkens [Development of thinking]. In R. Oerter, & L. Montada (Ed.), *Entwicklungspsychologie* (pp. 436-479). Beltz.
- Suryanti, M., Ibrahim, M., & Ledesma, N. S. (2018). Process skills approach to develop primary students' scientific literacy: A case study with low achieving students on water cycle. *IOP Conference Series: Materials Science and Engineering*, 296, Article 012030. <https://doi.org/10.1088/1757-899X/296/1/012030>
- Ullman, H., Almeida, R., & Klingberg, T. (2014). Structural maturation and brain activity predict future working memory capacity during childhood development. *The Journal of Neuroscience*, 34(5), 1592-1598. <https://doi.org/10.1523/JNEUROSCI.0842-13.2014>
- Ursavaş, N., & Genç, O. (2021). Enhancing middle school students' cognitive structure of water cycle through the use of water cycle educational game. *Kastamonu Educational Journal*, 29(1), 239-253. <https://doi.org/10.24106/kefdergi.808605>
- Vosniadou, S. (Ed.). (2008). *International handbook of research on conceptual change*. Routledge. <https://doi.org/10.4324/9780203874813>
- Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, 24, 535-585. [https://doi.org/10.1016/0010-0285\(92\)90018-W](https://doi.org/10.1016/0010-0285(92)90018-W)